UVOD

Teorija masovnog opsluživanja je dio teorije vjerovatnoće koji koristi njen matematički aparat, ali u posljednje vrijeme postaje samostalna nauka, jer se pomoću njenih modela mogu rješavati mnogi novi praktični zadaci. Primjeri takvih zadataka sreću se u različitim djelatnostima koje se bave opsluživanjem većeg broja ljudi, mašina ili prevoznih sredstava.

Masovno opsluživanje srećemo u slučajevima kada je potrebno organizovati opsluživanje većeg broja klijenata. Pod *klijentom* podrazumijevamo proizvoljan zahtjev ili porudžbinu za opsluživanjem. Klijent može da bude putnik koji želi da kupi kartu na biljetarnici, pacijent koji čeka ljekara u ambulanti, brod koji ulazi u luku i sa kojeg treba iskrcati, odnosno utovariti teret itd. Ako ne mogu biti odmah opsluženi, klijenti obrazuju *red.* Takvi su redovi u radnjama, na tramvajskim i autobuskim stanicama, u lukama (brodovi na vezovima) itd. Sredstva kojima se vrši opsluživanje nazivaju se *kanali opsluživanja.* To mogu biti automati za prodaju karata, biljetarnice, grupe radnika, luke za prijem brodova itd. Neravnomjernost u pojavljivanju klijenata na mjesto opsluživanja dovodi do nakupljanja zahtjeva za opsluživanjem, tj. do obrazovanja *redova klijenata*. Očigledno je da će se veći broj klijenata opslužiti ukoliko je veći broj kanala opsluživanja. Iso tako je jasno da je povećan broj kanala opsluživanja povezan sa većim rashodima, tj. sa angažovanjem većeg broja ljudi i materijalnih sredstava. Osim toga, mora se imati u vidu da ukupno vrijeme čekanja zajedno sa vremenom opsluživanja za klijenta bude prihvatljivo.

U vezi s naprijed navedenim, *teorija redova čekanja* u sklopu teorije masovnog opsluživanja istražuje *kriterijume optimalnosti* koji mogu biti različiti. Najvažniji kriterijumi sastoje se u zahtjevima da kanali opsluživanja što manje vremena stoje neangažovani i da klijenti uopšte ne čekaju ili čekaju na opsluživanje relativno kratko i za njih prihvatljivo vrijeme.

U glavi 1 definiše se Puasonova raspodjela koja spada u grupu diskretnih raspodjela sa beskonačno mnogo realizacija. Ta raspodjela je neposredno vezana za Puasonove procese koji igraju centralnu ulogu u opisivanju pomorskog transporta, kao i lučkih operacija, odnosno optimizacije odgovarajućih parametara. Nadalje se opisuje Puasonov potok događaja koji igra značajnu ulogu u teoriji redova čekanja. Zatim se definiše eksponencijalna raspodjela koja spada u grupu neprekidnih raspodjela. Daje se njena funkcija raspodjele i alternativna parametrizacija, kao i neka njena važna svojstva. Opisana je veza između *su vremena između pojavljivanja događaja međusobno statistički nezavisna i podliježu eksponencijalnoj raspodjeli sa parametrom* λ, *onda broj događaja u konačnom vremenskom intervalu t podliježe Puasonovoj raspodjeli sa parametrom* λt.

Nadalje se opisuje primjena ove raspodjele na Puasonovske procese opisane u prethodnom poglavlju. Na kraju ovog poglavlja daje se ocjena parametra normalne raspodjele metodom maksimalne vjerodostojnosti.

U šestom poglavlju razmatra se Erlangova raspodjela koja predstavlja uopštenje eksponencijalne raspodjele. Daje se formula za njenu gustinu i funkciju raspodjele. Konačno se opisuje primjena Erlangovih formula za određivanje vremena čekanja u teoriji redova čekanja.

U glavi 6 analiziran je model *M/D/k/N* reda čekanja sa nepromjenjivim prioritetima. Dobijen je izraz za vremena čekanja brodova u takvim redovima. Motivacija za analiziranje ovog tipa redova čekanja dolazi od problema koji nastaju iz dizajna ATM mreža u lukama ograničenog potencijala i sa višestrukim kanalima opsluživanja u proizvodnom okruženju trgovačkog tipa sa fiksiranim stovarišnim zonama i višestrukim klasama. Dobijeni numerički rezultati ilustruju varijacije vremena čekanja uporedo sa varijacijama parametara sistema, kao što su intenzitet dolaska, intenzitet opsluživanja, broj kanala opsluživanja, kapacitet luke i proporcija prioritetne klase. Ključna pretpostavka u izvođenju sastoji se u činjenici da je preostalo vrijeme za opsluživanje brodova nezavisno od vremena novog dolaska broda. Ovo je validna pretpostavka, jer uzimajući u obzir da se radi o saobraćaju visokog intenziteta u sistemu, brodovi ostaju zauzeti duži vremenski period.

1. ELEMENTI TEORIJE REDOVA ČEKANJA

1.1. Pojam i vrste redova čekanja

Teorija redova čekanja (*masovnog opsluživanja*) jedna je od metoda operacionih istraživanja koja proučava procese opsluživanja slučajno pristiglih jedinki ili zahtjeva za nekom uslugom, koristeći se pritom matematičkim modelima pomoću kojih se utvrđuje međuzavisnost između dolazaka jedinki, njihovog čekanja na uslugu, opsluživanja, te na kraju, izlaska jedinki iz sistema, s ciljem da se postigne optimalno funkcionisanje posmatranog sistema. Riješiti problem reda čekanja znači odrediti optimalan broj uslužnih mjesta za koji će vrijeme čekanja u redu ili troškovi (gubici) prouzrokovani čekanjem biti minimalni. Slijedi da se rješavanjem problema reda čekanja neće moći u potpunosti eliminisati čekanje, već samo gubici zbog čekanja svesti na minimum, jer da bi se eliminisalo čekanje jedinki, kapacitet uslužnih mjesta bi trebao biti veći od broja korisnika usluge, ali tada bi se povećala neiskorišćenost uslužnih mjesta, što je neracionalno.U teoriji redova čekanja se izučavaju fenomeni koji se odnose na zahtjeve za uslugama, s jedne, i analizu mogućnosti za zadovoljavanjem tih zahtjeva, s druge strane. Ovi zahtjevi se relativno često pojavljuju i najčešće su slučajnog (stohastičkog) karaktera.

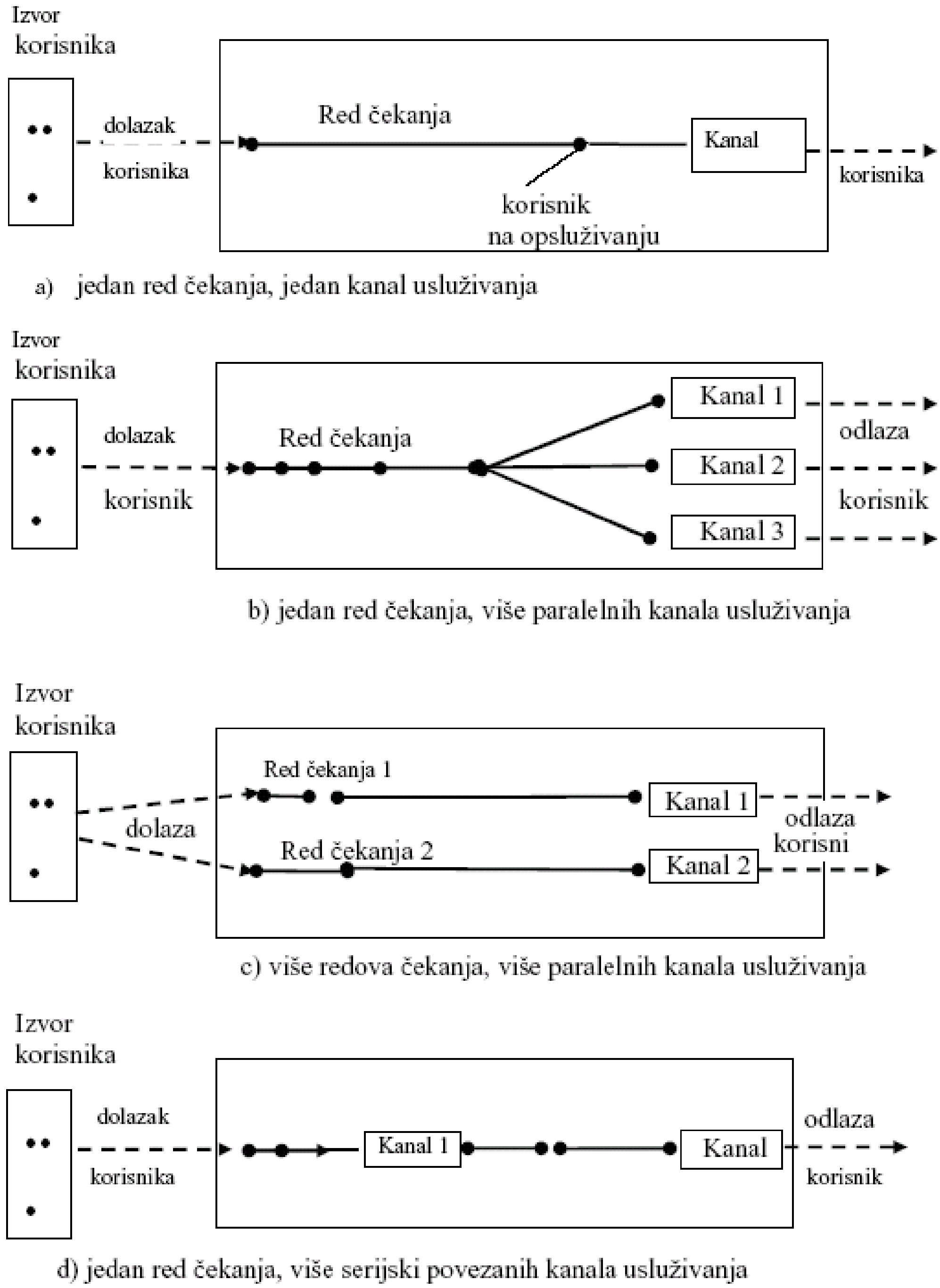
Sistem redova čekanja se može opisati kao jedan proces na sljedeći način: u nekim trenucima vremena koji su najčešće slučajnog (stohastičkog) karaktera, na mjesto usluživanja dolaze *korisnici* (*klijenti*) u nejednakim vremenskim razmacima, koje ćemo zvati *ulazni tok klijenata*. Ovi objekti, u zavisnosti od stanja u sistemu opsluživanja, odmah ili nakon izvjesnog vremena čekanja, pristupaju u *kanale opsluživanja* i nakon dobijene usluge napuštaju sistem redova čekanja i obrazuju *izlazni tok klijenata*.

Dakle, osnovni elementi koji karakterišu sistem redova čekanja su:

- ulazni tok klijenata,
- red čekanja,
- kanali opsluživanja i
- izlazni tok korisnika.

Što se tiče samih sistema redova čekanja - oni mogu biti različiti. Tako na primjer, u nekim sistemima se može dobiti samo jedna vrsta usluga, dok se u drugima može dobiti više različitih vrsta usluga. Prema broju kanala opsluživanja sisteme redova čekanja dijelimo na sisteme koji u svom sastavu imaju samo jedan kanal opsluživanja i zovu se *jednokanalni sistemi redova čekanja* i sisteme koji mogu imati više kanala opsluživanja i zovu se *višekanalni sistemi redova čekanja*.

Sljedećom šemom prikazane su različite vrste sistema redova čekanja:



Slika 1.1. Vrste sistema redova čekanja

Da bismo napravili analizu sistema redova čekanja potrebno je napraviti analizu svakog od njegovih elemenata. Ove analize su neophodna pretpostavka za formiranje odgovarajućeg matematičkog modela koji će na najbolji mogući način opisati funkcionisanje posmatranog sistema redova čekanja. Prilikom matematičkog modeliranja ulaznog toka korisnika, potrebno je imati u vidu da ovaj tok određuje spoljašnji svijet i da potencijalni broj korisnika, ili kako se često naziva populacija, koji u nekom trenutku mogu pripadati posmatranom sistemu redova čekanja, može biti konačan ili beskonačno veliki. Ako je ukupan broj potencijalnih korisnika *konačan*, onda je brojem korisnika u sistemu određen i preostali broj potencijalnih korisnika. Ako je broj potencijalnih korisnika *neograničen*, što se u praksi češće događa, onda broj korisnika u sistemu nema uticaja na ulazni tok korisnika. Taj broj potencijalnih korisnika, bez obzira da li je konačan ili je beskonačno veliki, smatraće se unaprijed poznatim.

Što se tiče broja korisnika i njihovog ponašanja u sistemu i to treba obuhvatiti matematičkim modelom, pri čemu treba imati u vidu sljedeće: broj korisnika u sistemu se stalno mijenja u vremenu, što zavisi od ulaznog i izlaznog toka korisnika. To je nepoznati broj, ali je potrebno znati *vjerovatnoću* sa kojom se može očekivati određeni broj korisnika u sistemu.

Broj korisnika u sistemu može se podijeliti na dva skupa. Jedan skup korisnika predstavljaće oni *korisnici koji se nalaze u kanalima opsluživanja*, u kojima se obavlja usluga, dok će drugi skup predstavljati *korisnike u redu čekanja*. Broj korisnika u jednom od ova dva skupa je uvijek poznat, dok je broj korisnika u drugom skupu nepoznat, ali je, isto kao i za ukupan broj korisnika u sistemu, poznata vjerovatnoća sa kojom se može očekivati određeni broj korisnika u ovom skupu. U slučaju da su svi kanali opsluživanja zauzeti, broj korisnika iz prvog skupa jednak je unaprijed poznatom broju kanala opsluživanja, dok je broj korisnika u redu čekanja nepoznat. U slučaju da svi kanali opsluživanja nisu zauzeti, broj korisnika u kanalima opsluživanja jednak je ukupnom broju korisnika u sistemu, koji je nepoznat, dok je broj korisnika u redu čekanja jednak nuli. To dalje implicira činjenicu da u tom slučaju i broj slobodnih kanala opsluživanja nije poznat.

Korisnici dolaze u sistem pojedinačno u različitim vremenskim trenucima. Vrijeme između dva dolaska je po pravilu *slučajna veličina* (promjenljiva), pa je prema tome potrebno odrediti očekivani broj dolazaka u jedinici vremena, kao i očekivano vrijeme između dva dolaska. Prilikom dolaska u sistem, korisnici se različito ponašaju. Neki korisnici obavezno staju u red i strpljivo čekaju na uslugu, dok drugi, čim naiđu na red odlaze iz sistema bez obavljene usluge. U najvećem broju slučajeva neki korisnici će ostati u sistemu, dok će drugi napustiti sistem, što znači da je i ponašanje korisnika u sistemu slučajno.

Kapacitet sistema redova čekanja predstavlja ukupan broj korisnika koje može da primi posmatrani sistem redova čekanja, a to je, naime, zbir broja korisnika koji se nalaze u redu čekanja i broja korisnika koji se opslužuju. U zavisnosti od samog sistema njegov kapacitet može biti *ograničen* ili *neograničen*. *Pravilo* ili *disciplina opsluživanja* se odnosi na redoslijed opsluživanja korisnika u sistemu redova čekanja. Neke manje-više standardne oznake se koriste za izražavanje redoslijeda opsluživanja korisnika u sistemu. Potrebno je pomenuti sljedeće:

- FIFO – redoslijed opsluživanja je određen redoslijedom pristizanja u sistem (First In First Out, tj. prvi prispjeli – prvi opslužen). Ovaj redoslijed opsluživanja je najčešći u praksi (česta je i oznaka FCFS-First Come First Served) .

- LIFO – redoslijed opsluživanja je obrnut redoslijedu dolazaka (Last In First Out, tj. posljednji prispjeli – prvi opslužen). Ovaj redoslijed se najčešće susreće npr. kod rada u skladištima, gdje se prvo uzimaju oni predmeti koji su posljednji dopremljeni u skladište (LCFS- Last Come First Served).

- PRI – redoslijed opsluživanja sa prioritetom. Npr. u zdravstvenoj ustanovi najprije se zbrinjavaju teži bolesnici, ili prilikom istovara prioritet ima lako kvarljiva roba, roba ograničenog vijeka trajanja.

- SIRO – slučajan izbor redoslijeda opsluživanja korisnika (Stohastic In Random Out). Ovo pravilo opsluživanja je tipično za mnoge telefonske sisteme.

Veličina populacije, odnosno potencijalni broj korisnika je *konačan* (prirodan broj) ili *beskonačna*. Ukoliko je broj korisnika ograničen, takvi sistemi se nazivaju *zatvoreni sistemi redova čekanja*. Osnovna karakteristika takvih sistema sastoji se u činjenici da broj korisnika u sistemu ima uticaj na ulazni tok korisnika, jer je brojem korisnika u sistemu određen i broj preostalih mogućih korisnika. Ako vrijeme između dva dolaska korisnika ne zavisi od toga koliko se korisnika nalazi u sistemu usvaja se da je veličina izvorne populacije neograničena. Takvi sistemi se nazivaju *otvoreni sistemi redova čekanja.*

Sve ove osobine sistema redova čekanja treba imati u vidu prilikom formiranja odgovarajućeg matematičkog modela redova čekanja, pomoću kojeg se može pristupiti izračunavanju određenih pokazatelja i veličina ili preformansi značajnih za posmatrani sistem redova čekanja, i to kako onih koje su bitne za korisnike usluga (klijente), tako i onih koje su bitne za davoce usluga (kanale opsluživanja). To znači da korišćenjem matematičkih modela treba odrediti i sljedeće pokazatelje za dati sistem redova čekanja:

- srednji broj korisnika koji pristižu u jedinici vremena,
- srednji broj korisnika koji mogu biti opsluženi u jedinici vremena,
- srednje vrijeme čekanja na uslugu,
- srednje vrijeme boravka korisnika u sistemu,
- srednji broj korisnika u sistemu,
- srednji broj korisnika u redu čekanja,
- srednji broj iskorišćenih kanala opsluživanja,
- srednji broj slobodnih kanala opsluživanja,
- prosječna iskorišćenost kanala opsluživanja,
- vjerovatnoća da će korisnik biti opslužen bez čekanja,
- odnos između srednjeg broj opsluženih i srednjeg broja prispjelih korisnika u jedinici vremena i dr.

Analizom ovih ukupnih rezultata treba pronaći optimalno rješenje koje će dati najbolje efekte i za primaoce (korisnike) i za davaoce usluga (kanale opsluživanja), iako su, što je ranije već rečeno, njihovi interesi suprotstavljeni jedni drugima.

1.2. Puasonova i eksponencijalna raspodjela

Puasonova raspodjela u teoriji vjerovatnoće i statistike je diskretna raspodjela vjerovatnoća koja izražava vjerovatnoću broja događaja koji se javljaju u fiksnom periodu vremena, pod uslovom da se ovi događaji javljaju sa poznatim *srednjom vrijednošću* (*intenzitetom*) i da su nezavisna od broja prošlih događaja.

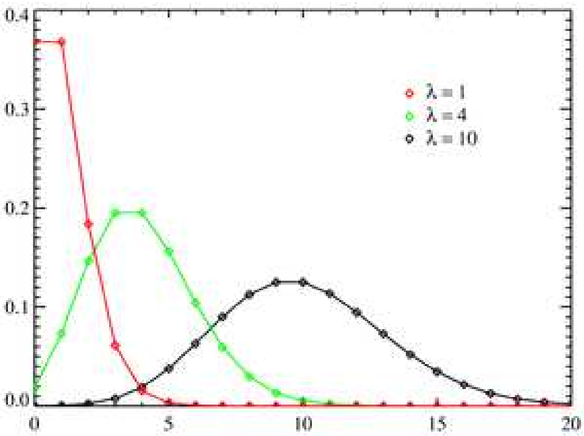
Za *diskretnu raspodjelu* $X = X(\lambda)$ sa beskonačno mnogo realizacija kažemo da ima Puasonovu raspodjelu ako je vjerovatnoća realizacija k događaja (k predstavlja nenegativan cijeli broj (k = 0, 1, 2, ...) data formulom

$$P(X = k) = f(k,\lambda) = \frac{e^{-\lambda}\lambda^k}{k!},\ k = 0,\ 1,\ 2,\ ...,$$

gdje je

- e osnova prirodnog logaritma (e = 2,71828...),
- k broj pojavljivanja glavnog događaja,
- k! faktorijel od k, tj. $k! = 1 \cdot 2 \cdots (k-1) \cdot k$ za $k > 0$ i $0! = 1$.
- λ pozitivan realan broj - *parametar raspodjele*, jednak očekivanom (srednjem) broju događaja koji se javljaju za vrijeme datog intervala.

Pokazuje se da je $\lambda = m(X) = \sigma^2(X)$, tj parametar λ je ujedno i *matematičko očekivanje* i *disperzija* pripadne Puasonove raspodjele.



Slika 1.2. Puasonova raspodjela - vjerovatnoće broja dogadjaja $f(k,\lambda)$. (na horizontalnoj osi su nanijete vrijednosti k. Funkcija $f(k,\lambda)$ je jedino različita od nule za cjelobrojne vrijednosti k).

Puasonova raspodjela igra značajnu ulogu u teoriji masovnog opsluživanja (teoriji redova čekanja), jer se u većini modela iz te teorije pokazuje da dolasci klijenata sa intenzitetom λ imaju Puasonovu raspodjelu $X(\lambda)$. U tom slučaju kažemo da se radi o *Puasonovom potoku događaja*, koji u poređenju sa drugim potocima, posjeduje osobine pogodne za efikasno rješavanje problema iz teorije masovnog opsluživanja.

Eksponencijalne raspodjela $T = T(\lambda)$ u teoriji vjerovatnoće spada u neprekidnu raspodjelu vjerovatnoća. *Funkcija gustine vjerovatnoća* eksponencijalne raspodjele data je formulom

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

gdje je $\lambda > 0$ parametar raspodjele.

Podsjetimo se da je funkcija raspodjele proizvoljne slučajne promjenljive X realna funkcija $F:(-\infty,+\infty) \to [0,1]$ definisana kao

$$F(x) = P(X < x) \text{ za svako } x \in (-\infty,+\infty).$$

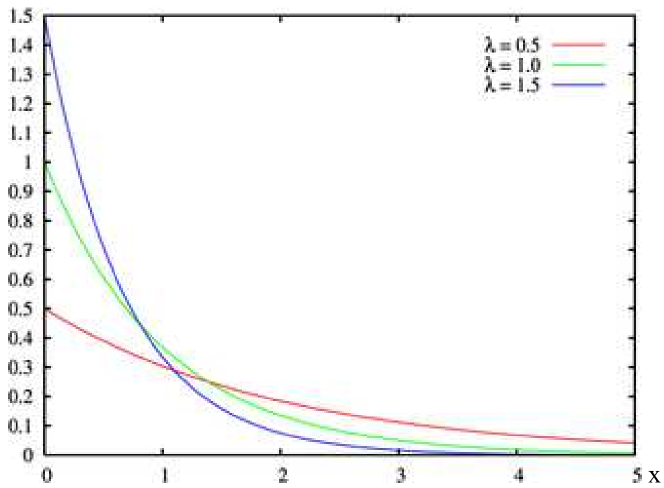
To znači da $F(x)$ predstavlja *vjerovatnoću da slučajna promjenljiva X poprimi vrijednost manju od x.* Preciznije, važi

$$F(x) = P(\{\omega : X(\omega) < x\}) \text{ za svako } x \in (-\infty,+\infty),$$

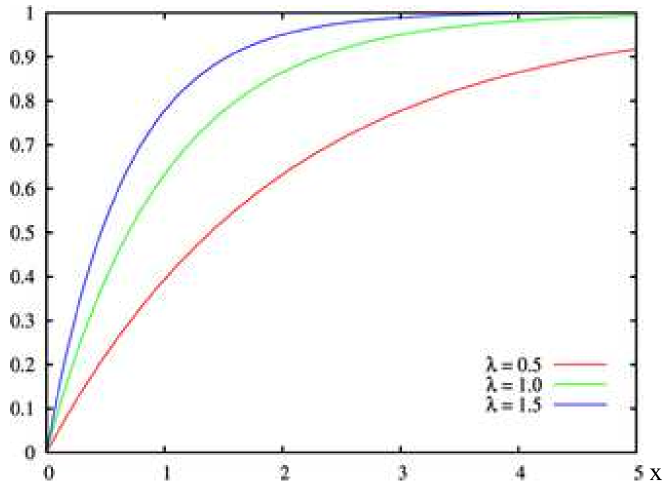
što čitamo kao: $F(x)$ predstavlja *vjerovatnoću događaja koji se sastoji od svih ishoda* ω *za koje slučajna promjenljiva X poprima vrijednost manju od x.*

Funkcija raspodjele eksponencijalne raspodjele data je na sljedeći način:

$$F(x;\lambda) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$



Slika 1.3. Eksponencijalna raspodjela.
Funkcija gustine vjerovatnoća $f(x,\lambda)$.



Slika 1.4. Funkcija raspodjele $F(x,\lambda)$ eksponencijalne raspodjele.

Uobičajeno korišćena alternativna parametrizacija definiše funkciju gustine vjerovatnoća eksponencijalne raspodjele formulom:

$$f(x;\beta)=\begin{cases}\frac{1}{\beta}e^{-x/\beta}, & x\geq 0,\\ 0, & x<0.\end{cases}$$

gdje $\beta > 0$ predstavlja recipročnu vrijednost naprijed definisanog parametra ($\beta = 1/\lambda$). U većini aplikativnih modela masovnog opsluživanja, β predstavlja *«vremenski parametar»*.

Eksponencijalna raspodjela kao i Puasonova raspodjela, ima veoma važnu ulogu u teoriji masovnog opsluživanja (teoriji redova čekanja), jer se u većini modela iz te teorije pokazuje da važi:

Vremena uzastopnih međudolazaka klijenata u sistem, a ponekad i vremena opsluživanja, imaju eksponencijalnu raspodjelu sa parametrom λ. *Pri tome je* $1/\lambda$ *prosječno vrijeme međudolazaka, odnosno srednje vrijeme opsluživanja.*

Drugim riječima, važi

$$m(X)=1/\lambda \quad \text{i} \quad \sigma^2(X)=1/\lambda^2 .$$

Veza između eksponencijalne i Puasonove raspodjele može se izraziti na sljedeći način:

Ako su vremena između pojavljivanja događaja međusobno statistički nezavisna i podliježu eksponencijalnoj raspodjeli sa parametrom λ, *onda broj događaja u* *konačnom vremenskom intervalu t podliježe Puasonovoj raspodjeli sa parametrom* λt. *Važi i obrnuto tvrđenje, tj. ako broj događaja u konačnom vremenskom intervalu t podliježe Puasonovoj raspodjeli sa parametrom* λt*, tada su vremena između pojavljivanja događaja međusobno statistički nezavisna i podliježu eksponencijalnoj raspodjeli sa parametrom* λ.

Važno svojstvo eksponencijalne raspodjele je njena *bezmemorijska osobina. To znači da ako je slučajna varijabla T eksponencijalno raspodijeljena, ona ispunjava uslov*

$$P(T < s + t | T > s) = P(T < t)\text{, za svako } s, t \geq 0.$$

Gornja matematička relacija znači da *vjerovatnoća da će se događaj desiti u vremenskom intervalu t zavisi samo od dužine intervala t, a ne zavisi od toga koliko dugo se događaj nije desio* (*tj. ne zavisi od s*).

1.3. Puasonovi potoci događaja

Kao što je rečeno u poglavlju 1.1, pojam klijenta u teoriji masovnog opsluživanja može se identifikovati sa događajem koji se realizuje na ulazu u sistem opsluživanja. Tako možemo govoriti o ulaznom potoku događaja (ulaznom toku klijenata).

Potok događaja je takav niz događaja koji proizilaze jedan za drugim u momentima vremena slučajno raspoređenim u posmatranom vremenskom intervalu.

Primjeri takvih potoka su: potok brodova koji dolaze u odredišnu luku, potok pješaka ka tramvajskoj stanici, potok telefonskih poziva preko automatske telefonske centrale itd.

Među potocima događaja poseban značaj ima *Puasonov potok događaja*, koji u poređenju sa drugim potocima, posjeduje osobine pogodne za efikasno rješavanje problema iz teorije masovnog opsluživanja. Puasonov potok događaja posjeduje osobine:

- *ordinarnost* i
- *odsustvo posljedica.*

Svojstvo ordinarnosti označava da klijenti pristupaju u sistem opsluživanja pojedinačno, tj. istovremena pojava dva klijenta u jednom momentu je skoro nemoguća.

Neordinarni potoci klijenata sreću se rjeđe u praksi nego ordinarni potoci klijenata. Primjer neordinarnog potoka klijenata je potok putnika koji izlaze iz lifta.. Ako u neordinarnom potoku klijenti pristupajau u parovima, trojkama itd. u ograničenim grupama, neordinarni potok se može svesti na ordinarni, ako se samo umjesto potoka pojedinih klijenata razmatraju potoci parova, trojki itd. Tako dolaske brodova u luku u baržama sa jednakim brojem brodova, možemo smatrati ordinarnim potokom događaja, ako svaku od barži smatramo jednim klijentom.

Sada objasnimo smisao izraza *odsustvo posljedica*. *Potok događaja posjeduje ovu osobinu ako broj događaja* X_1 *koji padaju na na interval vremena* t_1 *ne zavisi od broja događaja* X_2 *koji padaju na na interval vremena* t_2*, uz uslov da su intervali* t_1 *i* t_2 *disjunktni* (*da se ne poklapaju*). To drugim riječima znači da su slučajne promjenljive X_1 i X_2 međusobno nezavisne, tj. da važi

$$P(X_2 = m_2 / X_1 = m_1) = P(X_2 = m_2), \text{ za svako } m_1 = 0,1,2,...; \quad m_2 = 0,1,2,....$$

U kursevima teorije vjerovatnoće pokazuje se da

kod Puasonovog potoka događaja broj događaja X *koji padaju na proizvoljan fiksiran interval vremena* $(t, t+\Delta t)$ *ima Puasovu raspodjelu:*

$$P_{t,\Delta t}(X = m) = \frac{[\lambda(t,\Delta t)]^m}{m!} e^{-\lambda(t,\Delta t)}, \quad m = 0,1,2,...,$$

gdje je $\lambda = \lambda(t,\Delta t)$ parametar Puasonove raspodjele koji zavisi o t i Δt.

Ako parametar $\lambda = \lambda(t,\Delta t)$ *ovisi samo o dužini intervala* Δt*, kažemo da se radi o stacionarnom Puasonovom potoku događaja.*

Takav potok događaja, tj. potok koji posjeduje svojstva:

- *ordinarnost,*
- *odsustvo posljedica i*
- *stacionarnost*

naziva se prost potok događaja.

Prost potok događaja je tako nazvan zato što pretpostavka o prostom ulaznom potoku klijenata pri analizi različitih sistema masovnog opsluživanja dovodi do najjednostavnijih rješenja.

Ako se radi o prostom potoku događaja, tada srednji broj događaja $\lambda = \lambda(t,\Delta t)$ iz gornje formule ne zavisi o t i on zapravo iznosi $\lambda(\Delta t) = \lambda \cdot \Delta t$. Otuda slijedi da je vjerovatnoća da se na proizvoljno izabranom intervalu vremena Δt pojavi m događaja jednaka

$$P_{t,\Delta t}(X = m) = \frac{[\lambda \cdot (\Delta t)]^m}{m!} e^{-\lambda \cdot (\Delta t)}, \quad m = 0,1,2,...,$$

Vidimo da gornja formula predstavlja zapravo zakon Puasonove raspodjele sa parametrom $\lambda \cdot (\Delta t)$. Uočimo da za $\Delta t = 1$ taj parametar iznosi λ, pa se u tom slučaju radi o Puasonovoj raspodjeli sa parametrom λ.

Taj parameter λ predstavlja očekivani broj klijenata u prostom potoku i naziva se intenzitet ili *gustina toga potoka.*

Nadalje, na osnovu veze između Puasonove i eksponencijalne raspodjele date u poglavlju 1.2, slijedi važno tvrđenje:

U prostom potoku događaja intenziteta λ interval vremena između proizvoljna dva susjedna događaja (klijenta) ima eksponencijalnu raspodjelu sa parametrom λ.

1.4. Erlangova raspodjela i Erlangovi potoci događaja

Erlangova raspodjela spada u neprekidne raspodjele vjerovatnoća i predstavlja uopštenje eksponencijalne raspodjele. Erlangovu raspodjelu je otkrio danski inženjer A. K. Erlang ispitujući broj telefonskih poziva u telefonskom saobraćajnom inženjeringu, kao i vremena čekanja u sistemima *redova čekanja.* Kasnije je ova raspodjela našla široku primjenu u teoriji masovnog opsluživanja.

Erlangova raspodjela uzima pozitivne vrijednosti za sve realne brojeve veće od nule. Data je pomoću dva parametra: položaja *k*, koji je nenegativan cijeli broj i parametra λ, koji je pozitivan realan broj.

Funkcija gustine vjerovatnoća Erlangove raspodjele $E_k(\lambda)$ sa parametrima *k* i λ data je formulom

$$f(x;k,\lambda)=\frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!} \text{ za } x>0; \quad f(x;k,\lambda)=0 \text{ za } x\le 0.$$

Matematičko očekivanje Erlangove raspodjele $E_k(\lambda)$ dato je formulom

$$m(E_k(\lambda))=\frac{k}{\lambda}.$$

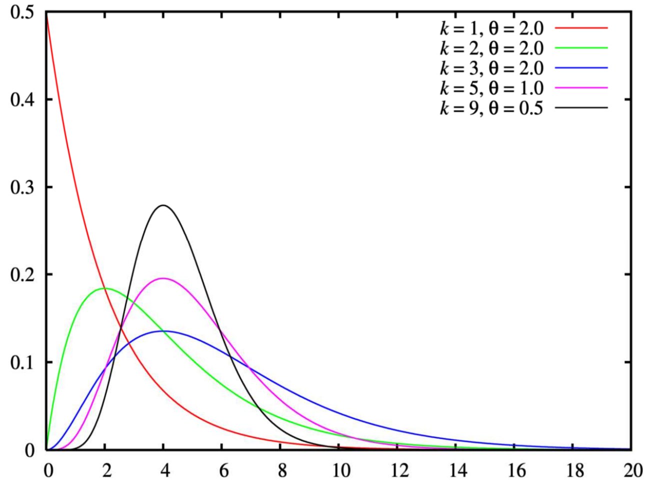
Disperzija od $E_k(\lambda)$ data je kao

$$\sigma^2(E_k(\lambda))=\frac{k}{\lambda^2}.$$

Uočimo da važi:

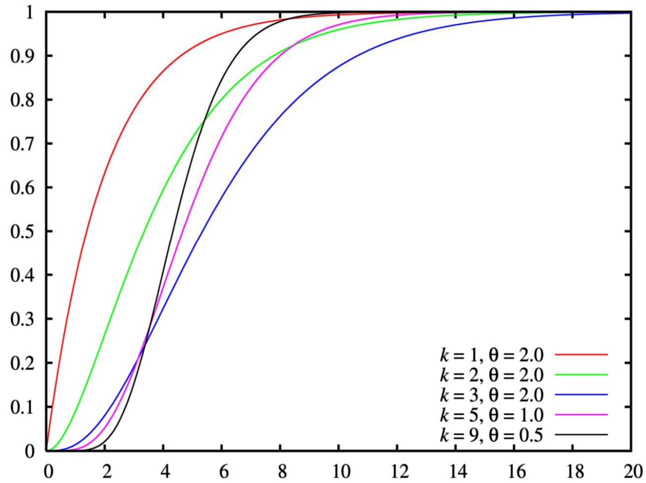
za $k=1$ Erlangova raspodjela svodi se na prethodno opisanu eksponencijalnu raspodjelu sa istim parametrom λ. Parametar k se često naziva i broj faza opsluživanja Erlangove raspodjele.

Erlangova raspodjela predstavlja specijalni slučaj *Gama raspodjele* za koju je *k* cijeli broj.



Slika 1.5. Funkcija gustine Erlangove raspodjela - $f(x,k,\lambda)=\dfrac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}$

(na horizontalnoj osi se nanose vrijednosti *k,*a na vertikalnoj vrijednosti $f(k,\lambda)$).



Slika 1.6. Funkcija raspodjele Erlangove raspodjele
(na horizontalnoj osi se nanose vrijednosti *k,*a na vertikalnoj vrijednosti $F(x,k,\lambda)$).

Naprijed definisanu Erlangovu raspodejelu sa parametrima λ i *k*, označavaćemo sa $E_k(\lambda)$ ili ukoliko se vrijednost λ smatra poznatom, kratko sa E_k.

Erlangova raspodjela $E_k(\lambda)$ *opisuje situaciju kada se od k međusobno nezavisnih i identično raspodijeljenih slučajnih prromjenljivih koje podliježu eksponencijalnoj raspodjeli sa parametrom* λ *posmatra svaki k-ti po redu.* To znači da Erlangova raspodjela opisuje raspodjelu sume

$$E_k = T_1 + T_2 + ... + T_k,$$

gdje su $T_1, T_2, ..., T_k$ međusobno nezavisne slučajne veličine raspodijeljene po eksponencijalnoj raspodjeli sa parametrom λ.

U praksi se prethodno svojstvo odnosi na slučajeve kada se jednom klijentu pruža usluga svakih *k* puta, pri čemu trajanje te usluge podliježe eksponencijalnoj raspodjeli. U vezi s tim, uvodimo pojam Erlangovih potoka događaja.

Erlangovi potoci događaja nastaju prosijavanjem ili prorjeđivanjem prostih potoka definisanih u poglavlju 1.3. *Ako u prostom potoku zadržimo svakog drugog klijenta, a ostale odbacujemo, onda tako prosijan potok predstavlja novi potok događaja koji se naziva Erlangov potok drugog reda. Uopšte, Erlangov potok k-tog* $(k = 1,2,...)$ *reda je potok događaja koji se dobije tako što se u prostom potoku zadrži svaki k-ti klijent po redu, a ostali se odbace.*

Na osnovu prethodno uočene veze između Erlangove i eksponencijalne raspodjele pokazuje se da važi tvrđenje:

U Erlangovom potoku događaja k-tog $(k = 1,2,...)$ ***reda koji se odnosi na prost potok intenziteta*** λ***, interval vremena između proizvoljna dva susjedna događaja (klijenta) ima Erlangovu raspodjelu sa parametrom*** λ***.***

2. MATEMATIČKO MODELIRANJE REDOVA ČEKANJA

2.1. Obilježavanje modela redova čekanja

Funkcionisanje sistema redova čekanja zavisi od niza različitih faktora, kao što su: broj klijenata u sistemu, broj kanala opsluživanja, broj potencijalnih klijenata, ulaznog i izlaznog toka klijenata, načina ponašanja klijenata u sistemu itd. Pomenuti faktori pored toga što utiču na uslove i način funkcionisanja sistema redova čekanja, omogućuju izvođenje i klasifikacije redova čekanja. Jedna od najpoznatijih, ujedno i najjednostavnijih klasifikacija sistema redova čekanja izvodi se pomoću *Kendal-Lijeve* notacije, koja obuhvaća sljedećih šest obilježja:

$$(a/b/c) : (/d/e/f),$$

gdje je:

a – raspodjela vremena dolazaka (ulazni tok klijenata)
b – raspodjela vremena opsluživanja (izlazni tok klijenata)
c – broj kanala opsluživanja
d – kapacitet sistema redova čekanja
e – pravilo ili disciplina opsluživanja
f - veličina populacije (broj potencijalnih klijenata).

Napomenimo da je prve tri klasifikacione oznake za sisteme masovnog opsluživanja (*a/b/c*) uveo D.G. Kendall 1953. godine, a nešto kasnije, 1966. godine A.M. Lee je uveo obilježja *d* i *e*.

Često se posljednje tri oznake ili neka od njih izostavljaju u označavanju sistema. U tim slučajevima implicitno se smatra da je kapacitet sistema neograničen, da je sistem opsluživanja FIFO ili da je izvorna populacija ograničena.

Prva oznaka a definiše tip toka dolazaka ili raspodjelu vremena dolazaka. Uobičajene su sljedeće oznake:

M – Puasonov tok ili eksponencijalna raspodjela
E_k – Erlangov tok ili gama raspodjela
D – deterministički tok
G – opšti ili bilo koji tok.

Iste oznake primjenjuju se i za tok odlazaka, odnosno raspodjele vremena opsluživanja koji su označeni sa b.

Broj kanala opsluživanja c je prirodan broj, tj. $c \in \{1,2,3,...\}$.

Kapacitet sistema d je, kao što smo već rekli, ukupan broj klijenata koji mogu da stanu u sistem, tj. zbir kljenata na opsluživanju i u redu čekanja. Ako se u oznaci ovaj broj izostavi, onda se pretpostavlja da je kapacitet neograničen.

Pravilo ili disciplina opsluživanja e određuje redoslijed redoslijed oplуživanja klijenata koji dolaze u sistem. Kao što je u poglavlju 1.1 opisano, taj redoslijed može biti FIFO, LIFO, PRI ili SIRO.

Veličina populacije f je prirodan broj, tj. $f \in \{1,2,3,...\}$. Ako vrijeme između dva dolaska klijenta ne zavisi od toga koliko se kljienata nalazi u sistemu, onda se usvaja da je veličina izvorne populacije neograničena.

Često se zadnje tri oznake (*d*, *e* i *f*) ili neka od njih, izostavljaju u označenim sistemima. U tim slučajevima se podrazumijeva da je kapacitet sistema neograničen, da je sistem opsluživanja FIFO ili da je izvorna populacija neograničena.

2.2. Osnovni pojmovi u matematičkom modeliranju redova čekanja

U matematičkom modeliranju koristićemo sljedeće pojmove i oznake:

n - stanje sistema, tj. broj klijenata u sistemu na opsluživanju i u redu čekanja

$P_n(t)$ -vjerovatnoća da se tačno n klijenata nalazi u sistemu opsluživanja u trenutku t

P_n - vjerovatnoća da se u stacionarnom (ustaljenom) režimu u sistemu nalazi n klijenata

s – broj kanala opsluživanja

λ_n - intenzitet dolazaka novih klijenata kada se u sistemu nalazi n klijenata (očekivani broj dolazaka u jedinici vremena)

μ_n - intenzitet opsluživanja kada se u sistemu nalazi n klijenata (očekivani broj klijenata kojima se završava usluga u jedinici vremena). Zapravo, μ_n predstavlja kombinovani intenzitet kojim zauzeti kanali opsluživanja (oni koji daju uslugu klijentu) završavaju usluge.

λ - intenzitet dolazaka, tj. broj dolazaka u jedinici vremena

μ - intenzitet opsluživanja, tj. broj usluga u jedinici vremena

L – očekivani broj klijenata u sistemu

L_q - očekivana dužina reda čekanja

w – vrijeme koje jedan klijent provede u sistemu. To je zapravo slučajna promjenljiva koja uključuje vrijeme čekanja i vrijeme opsluživanja.

$W = E(w)$ - matematičko očekivanje vremena w (prosječno vrijeme koje klijent provede u sistemu)

w_q – vrijeme koje jedan klijent provede u redu čekanja. To takođe predstavlja slučajnu promjenljivu.

$W_q = E(w_q)$ - matematičko očekivanje vremena w_q (prosječno vrijeme čekanja)

$P(A)$ - vjerovatnoća događaja A

$P(A \mid B)$ - vjerovatnoća događaja A pod uslovom da se desio događaj B.

2.3. Litlova formula

Jedan od važnih rezultata koji se mnogo koriste u teoriji redova čekanja predstavlja Litlova formula koja na jednostavan način *povezuje očekivani broj klijenata u sistemu opsluživanja L, intenzitet dolazaka* λ *i očekivano zadržavanje klijenta u sistemu W. Na osnovu Litlove formule važi*

$$L = \lambda W . \tag{2.1}$$

Njeno ime pripada autoru koji je prvi dao njen strogi dokaz (J. D. Little, 1961). Primjenjuje se u slučajevima kada je intenzitet dolazaka λ *konstantan* ili kada se može izračunati srednji intenzitet dolazaka λ_{sr}

Formula (2.1) je bila poznata i prije Litlovog rada sa njenim strogim matematičkim dokazom, jer se intuitivno može naslutiti i protumačiti. Zapravo formulu (2.1) iskazujemo kao:

Očekivani broj klijenata u sistemu jednak je proizvodu intenziteta (učestanosti) dolazaka klijenata i matematičkog očekivanja (prosječnog) vremena koje klijent provede u sistemu (u redu i na opsluživanju ukupno).

Primjer 1. pretpostavimo da u neki višekanalni sistem opsluživanja dolazi 12 klijenata na sat koji se u njemu prosječno zadržavaju po 20 minuta, odnosno po $1/3$ sata. Tada na osnovu Litlove formule prosječan broj klijenata za veoma dugi vremenski period u sistemu iznosi

$$L = \lambda W = 12 \cdot \frac{1}{3} = 4.$$

Do istog zaključka bismo došli i bez poznavanja Litlove formule.

Napomenimo da u navedenoj formuli L označava očekivani broj klijenata u sistemu koji uključuje i klijente na opsluživanju. Isto tako, W je očekivano ukupno vrijeme koje klijent provede u redu i na opsluživanju.

Analogna formuli (2.1) *je takođe Litlova formula koja daje vezu između između očekivane dužine reda* L_q *i očekivanog vremena* W_q *provedenog u redu. Na osnovu te formule važi*

$$L_q = \lambda . W_q \ . \tag{2.2}$$

Formulu (2.2) iskazujemo kao:

Očekivani broj klijenata u redu čekanja jednak je proizvodu intenziteta (učestanosti) dolazaka klijenata i matematičkog očekivanja (prosječnog) vremena koje klijent provede u redu čekanja.

Primjer 2. Pretpostavimo da u neki višekanalni sistem opsluživanja dolazi 12 klijenata na sat koji se u redu čekanja prosječno zadržavaju po 10 minuta, odnosno po $1/6$ sata. Tada na osnovu Litlove formule za veoma dugi vremenski period prosječan broj klijenata u redu čekanja iznosi

$$L_q = \lambda W_q = 12 \cdot \frac{1}{6} = 2.$$

Ako je intenzitet opsluživanja μ konstantan, tj. ako je matematičko očekivanje vremena opsluživanja konstantno, onda je očekivano ukupno vrijeme provedeno u sistemu dato formulom.

$$W = W_q + \frac{1}{\mu} \ . \tag{2.3}$$

Formulu (2.3) iskazujemo kao:

Prosječno vrijeme koje klijent provede u sistemu jednako je zbiru prosječnog vremena koje kljient provede u redu čekanja i prosječnog vremena koje klijent provede na opsluživanju.

Primjer 3. Pretpostavimo da u neki višekanalni sistem opsluživanja dolazi 12 klijenata na sat koji se u njemu prosječno zadržavaju po 20 minuta, odnosno po $1/3$ sata, dok se isti u redu čekanja prosječno zadržavaju po 10 minuta, odnosno po $1/6$ sata Tada na osnovu formule (2.3) dobijamo

$$\frac{1}{\mu} = W - W_q = \frac{1}{3} - \frac{1}{6} = \frac{1}{6},$$

odakle slijedi $\mu = 6$. Dakle, intenzitet opslučivanja u ovom sistemu iznosi 6 (klijenata na sat), a samim tim prosječno vrijeme opsluživanja iznosi $1/6$ sata, odnosno 10 minuta.

Formule (2.1), (2.2) i (2.3) *povezuju četiri važne veličine u analizi redova čekanja:* L, W, L_q *i* W_q.

Na osnovu formula (2.1) i (2.2) redom dobijamo $W = L/\lambda$ i $W_q = L_q/\lambda$, što uvršteno u formulu (2.3) daje

$$\frac{L}{\lambda} = \frac{L_q}{\lambda} + \frac{1}{\mu}.$$

Otuda slijedi

$$\frac{L - L_q}{\lambda} = \frac{1}{\mu},$$

odnosno

$$\mu = \frac{\lambda}{L - L_q}. \tag{2.4}$$

Primjer 4. Pretpostavimo da u neki višekanalni sistem opsluživanja prosječno dolazi 12 klijenata na sat ($\lambda = 12$). Nadalje pretpostavimo da za veoma dugi vremenski period prosječan broj klijenata u sistemu iznosi 4 ($L = 4$), a prosječan broj klijenata u redu čekanja iznosi 2 ($L_q = 2$). Tada na osnovu formule (2.4) dobijamo

$$\mu = \frac{\lambda}{L - L_q} = \frac{12}{4 - 2} = 6.$$

Dakle, intenzitet opsluživanja u ovakvom sistemu iznosi 6 (klijenata na sat), a samim tim prosječno vrijeme opsluživanja iznosi $1/6$ sata, odnosno 10 minuta. Uočimo da je taj rezultat u skladu sa rezultatom dobijenim u primjeru 3.

Konačno, na osnovu formule (2.4) imamo

$$\lambda = \mu(L - L_q) \tag{2.5}$$

Primjer 5. Pretpostavimo da u nekom višekanalnom sistemu opsluživanja iznosi 6 klijenata na sat. ($\mu = 6$). Nadalje pretpostavimo da za veoma dugi vremenski period prosječan broj klijenata u sistemu iznosi 4 ($L = 4$), a prosječan broj klijenata u redu čekanja iznosi 2 ($L_q = 2$). Tada na osnovu formule (2.5) dobijamo

$$\lambda = \mu(L - L_q) = 6(4 - 2) = 12.$$

Dakle, u ovakav sistem opsluživanja prosječno dolazi 12 klijenata na sat. Uočimo da je taj rezultat u skladu sa rezultatom dobijenim u primjeru 4.

3. STOHASTIČKI PROCESI I MARKOVLJEVI LANCI

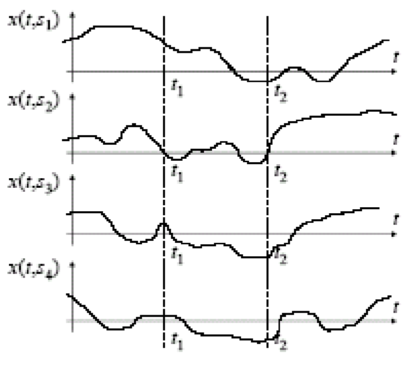
3.1 Pojam slučajnog procesa

Neke su pojave po prirodi determinističke, a neke slučajne ili stohastičke. Često se i složene determinističke pojave tretiraju kao slučajne, jer bi deterministički problem bio prekomplikovan za modeliranje. Prilikom rješavanja konkretnih problema iz oblasti redova čekanja koriste se različite matematičke metode. S obzirom da sistem redova čekanja predstavlja jedan proces koji karakterišu slučajne veličine, to sistemi redova čekanja u suštini nisu ništa drugo do različite vrste *slučajnih ili stohastičkih procesa* (*random process*). Upravo zbog toga, pri analizi i rješavanju različitih problema vezanih za sisteme redova čekanja, najčešće se koriste matematičke metode izgrađene za izučavanje stohastičkih procesa.

Postoji više različitih definicija i opisa slučajnih procesa, a najčešće se koristi definicija:

Slučajni proces je slučajna promjenljiva (veličina) koja zavisi od vremena. Pri tome slučajna promjenljiva x je preslikavanje koje svakom *ishodu eksperimenta s dodjeljuje broj x(s), koji predstavlja realizaciju te slučajne promjenljive. Slučajni proces je preslikavanje koje svakom ishodu eksperimenta s dodjeljuje funkciju vremena x(t,s)* - *jednu njegovu realizaciju.*

Dakle, pojam slučajnog procesa se dobija iz pojma slučajne promjenljive dodavanje vremenske varijable, što se grafički može predstaviti na sljedeći način:



Za $t = t_1$ $X(t_1 = s)$ je slučajna promjenljiva.

Može se koristiti i druga opštija i malo složenija definicija po kojoj je slučajni proces indeksirana kolekcija slučajnih promjenljivih $\{X_t\}$ koje se mijenjaju na slučajan način, u zavisnosti od vremena, tj. prolaskom indeksa t kroz dati skup T. Slučajni proces se obično označava relacijom

$$\{X(t) \mid t \geq 0\},$$

koja pokazuje da *posmatrana sluičajna promjenljiva može da uzima različite diskretne i kontinualne vrijednosti koje se nazivaju stanjima procesa ili stanjima sistema koji je predstavljen posmatranim procesom. Skup svih vrijednosti koje treba uzeti slučajna promjenljiva, ili skup svih stanja slučajnog procesa, naziva se prostor stanja.*

Klasifikacija slučajnog procesa zavisi od sljedeće tri veličine: *prostora stanja, indeksnog parametra t, a to je u sistemima redova čekanja vrijeme, i statističke zavisnosti između slučajnih promjenljivih X(t) za različite vrijednosti indeksnog parametra t.* U zavisnosti od toga da li slučajna promjenljiva uzima samo diskretne ili samo kontinualne vrijednosti, stanja slučajnog procesa mogu biti *diskretna* ili *kontinualna*. Slučajni proces sa diskretnim prostorom stanja naziva se *lanac*. Prostor stanja za lanac je obično skup prirodnih brojeva uključujući i nulu: $\{0,1,2,...\}$.

Ako se stanja slučajnog procesa mogu mijenjati samo u unaprijed određenim vremenskim trenucima koji su konačni ili prebrojivi, kažemo da se radi o *vremenski diskretnom slučajnom procesu*. U tom slučaju slučajni proces se može označiti sa

$$\{X_k \mid k \in K\},$$

gdje je sa K označen konačan ili prebrojiv skup svih stanja posmatranog procesa. Tada se X_k naziva *slučajni* ili *stohastički niz*. U protivnom, ako se stanja slučajnog procesa mogu mijenjati u bilo kom vremenskom trenutku na ograničenom ili neograničenom vremenskom intervalu, tada kažemo da se radi o *vremenski neprekidnom* ili *kontinualnom slučajnom procesu*. Bitna karakteristika slučajnog procesa je međusobni odnos slučajnih promjenljivih.

Kao primjer vremenski diskretnog slučajnog procesa iz oblasti teorije redova čekanja može se navesti broj klijenata u sistemu redova čekanja. U zavisnosti od vremena, broj klijenata u sistemu se mijenja i može uzimati bilo koju vrijednost iz skupa prirodnih brojeva uključujući i nulu, što znači da je broj klijenata u sistemu slučajni proces čija su stanja određena skupom prirodnih brojeva uključujući i nulu pa je to vremenski diskretan slučajni proces ili lanac.

3.2. Markovljevi lanci

U teoriji redova čekanja posebnu ulogu igraju slučajni procesi koji posjeduju svojstvo Markova i koji se po ruskom matematičaru A.A. Markovu (1856-1922), koji ih je prvi proučavao, nazivaju *lanci Markova* ili *Markovljevi lanci*. Svojstvo Markova se može opisati na više različitih načina, a njegova suština se sastoji u sljedećem.

Neka se posmatrani sistem, koji je opisan Markovljevim lancima, nalazi u nekom stanju koje ćemo nazvati *sadašnje stanje*. *Ako vjerovatnoća da sistem iz sadašnjeg stanja* *pređe u određeno buduće stanje zavisi samo od ova dva stanja* (*sadašnjeg i budućeg*), *a ne zavisi od toga kako je sistem dospio u sadašnje stanje, kaže se da takav sistem ima svojstvo Markova.*

Dakle, to su sistemi čija je čitava predistorija sadržana u sadašnjem stanju, a o ponašanju sistema u prethodnim stanjima ne postoje nikakve informacije. Takvi sistemi se nazivaju *sistemi bez posljedica* ili *sistemi bez memorije* (*pamćenja*), odnosno *sistemi sa odsustvom pamćenja* (*memorije*).

Svojstvo Markova se može iskazati i terminologijom redova čekanja na sljedeći način.

Vjerovatnoća da će u vremenskom intervalu $(t, t+\tau)$ *doći n klijenata u sistem redova čekanja ne zavisi od broja klijenata koji su došli u sistem prije trenutka t, nego zavisi samo od dužine vremenskog intervala τ, odnosno, ostvareni broj dolazaka klijenata u sistem redova čekanja u prethodnom vremenskom intervalu nema nikakvog uticaja na dolaske u narednom intervalu. To znači da su dolasci klijenata u sistem redova čekanja nezavisni.*

Ova osobina Markova može se izraziti analitički za vremenski diskretni stohastički proces $\{X_t\}$pomoću relacije:

$$P\{X_{t+1} = j \mid X_0 = k_0, \ldots, X_{t-1} = k_{t-1}, X_t = i\} = P\{X_{t+1} = j \mid X_t = i\},$$

za t = 0,1,2,... i svaki niz $\{i, j, k_0, k_1, \ldots, k_{t-1}\} \subset K$.

Uslovne vjerovatnoće $P\{X_{t+1} = j \mid X_t = i\}$ nazivaju se *jednostepene vjerovatnoće prelaza Markovljevog lanca*. Ako za svako i, j i t = 0,1,2,..., važi jednakost

$$P\{X_{t+1} = j \mid X_t = i\} = P\{X_1 = j \mid X_0 = i\},$$

onda kažemo da su jednostepene vjerovatnoće prelaza *ustaljene* ili *stacionarne*, *što znači da se te vjerovatnoće ne mjenjaju u vremenu.* Ovakve vjerovatnoće se obično obilježavaju sa p_{ij} i predstavljaju vjerovatnoće prelaza iz stanja i u stanje j u jednoj vremenskoj promjeni ili u jednom koraku. Mogu se definisati i *višestepene* (n-stepene) *stacionarne vjerovatnoće prelaza*, koje se obilježavaju sa $p_{ij}^{(n)}$ i koje predstavljaju uslovnu vjerovatnoću da slučajna promjenljiva X, koja počinje sa stanjem i, bude u stanju j poslije tačno n promjena u vremenu, odnosno nakon n koraka. Ako za svako i, j, n = 0,1,2,... i t = 0,1,2,..., važe sljedeće jednakosti:

$$P\{X_{t+n} = j \mid X_t = i\} = P\{X_n = j \mid X_0 = i\} = p_{ij}^{(n)}$$

kažemo da su vjerovatnoće $p_{ij}^{(n)}$ *n-stepene* (u opštem slučaju *višestepene*) *vjerovatnoće prelaza*.

Vrijednosti $p_{ij}^{(n)}$ kao uslovne vjerovatnoće su nenegativni brojevi. Proces se poslije promjene stanja mora naći u nekom od svojih stanja, pa otuda proizilazi da vjerovatnoće $p_{ij}^{(n)}$ moraju zadovoljavati sljedeće uslove:

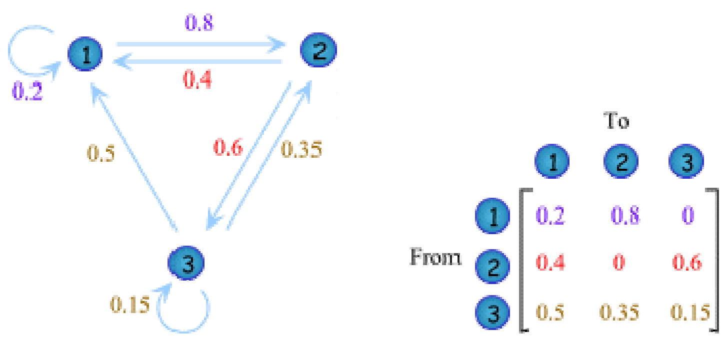
$p_{ij}^{(n)} \geq 0$ za svako i i j i za svako $n = 0,1,2,..,$

$\sum_j p_{ij}^{(n)} = 1$ za svako i i za svako $n = 1,2,....$

Tranzitivne vjerovatnoće, odnosno vjerovatnoće prelaza možemo zadati u vidu konačne ili beskonačne kvadratne matrice sa nenegativnim elementima, koja se naziva matrica tranzitivnih vjerovatnoća ili *stohastička matrica*, u oznaci $P = (P_{ij}),\ i, j \in S$. Zbir elemenata u svakom redu matrice P jednak je 1, s obzirom da bilo koji red predstavlja raspodjelu vjerovatnoća slučajne promjenljive. Matricu P pišemo u obliku:

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} & \dots \\ P_{21} & P_{22} & P_{23} & \dots \\ P_{31} & P_{32} & P_{33} & \dots \\ . & . & . & \dots \\ . & . & . & \dots \\ . & . & . & \dots \end{bmatrix}$$

Sljedeći primjer ilustruje dijagram i matricu prelaza:



Nakon svega prethodno izloženog, moguće je definisati Markovljev lanac.

Za slučajni proces $\{X_t\},\ t = 0,1,2,...,$ *kaže se da je lanac Markova ako ispunjava sljedeća tri uslova:*

1. *svojstvo Markova,*
2. *stacionarne vjerovatnoće prelaza,*
3. *poznat je skup početnih vjerovatnoća* $P\{X_0 = i\}$ *za svako i.*

Markovljevi lanci mogu imati *ograničen* ili *neograničen broj stanja*. Za teoriju redova čekanja od posebnog su značaja Markovljevi lanci sa beskonačnim brojem stanja.

U slučaju vremenski diskretnog lanca Markova trenuci u kojima se mogu dešavati promjene stanja su predodređeni i označeni brojevima 0,1,2..., a proces može da se zadrži u datom stanju samo za vrijeme koje mora biti geometrijski raspodijeljeno; to je jedina diskretna funkcija gustine raspodjele koja je bez pamćenja. Ovo svojstvo, *odsustvo pamćenja*, zahtijeva se kod svih lanaca Markova i ono ograničava opštost procesa koji se njima mogu razmatrati.

3.3. Procesi rađanja i umiranja

Najšire korišćeni modeli u teoriji RČ zasnivaju se neposredno na *procesima rađanja i umiranja*. S obzirom na odnos eksponencijalne i Puasonove raspodjele (vidjeti poglavlje 1.2), za ove modele se obično kaže da imaju *Puasonov proces dolazaka* i *eksponencijalno vrijeme opsluživanja*. Razlike između modela potiču samo od toga kako se λ i μ mijenjaju sa brojem stanja *n* i da li je broj stanja ograničen ili ne.

Ti procesi spadaju u veoma važnu klasu procesa Markova.

U procesima rađanja i umiranja moguće je da se iz nekog stanja n pređe samo u susjedno stanje $n-1$ *ili* $n+1$. *Prelazi u druga stanja nisu dozvoljeni, što se matematički izražava pretpostavkom da su te vjerovatnoće beskonačno male veličine višeg reda u odnosu na vjerovatnoće prelaza u susjedno stanje i u odnosu na vjerovatnoće ostajanja u stanju.*

Nazivi procesa rađanja i umiranja potiče upravo iz realnih sistema redova čekanja.

Pojavljivanje novog klijenta može se uslovno nazvati rađanjem jer povećava broj klijenata u sistemu za jedan. Napuštanje opsluženog klijenta znači smanjenje broja klijenata za jedan i može se uslovno nazvati umiranjem.

Stanje sistema u trenutku $t(t \geq 0)$, određeno je brojem klijenata $N(t)$ u sistemu opsluživanja u trenutku *t*. Proces rađanja i umiranja daje vjerovatnosni (probabilistički) opis promjene $N(t)$ u vremenu.

Pojedinačna rađanja i umiranja se dešavaju slučajno, a njihovi intenziteti (*učestanosti događanja*) *zavise od stanja sistema.*

Pretpostavke da posmatrani proces predstavlja proces rađanja i umiranja su sljedeće:

1. Za dato $N(t) = n$, raspodjela vjerovatnoće vremena koje će sistem provesti u stanju *n* dok se ne desi sljedeće rađanje (dolazak) je eksponencijalna sa parametrom $\lambda_n, n = 0,1,2,...$

2. Za dato $N(t) = n$, raspodjela vjerovatnoća vremena koje će sistem provesti u stanju n dok se ne desi sljedeće umiranje (odlazak) je eksponencijalna sa parametrom $\mu_n, n = 0,1,2,...$

3. Može se desiti samo jedno rađanje ili jedno umiranje u jednom trenutku.

Proces rađanja i umiranja zbog pretpostavki 1 *i* 2 *predstavlja poseban tip vremenski neprekidnog lanca Markova.*

Pretpostavka 3 značajno olakšava matematičku analizu procesa.

Analiza procesa rađanja i umiranja u prelaznom režimu, tj. određivanje vjerovatnoća $P_n(t)$ da će se u sistemu opsluživanja u trenutku t nalaziti tačno n klijenata, predstavlja izuzetno težak zadatak izuzevši u nekoliko posebnih slučajeva. Relativno je lako izračunati raspodjelu vjerovatnoća P_n u uslovima ravnoteže, odnosno u stacionarnom ili ustaljenom stanju ($P_n = \lim_{t\to\infty} P_n(t)$), ako takvo stanje postoji.

Osnovni princip za analizu procesa rađanja i umiranja sastoji se u sljedećoj činjenici:

U stacionarnom stanju očekivani broj ulazaka procesa u stanje n u jedinici vremena jednak je očekivanom intenzitetu toka izlazaka iz stanja n.

Kako je naprijed rečeno, proces može da uđe u stanje n ako se nalazi u stanju $n-1$ ili $n+1$. Iz stanja $n-1$ ući će u stanje n ako se desi rađanje. Iz stanja $n+1$ ući će u stanje n ako se desi umiranje. Prema tome, očekivani broj ulazaka u stanje n u jedinici vremena je

$$\lambda_{n-1} P_{n-1} + \mu_{n+1} P_{n+1}.$$

Slično, proces može da izađe iz stanja n u stanje $n+1$ ako se desi rađanje ili u stanje $n-1$ ako se desi umiranje. Prema tome, očekivani broj izlazaka iz stanja n u jedinici vremena je

$$\lambda_n P_n + \mu_n P_n.$$

U stacionarnom (ustaljenom) stanju očekivani intenziteti ulazaka i izlazaka su jednaki, tj. na osnovu prethodna dva izraza, mora biti

$$\lambda_{n-1} P_{n-1} + \mu_{n+1} P_{n+1} = \lambda_{n-1} P_{n-1} + \mu_{n+1} P_{n+1}, \; n = 0,1,2,.... \qquad (3.1)$$

Kako sistem mora da bude u jednom od stanja, to mora biti zadovoljen normirajući uslov

$$\sum_{n=0}^{\infty} P_n = 1.$$

Iz jednačina ravnoteže (3.1) dobija se sljedeći sistem jednačina ravnoteže:

$$\begin{aligned} \mu_1 P_1 &= \lambda_0 P_0 \\ \lambda_0 P_0 + \mu_2 P_2 &= (\lambda_1 + \mu_1) P_1 \\ \lambda_1 P_1 + \mu_3 P_3 &= (\lambda_2 + \mu_2) P_2 \\ &\vdots \\ \lambda_{n-2} P_{n-2} + \mu_n P_n &= (\lambda_{n-1} + \mu_{n-1}) P_{n-1} \\ \lambda_{n-1} P_{n-1} + \mu_{n+1} P_{n+1} &= (\lambda_n + \mu_n) P_{n1} \\ &\vdots \end{aligned} \tag{3.2}$$

Da bismo riješili gornji sistem postupamo na sljedeći način. Iz prve jednačine se vjerovatnoća P_1 izrazi preko P_0, zatim se tako dobijena vrijednost od P_1 zamijeni u drugu jednačinu, a potom se P_2 izrazi preko P_0 itd. Tako se redom nakon sređivanja dobijaju rješenja

$$\begin{aligned} P_1 &= \frac{\lambda_0}{\mu_1} P_0 \\ P_2 &= \frac{\lambda_1 \lambda_0}{\mu_2 \mu_1} P_0 \\ P_3 &= \frac{\lambda_2 \lambda_1 \lambda_0}{\mu_3 \mu_2 \mu_1} P_0 \\ &\vdots \\ P_n &= \frac{\lambda_{n-1} \lambda_{n-2} \cdots \lambda_0}{\mu_n \mu_{n-1} \cdots \mu_1} P_0 \\ P_{n+1} &= \frac{\lambda_n \lambda_{n-1} \cdots \lambda_0}{\mu_{n+1} \mu_n \cdots \mu_1} P_0 . \end{aligned} \tag{3.3}$$

Radi kraćeg zapisa gornjih formula, uvedimo oznaku koju ćemo ubuduće koristiti.

$$C_n = \frac{\lambda_{n-1} \lambda_{n-2} \cdots \lambda_0}{\mu_n \mu_{n-1} \cdots \mu_1}, \quad n = 1,2,... \tag{3.4}$$

U skladu sa gornjom oznakom rješenja sistema (3.2) data formulama (3.3) pišemo kao

$$P_n = C_n P_0, \quad n = 1,2,... \tag{3.5}$$

Otuda i iz normirajućeg uslova

$$\sum_{n=0}^{\infty} P_n = 1$$

dobijamo

$$\left(1+\sum_{n=1}^{\infty} C_n\right) P_0 = 1,$$

odakle slijedi

$$P_0 = \frac{1}{1+\sum_{n=0}^{\infty} C_n}. \tag{3.6}$$

Kada su poznate vjerovatnoće stanja P_n, $n = 0,1,2,...,$ možemo izračunati karakteristike sistema (vidjeti M. Vujošević [3] i S. Vukadinović [5]), i to:

1) *očekivani broj klijenata u sistemu L*:

$$L = \sum_{n=0}^{\infty} nP_n; \tag{3.7}$$

2) *ako se radi o sistemu sa s kanala opsluživanja, očekivani broj klijenata u redu čekanja L_q je*

$$L_q = \sum_{n=s}^{\infty} (n-s)P_n;$$

nadalje, na osnovu *Litlove formule* (vidjeti poglavlje 2.3) možemo izračunati

3) *vrijeme koje klijent provede u sistemu usluživanja W i ono iznosi*

$$W = \frac{L}{\lambda_{sr}}; \tag{3.8}$$

4) *vrijeme koje klijent provede u redu čekanja W_q iznosi*

$$W_q = \frac{L_q}{\lambda_{sr}}, \tag{3.9}$$

pri čemu je λ_{sr} *prosječni intenzitet dolazaka u sistem opsluživanja koji se računa po formuli*

$$\lambda_{sr} = \sum_{n=s}^{\infty} \lambda_n P_n \ . \tag{3.10}$$

Izvedene formule (3.5)-(3.10) *predstavljaju osnovu za rješavanje velikog broja sistema redova čekanja koji se modeliraju kao procesi rađanja i umiranja, od kojih su neki opisani u narednim poglavljima.*

4. MODELI REDOVA ČEKANJA KOJI SE ODNOSE NA DISKRETNE PROCESE RAĐANJA I UMIRANJA

Najšire korišćeni modeli u teoriji redova čekanja (u skraćenoj oznaci RČ) zasnivaju se neposredno na procesima rađanja i umiranja. S obzirom na odnos eksponencijalne i Puasonove raspodjele (vidjeti poglavlje 1.2), za ove modele se obično kaže da imaju Puasonov proces dolazaka i eksponencijalno vrijeme usluživanja. Razlike između modela potiču samo od toga kako se λ_n i μ_n mijenjaju sa brojem stanja n i da li je broj stanja ograničen ili ne.

4.1. Model $M/M/s$

U skladu sa *Kendal-Lijevom notacijom* (vidjeti poglavlje 2.1) *model* $M/M/s$ *opisuje sistem u kome su*:

- *dolasci međusobno statistički nezavisni i to sa vremenom između dva dolaska koje je eksponencijalno raspodijeljeno, tj. tok dolazaka je Puasonov,*
- *vremena opsluživanja su takođe međusobno statistički nezavisna i podliježu nekoj drugoj eksponencijalnoj raspodjeli i*
- *broj kanala opsluživanja je pozitivan prirodan broj s.*

Ovaj model je poseban slučaj procesa rađanja i umiranja u kome su intenziteti dolazaka (rađanja) i intenziteti odlazaka (umiranja) za zauzeti kanal konstantni i ne zavise od stanja sistema.

Kada u sistemu postoji samo jedan kanal opsluživanja (s =1) slijedi da su parametri procesa rađanja i umiranja $\lambda_n = \lambda (n = 0,1,2,...)$ i $\mu_n = \mu (n = 1,2,3,...)$.

Kada sistem ima više kanala opsluživanja ($s > 1$), tada μ_n predstavlja intenzitet opsluživanja za cijeli sistem, odnosno intenzitet odlazaka (umiranja) za sistem kao cjelinu kada se u njemu nalazi n klijenata.

Kada je maksimalni intenzitet opsluživanja (sμ) veći od intenziteta dolazaka λ, tj. kada je ispunjen uslov

$$\rho = \frac{\lambda}{s\mu} < 1, \tag{4.1}$$

onda sistem ima ustaljeno stanje.

Navedeni uslov za ustaljeno stanje je razumljiv po sebi i iskazujemo ga kao:

srednji broj klijenata koji dospijevaju u sistem u jedinici vremena treba da bude manji od srednjeg broja klijenata koji mogu biti opsluženi u jedinici vremena. U suprotnom slučaju desiće se pojava gomilanja i red će u vremenu neograničeno rasti.

Rezultati za model *M/M/*1

U slučaju modela faktori C_n dati formulom (3.4) svode se na

$$C_n = (\lambda / \mu)^n = \rho^n, \quad \text{za } n = 1,2,3,....$$

Sada je

$$P_n = \rho^n P_0, \qquad \text{za } n = 1,2,3,....,$$

pri čemu je

$$P_0 = \frac{1}{1+\sum_{n=1}^{\infty}\rho^n}$$

$$= \left(\sum_{n=0}^{\infty}\rho^n\right)^{-1}$$

$$= \left(\frac{1}{1-\rho}\right)^{-1}$$

$$= 1-\rho .$$

Prema tome, $P_0 = 1-\rho$, što uvršteno u prethodnu formulu daje

$$P_n = \rho^n (1-\rho), \quad \text{za } n = 1,2,3,.... \tag{4.2}$$

Dalje imamo

$$L = \sum_{n=0}^{\infty} nP_n$$

$$= \sum_{n=0}^{\infty} n(1-\rho)\rho^n$$

$$= (1-\rho)\rho\sum_{n=0}^{\infty}\frac{d}{d\rho}(\rho^n),$$

odakle dalje slijedi

$$\begin{aligned} L &= (1-\rho)\,\rho\frac{d}{d\rho}\left(\sum_{n=0}^{\infty}\rho^{n}\right) \\ &= (1-\rho)\rho\frac{d}{d\rho}\left(\frac{1}{1-\rho}\right) \\ &= \frac{\rho}{1-\rho} \\ &= \frac{\lambda}{\mu-\lambda}\ . \end{aligned} \tag{4.3}$$

Na sličan način dobijamo

$$\begin{aligned} L_q &= \sum_{n=1}^{\infty}(n-1)P_n \\ &= L-(1-P_0) \\ &= \frac{\lambda^2}{\mu(\mu-\lambda)}\,. \end{aligned}$$

(4.4)

Koristeći Litlovu formulu neposredno dobijamo

$$W = \frac{1}{\mu-\lambda}\,. \tag{4.5}$$

Primjenjujući formule $L_q = \lambda W_q$ ili $W_q = W - 1/\mu$, dobijamo

$$W_q = \frac{\lambda}{\mu(\mu-\lambda)}\,. \tag{4.6}$$

Vjerovatnoća da klijent neće čekati u redu je

$$P\{w_q = 0\} = P_0 = 1-\rho. \tag{4.7}$$

Iz gornje formule zaključujemo da parametar $\rho\ (=1-P_0)$ *u ovom modelu predstavlja vjerovatnoću da je sistem opsluživanja zauzet i obično se naziva koeficijentom iskorišćenosti ili faktorom opsluživanja.*

4.2. Rezultati za model *M/M/s*

Kada je $s > 1$, faktori C_n dati formulom (3.4) postaju

$$C_n = \frac{(\lambda/\mu)^n}{n!}, \qquad \text{za } n = 1,2,3,\ldots,s,$$

$$C_n = \frac{(\lambda/\mu)}{s!}\left(\frac{\lambda}{s\mu}\right)^{n-s} = \frac{(\lambda/\mu)^n}{s!\,s^{n-s}}, \qquad \text{za } n = s,\ s+1,\ s+2,\ldots$$

Prema tome, ako je $\lambda < s\mu$, onda imamo

$$P_0 = \left[\sum_{n=0}^{s-1}\frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^s}{s!}\sum_{n=s}^{\infty}\left(\frac{\lambda}{s\mu}\right)^{n-s}\right]^{-1}$$

$$= \left[\sum_{n=0}^{s-1}\frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^s}{s!}\frac{1}{1-(\lambda/s\mu)}\right]^{-1} \tag{4.8}$$

$$P_n = \begin{cases} \dfrac{(\lambda/\mu)^n}{n!}P_0 \\ \\ \dfrac{(\lambda/\mu)^n}{s!\,s^{n-s}}P_0 \end{cases} \qquad \text{za } n = 0,1,2,\ldots,s \ \text{ i za } n = s+1,\ s+2,\ldots \tag{4.9}$$

Uvodeći oznaku $\rho = \lambda / s\mu$, dobijamo

$$L_q = \sum_{n=s}^{\infty}(n-s)P_n$$

$$= \sum_{j=0}^{\infty} jP_{s+j}$$

$$
\begin{aligned}
&= \sum_{j=0}^{\infty} j \frac{(\lambda/\mu)^s}{s!} \rho^j P_0 \\
&= P_0 \frac{(\lambda/\mu)^s}{s!} \rho \sum_{j=0}^{\infty} \frac{d}{d\rho}\left(\rho^j\right) \\
&= P_0 \frac{(\lambda/\mu)^s}{s!} \rho \frac{d}{d\rho}\left(\sum_{j=0}^{\infty} \rho^j\right) \\
&= P_0 \frac{(\lambda/\mu)^s}{s!} \rho \frac{d}{d\rho}\left(\frac{1}{1-\rho}\right) \\
&= \frac{(\lambda/\mu)^s \rho}{s!(1-\rho)^2} P_0
\end{aligned}
$$

(4.10)

$$W_q = L_q / \lambda \tag{4.11}$$

$$W = W_q + 1/\mu \tag{4.12}$$

$$L = \lambda\left(W_q + 1/\mu\right) = L_q + \frac{\lambda}{\mu} \tag{4.13}$$

Vjerovatnoća da klijent neće čekati na uslugu jednaka je vjerovatnoći da se sistem nalazi u nekom od stanja $n < s$, pa prema tome ona iznosi iznosi

$$P\{w_q = 0\} = \sum_{n=0}^{s-1} P_n \tag{4.14}$$

Primjer 1. Utvrđeno je da radnim danima čitaoci dolaze u biblioteku po knjige slučajno i da je vrijeme između dva dolaska eksponencijalno raspodijeljeno sa srednjom vrijednošću 20 min. U biblioteci radi jedan službenik. Pretpostavlja se da je vrijeme usluge eksponencijalno raspodijeljeno i da prosječno traje 12 min. Subotom se broj dolazaka u biblioteku udvostručava ali tada rade dva službenika. Uporediti performanse ova dva sistema opsluživanja.

Rješenje: Na osnovu opisa problema zaključuje se da za radne dane odgovara model sistema $M/M/1$ sa parametrima $\lambda = 3$ dolaska/sat i $\mu = 5$ usluga/sat. Korišćenjem obrazaca (4.2)-(4.7) neposredno određujemo sljedeće performanse:

$$\rho = \lambda / \mu = 3/5 = 0{,}6$$

$$P_0 = 1 - \rho = 0{,}4$$

$$L = \lambda / (\mu - \lambda) = 1,5$$

$$L_q = L - (1 - P_0) = 0,9$$

$$W = 1/(\mu - \lambda) = 1/2 = 30 \text{ minuta}$$

$$W_q = \lambda / \mu(\mu - \lambda) = 0,3 = 18 \text{ minuta.}$$

Tumačenje ovih podataka sastoji se u sljedećem. Zauzetost službenika pružanjem usluga čitaocima je 60% ukupnog vremena (ρ = 0,6); vjerovatnoća da će čitalac biti uslužen bez čekanja u redu je $P_0 = 0,4$; prosječan broj čitalaca u biblioteci je L = 1,5, a u redu čekanja L_q = 0,9; prosječno zadržavanje čitaoca u biblioteci je W = 30min, od čega u redu čekanja W_q =18 min.

Za subotu odgovara model $M/M/2$ sa parametrima λ = 6 dolazaka/sat i μ = 5 usluga/sat. Korišćenjem obrazaca (4.8)-(4.13) dobijamo sljedeće performanse sistema.

$$\lambda / \mu = 6/5 = 1,2$$

$$\rho = \lambda / 2\mu = 6/10 = 0,6$$

$$P_0 = \left[\sum_{n=9}^{1} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^2}{2!} \frac{1}{1-(\lambda/s\mu)^n} \right]^{-1} = 0,25$$

$$L_q = \frac{(\lambda/\mu)^2 \rho}{2/(1-\rho)^2} P_0 = 0,675$$

$$L = L_q + \lambda / \mu = 1,875$$

$$W_q = L_q / \lambda = 6,75 \text{ minuta}$$

$$W = W_q + 1/\mu = 6,75 + 12 = 18,75 .$$

Vjerovatnoće stanja sistema za ova dva sistema prikazane su u sljedećoj tabeli.

Tabela 4.1.

	P_0	P_1	P_2	P_3	P_4	P_5	P_6
$M/M/1$	0,4	0,24	0,144	0,0864	0,05184	0,03110	0,01866

M/*M*/2	0,25	0,30	0,18	0,108	0,0648	0,0388	0,02333

Zanimljivo je uočiti da ova dva sistema imaju isti koeficijent ρ = 0,6, odnosno u oba slučaja 60% ukupnog vremena sistemi su zauzeti davanjem usluge, ali se pri tome u njima postižu različiti nivoi kvaliteta usluge. Naime, u sistemu *M*/*M*/2 mušterija prosječno provede manje vremena u čekanju (6,75 min. u odnosu na 18 minuta, a vjerovatnoća da će biti uslužen bez čekanja je veća (iznosi 0,55, što je veće od 0,4).

4.3. Modeli sistema ograničenog kapaciteta *M*/*M*/*s*/*K*

U skladu sa *Kendal-Lijevom notacijom* (vidjeti poglavlje 2.1), *model M*/*M*/*s*/*K opisuje sistem u kome su*:

- dolasci međusobno statistički nezavisni i to sa vremenom između dva dolaska koje je eksponencijalno raspodijeljeno, tj. tok dolazaka je Puasonov,

- vremena opsluživanja su takođe međusobno statistički nezavisna i podliježu nekoj drugoj eksponencijalnoj raspodjeli,

- broj kanala opsluživanja je pozitivan prirodan broj s i

- sistem opsluživanja je ograničenog kapaciteta koji iznosi K.

Ovi modeli se razlikuju od prethodnih po tome što se pretpostavlja da je sistem opsluživanja ograničenog kapaciteta i iznosi *K*. To znači da je ukupan broj klijenata koji mogu da se nađu u sistemu, odnosno u redu čekanja i na opsluživanju, jednak *K*. Veliki broj realnih sistema ima upravo ovakvo ograničenje. U ovakvim sistemima važi

$$\lambda_n = \begin{cases} \lambda, n = 0,1,2,..,K-1 \\ 0, n \geq K. \end{cases}$$

Pošto je $\lambda_n = 0$ za $n \geq K$, ovaj sistem dospijeva u ustaljeno stanje.

Rezultati za jednokanalni sistem *M/M/*1*/K*

U slučaju sistema *M/M/*1*/K,* na osnovu prethodnih formula, relacije (3.4) za C_n postaju

$$C_n = \begin{cases} \left(\dfrac{\lambda}{\mu}\right)^n = \rho^n, n = 1,2,...,K \\ 0, n > K \end{cases}$$

Otuda za $\rho \neq 1$, dobijamo

$$P_0 = \frac{1}{\sum_{n=0}^{K} (\lambda/\mu)^n}$$

$$= 1 / \left[\frac{1-(\lambda/\mu)^{K+1}}{1-(\lambda/\mu)} \right]$$

$$= \frac{1-\rho}{1-\rho^{K+1}}, \tag{4.15}$$

odakle slijedi

$$P_n = \frac{1-\rho}{1-\rho^{K+1}} \rho^n, \qquad \text{za } n = 0,1,2,...,K. \tag{4.16}$$

Otuda dobijamo

$$L = \sum_{n=0}^{K} n P_n$$

$$= \frac{1-\rho}{1-\rho^{K+1}} \rho \sum_{n=0}^{K} \frac{d}{d\rho} (\rho^n)$$

$$= \frac{1-\rho}{1-\rho^{K+1}} \rho \frac{d}{d\rho} \left(\sum_{n=0}^{K} \rho^n \right)$$

$$= \frac{1-\rho}{1-\rho^{K+1}} \rho \frac{d}{d\rho} \left(\frac{1-\rho^{K+1}}{1-\rho} \right)$$

$$= \rho \frac{-(K+1)\rho^K + K\rho^{K+1} + 1}{(1-\rho^{K+1})(1-\rho)}$$

$$= \frac{\rho}{1-\rho} - \frac{(K+1)\rho^{K+1}}{(1-\rho^{K+1})}. \tag{4.17}$$

Za jednokanalne sisteme je

$$L_q = L - (1 - P_0)..$$

Navedeni rezultati izvedeni su bez uslova $\lambda < \mu$. Ako je $\lambda = \mu$, tada je

$$\rho_n = \frac{1}{K+1}, \qquad \text{za } n = 0,1,2,...,K, \tag{4.18}$$

tako da je $L = \dfrac{K}{2}$.

Za računanje očekivanih čekanja koriste se Litlove formule koje u ovom modelu imaju oblik

$$W = \frac{L}{\lambda_{sr}} \tag{4.19}$$

$$W_q = \frac{L_q}{\lambda_{sr}}, \tag{4.20}$$

gdje je

$$\begin{aligned} \lambda_{sr} &= \sum_{n=0}^{\infty} \lambda_n P_n \\ &= \sum_{n=0}^{K-1} \lambda_n P_n \\ &= \lambda(1 - P_K) \end{aligned} \tag{4.21}$$

Rezultati za višekanalni sistem *M/M/s/K*

Jasno je da je u sistemu *M/M/s/K* mora biti $s \le K$ jer kapacitet sistema ne može biti manji od broja kanala opsluživanja. Prema tome, važi

$$C_n = \frac{(\lambda/\mu)^n}{n!} \qquad \text{za } n = 1,2,...,s$$

$$C_n = \frac{(\lambda/\mu)}{s!}\left(\frac{\lambda}{s\mu}\right)^{n-s} = \frac{(\lambda/\mu)^n}{s!s^{n-s}}, \quad \text{za } n = s, s+1, ..., K,$$

$$C_n = 0, \qquad \text{za } n > K,$$

odakle slijedi

$$P_n = \begin{cases} \dfrac{(\lambda/\mu)^n}{n!} P_0, & \text{za } n = 1,2,...,s \\ \dfrac{(\lambda/\mu)^s}{s!s^{n-s}} P_0, & \text{za } n = s, s+1, ...,K \\ 0 & \text{za } n > K, \end{cases}$$

gdje je

$$P_0 = \left[\sum_{n=0}^{s} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^s}{s!} \sum_{n=s+1}^{K} \left(\frac{\lambda}{s\mu} \right)^{n-s} \right]^{-1} \tag{4.22}$$

Nadalje se može se izvesti formula

$$L_q = \frac{(\lambda/\mu)^s \rho P_0}{s!(1-\rho)^2} \left[1 - \rho^{K-s} - (K-s)\rho^{K-s}(1-\rho) \right], \tag{4.23}$$

gdje je $\rho = \lambda / s\mu.$ Dalje slijedi

$$L = \sum_{n=0}^{s-1} nP_n + L_q + s\left(1 - \sum_{n=0}^{s-1} P_n \right). \tag{4.24}$$

4.4. Modeli sistema sa konačnom izvornom populacijom *M/M/s///N*

Potpuna Kendal-Lijeva notacija za ove sisteme je *M*/*M*/*s*/∞/FIFO/*N* (FIFO – redoslijed opsluživanja je određen *redoslijedom pristizanja u sistem* (First In First Out)). Ovaj redoslijed opsluživanja je najčešći u praksi (česta je i oznaka FCFS-First Come First Served), a skraćenica za iste je *M/M/s///N* jer se podrazumijeva da je *kapacitet sistema neograničen i pravilo opsluživanja „prvi stigao prvi opslužen"* kad god drugačije nije navedeno. Posljednja oznaka *N* predstavlja broj klijenata koji postoji u izvornoj populaciji. Svaki od klijenata dolazi u sistem opsluživanja nezavisno i to sa vremenima između dva dolaska koja podliježu eksponencijalnoj raspodjeli sa parametrom λ. Ako je sistem prazan, tj. stanje je $n = 0$, tada je intenzitet dolazaka klijenata u sistem jednak $N\lambda$; kada se u sistemu nalazi n klijenata, tj. stanje je n, u izvornoj populaciji ima N - n

klijenata čiji je intenzitet dolazaka (N- n)λ. Naravno, sistem ne može da se nađe u stanju $n > N$.

Ovakvi sistemi su karakteristični za održavanje relativno malog broja određene vrste mašina ili uređaja. Intenzitet zahtijeva za opravku mašina u određenom trenutku zavisi od toga koliko je njih ispravno

Rezultati za *M/M/*1*///N*

Za ovaj model je

$$\lambda_n = \begin{cases} (N-n)\lambda, n = 0,1,2,\ldots, N-1 \\ 0, n > N \end{cases}$$

$$\mu_n = \mu, \qquad \text{za } n = 1, 2, \ldots, N.$$

Faktori C_n dati formulama (3.4) sada iznose

$$C_n = N(N-1)\cdots(N-n+1)\left(\frac{\lambda}{\mu}\right)^n = \frac{N!}{(N-n)!}\left(\frac{\lambda}{\mu}\right)^n, \quad \text{za } n \le N$$

$$C_n = 0, \qquad \text{za } n > N.$$

Otuda se dobija

$$P_0 = \sum_{n=0}^{N}\left[\frac{N!}{(N-n)!}\left(\frac{\lambda}{\mu}\right)^n\right]^{-1} \tag{4.25}$$

$$P_n = \frac{N!}{(N-n)!}\left(\frac{\lambda}{\mu}\right)^n P_{0,} \qquad n = 1, 2, .., N \tag{4.26}$$

$$L_q = \sum_{n=0}^{N}(n-1)P_n \; ,$$

odakle nakon sređivanja slijedi

$$L_q = N - \frac{\lambda+\mu}{\lambda}(1-P_0) \tag{4.27}$$

$$L = \sum_{n=0}^{N} nP_n = L_q + (1-P_0)$$

$$= N - \frac{\mu}{\lambda}\left(1 - P_0\right). \tag{4.28}$$

Konačno imamo

$$W = \frac{L}{\lambda_{sr}},$$

$$W_q = \frac{L_q}{\lambda_{sr}}, \tag{4.29}$$

gdje je

$$\lambda_{sr} = \sum \lambda_n P_n = \sum_{n=0}^{N} (N-n)\lambda P_n = \lambda(N-L). \tag{4.30}$$

Rezultati za *M/M/s///N*

Za $N \geq s > 1$ koristeći formule (3.4) , dobija se

$$C_n = \begin{cases} \dfrac{N!}{(N-n)!\,n!}\left(\dfrac{\lambda}{\mu}\right)^n, & n = 1, 2, .., s \\ \dfrac{N!}{(N-n)!\,s!\,s^{n-s}}\left(\dfrac{\lambda}{\mu}\right)^n, & n = s+1, s+2, ..., N \\ 0, n > N & \end{cases}$$

$$P_n = \begin{cases} \dfrac{N!}{(N-n)!\,n!}\left(\dfrac{\lambda}{\mu}\right)^n P_0, & n = 1,2,...,s \\ \dfrac{N!}{(N-n)!\,s!\,s^{n-s}}\left(\dfrac{\lambda}{\mu}\right)^n P_0, & n = s+1, s+2, ..., N \\ 0, & n > N \end{cases}$$

$$P_0 = \left[\sum_{n=0}^{s-1} \frac{N!}{(N-n)!n!} \left(\frac{\lambda}{\mu} \right)^n + \sum_{n=s}^{N} \frac{N!}{(N-n)!s!s^{n-s}} \left(\frac{\lambda}{\mu} \right)^n \right]^{-1} \tag{4.31}$$

$$L_q = \sum_{n=s}^{N} (n-s) P_n \tag{4.32}$$

$$L = \sum_{n=0}^{s-1} n P_n + L_q + s\left(1 - \sum_{n=0}^{s-1} P_n \right), \tag{4.33}$$

a W i W_q se računaju kao i za model sistema *M/M/*1*///N.*

Interesantno je napomenuti da navedene formule za P_0, P_n, L_q, L, W_q i W važe i za modele *G/M/s///N* pod uslovom da su raspodjele vremena između dva dolaska klijenta identične i imaju konstantno očekivanje $1/\lambda$.