Svojstvene vrijednosti prijelaznih matrica Markovljevih lanaca i primjene

Batković, Anna-Lena

SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO–MATEMATIČKI FAKULTET
MATEMATIČKI ODSJEK

Anna-Lena Batković

SVOJSTVENE VRIJEDNOSTI PRIJELAZNIH MATRICA MARKOVLJEVIH LANACA I PRIMJENE

Diplomski rad

Voditelj rada:
prof. dr. sc. Zoran Vondraček

Zagreb, srpanj 2023.

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

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

Povjerenstvo je rad ocijenilo ocjenom ________________.

Potpisi članova povjerenstva:

1. ____________________________
2. ____________________________
3. ____________________________

Sadržaj

Sadržaj **iii**

Uvod **1**

1 Osnovni pojmovi o Markovljevim lancima i spektru linearnih operatora **3**
1.1 Linearni operatori i spektar 3
1.2 Markovljevi lanci 7

2 Vrijeme miješanja Markovljevog lanca **13**
2.1 Udaljenost potpune varijacije 13
2.2 Vrijeme miješanja 15

3 Miješanje karata **19**
3.1 Slučajna šetnja na grupi permutacija 19
3.2 Miješanje karata 20

4 Svojstvene vrijednosti prijelazne matrice Markovljevog lanca **25**
4.1 Spektralna reprezentacija prijelazne matrice 25
4.2 Ograde na vrijeme miješanja 29

5 Ocjene uz granično ponašanje lanaca **33**

Bibliografija **39**

Uvod

Teorija vjerojatnosti dala je mnoštvo vrijednih rezulata, ali pod pretpostavkom da su promatrani nizovi slučajnih varijabli nezavisni. Nezavisnost je ipak snažna pretpostavka i primjene nekad zahtijevaju zavisnost u promatranim procesima. Markovljevi lanci primjer su jednog jednostavnog modela slučajnog procesa koji dozvoljava zavisnost među uzastopnim varijablama u nizu. Markovljeve lance karakterizira svojstvo da je neposredna budućnost nezavisna od prošlosti uz danu sadašnjost. Takav model našao je svoje mjesto u mnogim primjenama. Matematičari Persi Diaconis, David Aldous i Dave Beyer modelirali su miješanje karata Markovljevim lancem i došli do optimalnog broja miješanja za dobro promiješan špil. Taj primjer je tema Poglavlja 3. S druge strane, Markovljevi lanci su temelj cijele jedne klase metoda simulacija – *Monte Carlo Markov Chain* (MCMC). U Poglavlju 5 govorit ćemo više o poopćenju jakog zakona velikih brojeva na Markovljeve lance i teoriji koja opravdava valjanost uzoraka dobivenih MCMC metodama.

No, da bismo razumjeli spomenute primjere, kroz Poglavlja 2 i 4 analizirat ćemo granično ponašanje Markovljevih lanaca. Naime, uz posebne pretpostavke na sam lanac i nakon dovoljno vremena, vjerojatnosna svojstva Markovljevog lanca postaju stabilna i lanac počinje otprilike pratiti istu distribuciju u svakom koraku, što omogućava točnije predikcije o njegovom ponašanju kada pustimo da vrijeme teče. Trenutak kad lanac se lanac prvi put nađe blizu stacionarnosti zvat ćemo vremenom miješanja. Svojstvene vrijednosti prijelazne matrice lanca odredit će gornju i donju ogradu na vrijeme miješanja, odnosno pokazat ćemo da imaju neposredan utjecaj na brzinu konvergencije lanca prema stacionarnosti.

Poglavlje 1

Osnovni pojmovi o Markovljevim lancima i spektru linearnih operatora

U ovom poglavlju dat ćemo kratak pregled definicija i tvrdnji o spektrima linearnih operatora prema [2] i [3], gdje se mogu naći i dokazi koji će ovdje biti izostavljeni. Također ćemo navesti definiciju i neka svojstva Markovljevih lanaca koja su potrebna za razumijevanje daljnjeg sadržaja u radu. Iako su tvrdnje u ovom poglavlju fundamentalne za teoriju Markovljevih lanaca, njihovi dokazi izlaze van okvira ovog rada, ali se mogu pronaći u [8]; osim potpoglavlja *Vrijeme zaustavljanja i jako stacionarno vrijeme*, koje je prezentirano prema [5].

1.1 Linearni operatori i spektar

Neka su V i W vektorski prostori nad poljem $\mathbb{F}$ ($\mathbb{R}$ ili $\mathbb{C}$) takvi da $\dim V = n$ i $\dim W = m$.

Unitarni i normirani prostori

Definicija 1.1.1. ***Norma*** *na vektorskom prostoru V nad poljem $\mathbb{F}$ je preslikavanje $\|\cdot\| : V \to \mathbb{R}$ za koju vrijede sljedeća svojstva:*

a) $\|x\| \geq 0, \forall x \in V$*;*

b) $\|x\| = 0 \iff x = 0$*;*

c) $\|\alpha x\| = |\alpha| \, \|x\|, \forall \alpha \in \mathbb{F}, \forall x \in V$*;*

d) $\|x+y\| \leq \|x\| + \|y\|, \forall x, y \in V$.

Normiran prostor *je uređeni par* $(V, \|\cdot\|)$.

Definicija 1.1.2. ***Skalarni produkt*** *na vektorskom prostoru* V *nad poljem* $\mathbb{F}$ *je preslikavanje* $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$ *sa svojstvima:*

a) $\langle x, x \rangle \geq 0, \forall x \in V$;

b) $\langle x, x \rangle = 0 \iff x = 0$;

c) $\langle \alpha x, y \rangle, \forall \alpha \in \mathbb{F}, \forall x, y \in V$;

d) $\langle x_1 + x_2, x \rangle = \langle x_1, y \rangle + \langle x_2, y \rangle, \forall x_1, x_2, y \in V$;

e) $\langle x, y \rangle = \overline{\langle y, x \rangle}, \forall x, y \in V$.

Unitaran prostor *je uređeni par* $(V, \langle \cdot, \cdot \rangle)$.

U svakom unitarnom prostoru vrijedi *Cauchy-Schwarzova nejednakost*:

$$|\langle x, y \rangle|^2 \leq \langle x, x \rangle \langle y, y \rangle, \forall x, y \in V,$$

uz čiju pomoć se može pokazati da na svakom unitarnom prostoru postoji i norma definirana sa $\|x\| = \sqrt{\langle x, x \rangle}$. Obrat ne vrijedi, tj. proizvoljna norma ne mora biti inducirana nekim skalarnim produktom.
Od svih baza na unitarnom prostoru V, posebno će nas zanimati *ortonormirane*. Može se pokazati da svaki konačnodimenzionalan unitaran prostor ima ortonormiranu bazu. Naime, od svake baze prostora V može se dobiti ortonormirana baza pomoću Gram-Schmidtovog postupka dijagonalizacije (vidi [2]). Ortonormirane baze imaju puno korisnih svojstava koja "obične" baze nemaju. Ovdje navodimo definiciju i samo neka svojstva koja ćemo kasnije eksplicitno koristiti.

Definicija 1.1.3. *Neka je* $(V, \langle \cdot, \cdot \rangle)$ *unitaran prostor.*

i) Za vektore $x, y \in V$ *kažemo da su međusobno* ***ortogonalni*** *ako vrijedi* $\langle x, y \rangle = 0$.

ii) Za bazu $\{e_1, e_2, \dots, e_n\}$ *vektorskog prostora kažemo da je* ***ortonormirana baza*** *ako su elementi baze u parovima ortogonalni, te vrijedi* $\|e_i\| = 1, 1 \leq i \leq n$. *Ortonormiranu bazu kraće označavamo sa ONB.*

Propozicija 1.1.4. *U zapisu vektora* $x \in V$ *u ONB* $\{e_1, e_2, \dots, e_n\}$, $x = \sum_{i=1}^{n} \alpha_i e_i$, *koeficijenti* α_i *dani su sa:*

$$\alpha_i = \langle x, e_i \rangle, \forall i = 1, \dots, n.$$

Propozicija 1.1.5. *Neka je* $\{e_1, e_2, \dots, e_n\}$ *ortonormiran skup u unitarnom prostoru* $(V, \langle\cdot,\cdot\rangle)$. *Ekvivalentno je:*

i) $\{e_1, e_2, \dots, e_n\}$ *je ortonormirana baza;*

ii) vrijedi Parsevalova jednakost:

$$\|x\|^2 = \sum_{i=1}^{n} |\langle x, e_i\rangle|^2, \forall x \in V.$$

Linearni operatori i spektar

Definicija 1.1.6. *Linearan operator sa* V *u* W *je funkcija* $A\colon V \to W$ *takva da vrijedi* $A(\alpha x + \beta y) = \alpha A(x) + \beta A(y), \forall \alpha, \beta \in \mathbb{F}, \forall x, y \in V$. *Skup svih linearnih operatora označavamo sa* $L(V, W)$.

Skup $L(V, W)$ je također vektorski prostor nad poljem $\mathbb{F}$. Prostor $L(V, W)$ izomorfan je prostoru $M_{mn}(\mathbb{F})$, prostoru svih matrica dimenzije $m \times n$, pa zato svaki operator ima svoju matričnu reprezentaciju (koja ovisi o fiksiranim bazama u V i W). Zato sve tvrdnje iskazane za operatore vrijede i za matrice, i obratno. Posebno, kad je $W = V$, pišemo $L(V) = L(V, V)$.

Za linearni operator $A \in L(V, W)$ definiramo i *jezgru* $\operatorname{Ker} A \coloneqq \{x \in V : Ax = 0\}$ i *sliku* $\operatorname{Im} A \coloneqq \{Ax \in W : x \in V\}$. Jezgra i slika operatora su potprostori od V i W redom. Dodatno, definiramo *defekt* $d(A) \coloneqq \dim \operatorname{Ker} A$ i *rang* $r(A) \coloneqq \dim \operatorname{Im} A$.

Teorem 1.1.7 (Teorem o rangu i defektu)**.** *Neka je* $A \in L(V, W)$ *i* $\dim V < \infty$. *Tada je* $d(A) + r(A) = \dim V$.

Definicija 1.1.8. ***Svojstvena vrijednost*** *linearnog operatora* $A \in L(V)$ *je broj* $\lambda \in \mathbb{F}$ *za koji postoji netrivijalan* $x \in V$ *takav da vrijedi* $Ax = \lambda x$. *Vektor* x *onda nazivamo* ***svojstvenim vektorom*** *uz svojstvenu vrijednost* λ. *Skup svih svojstvenih vrijednosti operatora* A *zove se* ***spektar*** *operatora* A, *u oznaci* $\sigma(A)$.

Svi svojstveni vektori jedne svojstvene vrijednosti λ_0 čine *svojstveni potprostor* $V_A(\lambda_0) = \{x \in V : Ax = \lambda_0 x\}$. Dimenziju svojstvenog potprostora $V_A(\lambda_0)$ nazivamo *geometrijskom kratnošću* svojstvene vrijednosti λ_0.

Svojstvene vrijednosti se pronalaze kao nultočke *svojstvenog polinoma* $k_A(\lambda) = \det(A - \lambda I)$, gdje je A matrični prikaz operatora A u bilo kojoj bazi. Svojstveni polinom ne ovisi o izboru baze u matričnom prikazu jer su matrice nastale kao prikazi istog operatora u različitim bazama slične, pa imaju iste determinante. Kratnost nultočke λ_0 u svojstvenom polinomu nazivamo *algebarskom kratnošću* svojstvene vrijednosti λ_0.

Teorem 1.1.9. *Neka je $A \in L(V)$ linearan operator, $\sigma(A) = \{\lambda_1, \ldots, \lambda_k\}$, te $e^{(i)} = \{e_1, \ldots, e_{d_i}\}$ baza za svojstveni potprostor $V_A(\lambda_i)$ dimenzije d_i, $\forall i = 1, \ldots, k$. Unija $\cup_{i=1}^k e^{(i)}$ je linearno nezavisan skup u V.*

Iz Teorema 1.1.9 slijedi da se matrica može dijagonalizirati ako i samo ako za svaku svojstvenu vrijednost vrijedi da su njena algebarska i geometrijska kratnost međusobno jednake. U tom slučaju, baza u kojoj se matrica dijagonalizira je skup svojstvenih vektora.

Kako se svojstvene vrijednosti nalaze kao nultočke svojstvenog polinoma, važno je polje $\mathbb{F}$ nad kojim promatramo V, pa posljedično i $k_A(\lambda)$. Kad je $\mathbb{F} = \mathbb{C}$, neupitno je postojanje n svojstvenih vrijednosti jer sve nultočke polinoma sadržane su u $\mathbb{C}$, no kad je $\mathbb{F} = \mathbb{R}$, to ne mora biti slučaj, tj. nultočke polinoma ne moraju nužno biti sadržane u $\mathbb{R}$. Općenito, matrica nad poljem $\mathbb{R}$ ne mora imati svojstvenih vrijednosti i ne mora biti dijagonalizibilna. U nastavku ćemo opisati jednu posebnu vrstu matrica za koje su sve svojstvene vrijednosti realne i matrica se može dijagonalizirati.

Može se pokazati da za $A \in L(V)$, gdje je V unitaran prostor, postoji jedinstveni $A^* \in L(V)$ takav da za svaki $x, y \in V$ vrijedi $\langle Ax, y\rangle = \langle x, A^*y\rangle$.

Definicija 1.1.10. *Neka je V unitaran prostor i $A \in L(V)$.*

i) Operator A^* za koji vrijedi $\langle Ax, y\rangle = \langle x, A^*y\rangle, \forall x, y \in V$ zove se ***hermitski adjungiran operator*** *operatoru A.*

ii) Za operator A kažemo da je ***hermitski*** *ako je $A = A^*$.*

Teorem 1.1.11. *Neka je V konačnodimenzionalan unitaran prostor nad $\mathbb{R}$ i $A \in L(V)$ hermitski operator. Tada je spektar operatora A neprazan i sve svojstvene vrijednosti su realne.*

Teorem 1.1.12. *Neka je V konačnodimenzionalan unitaran prostor i $A \in L(V)$ hermitski operator. Postoji ONB $e = \{e_1, \ldots, e_n\}$ za V u kojoj je matrični zapis operatora dijagonalna matrica.*

Na kraju, napomenimo da ćemo u ovom radu promatrati prostor $V = \mathbb{R}^n$ kao unitaran prostor sa standardnim skalarnim produktom $\langle\cdot,\cdot\rangle\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ danim sa:

$$\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n)\rangle = \sum_{i=1}^n x_i y_i.$$

Taj skalarni produkt inducira 2-normu na $\mathbb{R}^n$ danu sa:

$$\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}.$$

1.2 Markovljevi lanci

Definicija Markovljevog lanca i osnovna svojstva

Definicija 1.2.1. *Neka je $(\mathcal{X}, \mathcal{G})$ izmjeriv prostor.* ***Slučajan proces*** *s diskretnim vremenom na prostoru stanja $\mathcal{X}$ je niz slučajnih varijabli $(X_n)_{n\in\mathbb{N}_0}$, gdje su $X_n\colon \Omega \to \mathcal{X}$ slučajne varijable na nekom vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$.*

Definicija 1.2.2. *Neka je $\mathcal{X}$ prebrojiv skup. Slučajni proces $(X_n)_{n\in\mathbb{N}_0}$ je* ***Markovljev lanac*** *ako za njega vrijedi:*

$$\mathbb{P}(X_{n+1} = y | X_n = x, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = \mathbb{P}(X_{n+1} = y | X_n = x), \tag{1.1}$$

za sve $n \geq 0$ i za sve $x, y, x_0, \dots, x_{n-1} \in \mathcal{X}$ za koje su uvjetne vjerojatnosti dobro definirane.

Uvjet u (1.1) zove se *Markovljevo svojstvo* i govori da ponašanje lanca u neposrednoj budućnosti ovisi samo o sadašnjosti, a ne ovisi o prošlosti.
Stohastička matrica je matrica $P = [P(x, y)]_{x,y\in\mathcal{X}}$ za koju vrijedi da je $P(x, y) \geq 0$, $\forall x, y \in \mathcal{X}$ te $\sum_{y\in\mathcal{X}} P(x, y) = 1$, $\forall i \in \mathcal{X}$. Pod pojmom distribucije u ovom radu smatrat ćemo niz nenegativnih brojeva $\mu = (\mu(x) : x \in \mathcal{X})$ takav da $\sum_{x\in\mathcal{X}} \mu(x) = 1$. Takve matrice i vektori važni su jer sadrže prijelazne i početne vjerojatnosti za homogene lance koji su precizno opisani u sljedećoj definiciji.

Definicija 1.2.3. *Neka je $\mu = (\mu(x) : x \in \mathcal{X})$ distribucija na $\mathcal{X}$ te neka je $P = [P(x, y)]_{x,y\in\mathcal{X}}$ stohastička matrica. Slučajni proces $X = (X_n)_{n\in\mathbb{N}_0}$ je* ***homogen Markovljev lanac*** *s početnom distribucijom μ i prijelaznom matricom P, ili kraće (μ, P)-Markovljev lanac, ako vrijedi:*

a) $\mathbb{P}(X_0 = x) = \mu(x), \forall x \in \mathcal{X}$

b) $\forall x_0, \dots, x_{n-1}, x, y \in \mathcal{X}, \forall n \geq 0$ *vrijedi:*

$$\mathbb{P}(X_{n+1} = y | X_n = x, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = P(x, y).$$

Za stanje $x \in \mathcal{X}$ takvo da $\mu(x) > 0$ možemo definirati uvjetnu vjerojatnost sa $\mathbb{P}_x(A) := \mathbb{P}(A | X_0 = x), A \in \mathcal{F}$. Isto tako ćemo očekivanje uz vjerojatnost $\mathbb{P}_i$ označavati sa $\mathbb{E}_x$. Ako je μ koncentrirana u jednom stanju, tj. $\mu(x) = 1$ za neki $x \in \mathcal{X}$, onda je $\mathbb{P}(A) = \mathbb{P}_x(A)$. Takvu distribuciju koncentriranu u stanju x označavamo sa $\delta^x := (\delta_{xy} : x, y \in \mathcal{X})$.
U nastavku govorimo samo o homogenim lancima, iako to nećemo eksplicitno navoditi. Nije odmah jasno iz definicije da homogeni Markovljevi lanci zadovoljavaju Markovljevo svojstvo (1.1), ali to je jednostavna posljedica sljedećeg teorema.

Teorem 1.2.4. *Neka je X (μ, P)-Markovljev lanac. Tada za sve $n \geq 0$ i sva stanja $i_0, \ldots, i_n$ vrijedi:*

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \mu_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}. \tag{1.2}$$

Vrijedi i obrat: ako je X slučajni proces s konačnodimenzionalnim distribucijama (1.2)*, onda je X (μ, P)-Markovljev lanac.*

Teorem 1.2.5. *Neka je X (μ, P)-Markovljev lanac. Uvjetno na $X_m = x$, slučajni proces $(X_{m+n})_{n \in \mathbb{N}_0}$ je (δ^x, P)-Markovljev lanac koji je nezavisan od slučajnih varijabli $X_0, \ldots, X_m$.*

Ako lanac kreće iz stanja x, vjerojatnost da se nakon n koraka nađe u stanju y dana je sa $\mathbb{P}_x(X_n = y) = (P^n)(x, y)$, što je također posljedica Teorema 1.2.5. Brojevi $P^n(x, y)$ nazivaju se n-koračne prijelazne vjerojatnosti. Ako je početna distribucija μ, onda je $\mathbb{P}(X_n = y) = (\mu P^n)(x, y)$.

Analiza prostora stanja

Da bismo mogli reći nešto o dugoročnom ponašanju lanca, važno je razumjeti kakve su mogućnosti prijelaza između dostupnih stanja. Zato definiramo neka važna svojstva koja lanac može imati, ireducibilnost i aperiodičnost, koja ovise o prijelaznim vjerojatnostima među stanjima u $\mathcal{X}$, a bit će od važnosti kad budemo govorili o graničnom ponašanju lanca.

Prvo vrijeme pogađanja skupa $B \subset \mathcal{X}$ je slučajna varijabla:

$$T_B \coloneqq \min\{n \geq 0 : X_n \in B\},$$

uz konvenciju $\min \varnothing = \infty$. Kad je B jednočlan, npr. $B = \{x\}$, pišemo T_x umjesto $T_{\{x\}}$.

Definicija 1.2.6. *Stanje $y \in \mathcal{X}$ je* ***dostižno*** *iz stanja $x \in \mathcal{X}$ ako je $\mathbb{P}_x(T_y < \infty) > 0$, u oznaci $x \longrightarrow y$. Ako je $x \longrightarrow y$ i $y \longrightarrow x$, onda kažemo da stanja x i y komuniciraju, u oznaci $x \longleftrightarrow y$. Za Markovljev lanac $(X_n)_{n \in \mathbb{N}_0}$ s prostorom stanja $\mathcal{X}$ kažemo da je* ***ireducibilan*** *ako svaka dva stanja u $\mathcal{X}$ komuniciraju.*

Dostižnost stanja y iz stanja x ekvivalentno je egzistenciji $t \in \mathbb{N}$ za koji je $P^t(x, y) > 0$. Prvo vrijeme povratka u stanje $x \in \mathcal{X}$ definiramo sa:

$$T_x \coloneqq \min\{n > 0 : X_n = x\}.$$

Definicija 1.2.7. *Stanje $x \in \mathcal{X}$ je* ***povratno*** *ako vrijedi $\mathbb{P}_x(T_x < \infty) = 1$, a* ***prolaznim*** *ga nazivamo ako je $\mathbb{P}_x(T_x < \infty) < 1$. Stanje $x \in \mathcal{X}$ je* ***pozitivno povratno*** *ako je $\mathbb{E}_x[T_x] < \infty$.*

Pozitivno povratno stanje je uvijek povratno, ali obrat ne vrijedi. Povratno stanje koje nije pozitivno povratno je *nul-povratno.*

Definicija 1.2.8. *Neka je $(X_n)_{n\in\mathbb{N}_0}$ Markovljev lanac na skupu stanja $\mathcal{X}$ s prijelaznom matricom P. Neka je $T(x) \coloneqq \{t \geq 1 : P^t(x,x) > 0\}$ skup broja koraka kad je moguće da se lanac vrati u početno stanje $x \in \mathcal{X}$. Neka je $d(x) \coloneqq \gcd T(x)$. Za stanje $x \in \mathcal{X}$ kažemo da je **aperiodično** ako je $d(x) = 1$. U suprotnom je stanje x **periodičko**, a $d(x)$ je **period** tog stanja.*

Ako su sva stanja aperiodična, onda i za lanac kažemo da je aperiodičan. Analogno tomu, za lanac kažemo da je povratan ako su sva stanja povratna. Svojstva perioda i povratnosti se prenose na druga stanja po komuniciranju, pa kod ireducibilnih lanaca sva stanja nose ista svojstva periodičnosti i povratnosti, te ta svojstva odmah pripisujemo i samom lancu. Ako lanac nije aperiodičan, onda se skup stanja može dekomponirati na klase ovisno o periodima stanja. Takva particija skupa stanja naziva se ciklička dekompozicija.

Definicija 1.2.9. *Za (μ, P)-Markovljev lanac reći ćemo da je **tranzitivan** ako za svaki $x, y \in \mathcal{X}$ postoji bijekcija $\varphi = \varphi_{xy} \colon \mathcal{X} \to \mathcal{X}$ takva da vrijedi:*

a) $\varphi(x) = y$,

b) $P(z, w) = P(\varphi(z), \varphi(w)), \forall z, w \in \mathcal{X}$.

Intuitivno, svojstvo tranzitivnosti govori da lanac (u smislu prijelaznih vjerojatnosti) izgleda jednako iz svakog stanja u $\mathcal{X}$.

Granično ponašanje i stacionarna distribucija

Definicija 1.2.10. *Neka je $(X_n)_{n\in\mathbb{N}_0}$ Markovljev lanac sa skupom stanja $\mathcal{X}$ i prijelaznom matricom P. Distribucija $\pi = (\pi(x) : x \in \mathcal{X})$ je **stacionarna distribucija** Markovljevog lanca $(X_n)_{n\in\mathbb{N}_0}$ (odnosno matrice P) ako vrijedi $\pi = \pi P$, tj. po komponentama $\pi(x) = \sum_{z\in\mathcal{X}} \pi(z)P(z, y), \forall y \in \mathcal{X}$.*

Pojam stacionarnosti označava stabilnost vjerojatnosnih svojstava kroz vrijeme, ili drugim riječima, ako Markovljev lanac ima stacionarnu distribuciju u nekom trenutku, ta distribucija ostaje ista i u sljedećem koraku. Stacionarna distribucija za neki Markovljev lanac općenito ne mora postojati, ali u posebnom slučaju kad je lanac ireducibilan i pozitivno povratan, znamo kako stacionarna distribucija izgleda i može se pokazati da je ona jedinstvena.

Teorem 1.2.11. *Neka je $(X_n)_{n\in\mathbb{N}_0}$ ireducibilan Markovljev lanac na skupu stanja $\mathcal{X}$. Ekvivalentno je:*

a) svako stanje je pozitivno povratno,

b) postoji pozitivno povratno stanje,

c) $(X_n)_{n\in\mathbb{N}_0}$ *ima stacionarnu distribuciju* π.

Ako vrijedi c), onda vrijedi:

$$\mathbb{E}_x[T_x] = \frac{1}{\pi(x)},$$

zbog čega je π *jedinstvena stacionarna distribucija.*

Teorem 1.2.12. *Neka je* μ *proizvoljna distribucija na* $\mathcal{X}$ *te neka je* $(X_n)_{n\in\mathbb{N}_0}$ *ireducibilan i aperiodičan* (μ, P)*-Markovljev lanac sa stacionarnom distribucijom* π. *Tada vrijedi:*

$$\lim_{n\to\infty} \mathbb{P}(X_n = y) = \lim_{n\to\infty} P^n(x, y) = \pi(y), \forall x, y \in \mathcal{X}.$$

Teorem 1.2.13 (Ergodski[1] teorem)**.** *Pretpostavimo da je Markovljev lanac* $(X_n)_{n\in\mathbb{N}_0}$ *ireducibilan i pozitivno povratan, te neka je* π *njegova jedinstvena stacionarna distribucija. Pretpostavimo da je* f *nenegativna ili ograničena realna funkcija definirana na* $\mathcal{X}$. *Tada vrijedi:*

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(X_k) = \sum_{x\in\mathcal{X}} f(x)\pi(x)\right) = 1.$$

Od lanca koji konvergira prema stacionarnoj distribuciji očekujemo nakon nekog vremena stabilno kretanje, u skladu sa stacionarnom distribucijom u svakom koraku. Motivacija Poglavlja 2 je pronaći broj koraka kada je lanac dovoljno blizu stacionarnosti.

S druge strane, ergodski teorem poopćuje jaki zakon velikih brojeva (Teorem 5.1) do Markovljevih lanaca. U Poglavlju 5 ćemo reći nešto više o toj konvergenciji uz pomoć Čebiševljeve nejednakosti.

Markovljevi lanci unatrag

Markovljevo svojstvo (1.1) može se iskoristiti da se pokaže da su u Markovljevom lancu budućnost i prošlost nezavisne uvjetno na sadašnjost. To sugerira da postoji neka simetrija između prošlosti i budućnosti te da onda lanac možemo promatrati i

[1]Ergodičnost - svojstvo da je proizvoljan niz reprezentativan za cjelinu; u kontekstu Teorema 1.2.13, gotovo svaka putanja Markovljevog lanca je reprezentativan uzorak za stacionarnu distribuciju

kad vrijeme teče unazad. Treba samo uzeti u obzir da po Teoremu 1.2.12 distribucija lanca teži ka stacionarnoj distribuciji, pa ako bismo pustili vrijeme da teče unatrag, simetriju možemo očuvati samo ako početna distribucija bude i stacionarna.

Definicija 1.2.14. *Stohastička matrica $P = [p_{ij}]_{i,j\in\mathcal{X}}$ i distribucija $\mu = (\mu_i : i, j \in \mathcal{X})$ su u **detaljnoj ravnoteži** ako za njih vrijedi:*

$$\mu(x)P(x,y) = \mu(y)P(y,x), \forall x, y \in \mathcal{X}.$$

Propozicija 1.2.15. *Ako su μ i P u detaljnoj ravnoteži, onda je μ stacionarna za P.*

Motivirani uvjetom detaljne ravnoteže, definiramo matricu $\widehat{P} = \left[\widehat{P}(x,y)\right]_{x,y\in\mathcal{X}}$ po elementima sa:

$$\widehat{P}(x,y) := \frac{\pi(y)P(y,x)}{\pi(x)}, \forall x, y \in \mathcal{X}.$$

Teorem 1.2.16. *Neka je $(X_n)_{n\in\mathbb{N}_0}$ ireducibilan (π, P)-Markovljev lanac, gdje je π stacionarna distribucija. Za $N \in \mathbb{N}$ i $0 \leq n \leq N$ definiramo $Y_n = X_{N-n}$. Tada je $(Y_n)_{0\leq n\leq N}$ ireducibilan $(\pi, \widehat{P})$-Markovljev lanac sa stacionarnom distribucijom π.*

Definicija 1.2.17. *Za ireducibilan (μ, P)-Markovljev lanac $(X_n)_{n\in\mathbb{N}_0}$ kažemo da je **reverzibilan** ako je za sve $N \geq 1$ vektor $(X_{N-n})_{0\leq n\leq N}$ ponovno (μ, P)-Markovljev lanac.*

Teorem 1.2.18. *Neka je $(X_n)_{n\in\mathbb{N}_0}$ ireducibilan (μ, P)-Markovljev lanac. Vrijedi da je lanac $(X_n)_{n\in\mathbb{N}_0}$ reverzibilan ako i samo ako su μ i P u detaljnoj ravnoteži.*

Vrijeme zaustavljanja i jako stacionarno vrijeme

Definicija 1.2.19. *Neka je $\{\mathcal{F}_t\}_{t\in\mathbb{N}_0}$ rastuća familija σ-algebri, tj. za njih vrijedi $\mathcal{F}_t \subset \mathcal{F}_{t+1}, \forall t \in \mathbb{N}$. Familiju $\{\mathcal{F}_t\}_t$ nazivamo **filtracijom**. Za Markovljev lanac $(X_n)_{n\in\mathbb{N}_0}$ kažemo da je **adaptiran** filtraciji $\{\mathcal{F}_t\}_t$ ako je X_t $\mathcal{F}_t$-izmjeriva za sve t. Ako je $\mathcal{F}_t = \sigma(X_1, \ldots, X_t)$, filtraciju nazivamo **prirodnom**.*

Svaki Markovljev lanac je po svojoj definiciji adaptiran prirodnoj filtraciji.

Definicija 1.2.20. *Neka je $(X_n)_{n\in\mathbb{N}_0}$ adaptiran filtraciji $\{\mathcal{F}_t\}_t$. **Vrijeme zaustavljanja** uz filtraciju $\{\mathcal{F}_t\}_t$ je slučajna varijabla τ s vrijednostima u $\mathbb{N} \cup \{\infty\}$ t.d. $\{\tau = t\} \in \mathcal{F}_t$. Ako drugačije nije navedeno, smatra se da je riječ o prirodnoj filtraciji.*

Intuitivno objašnjenje definicije vremena zaustavljanja je da odluka hoćemo li neki proces zaustaviti ili ne, ovisi isključivo o prošlosti i sadašnjosti, ali ne i o budućnosti.

Definicija 1.2.21. *Neka je $(X_n)_{n\in\mathbb{N}_0}$ ireducibilan Markovljev lanac sa stacionarnom distribucijom π. Neka je $(X_n)_{n\in\mathbb{N}_0}$ adaptiran nekoj filtraciji $\{\mathcal{F}_t\}_t$.* ***Stacionarno vrijeme*** *τ za $(X_n)_{n\in\mathbb{N}_0}$ je $\{\mathcal{F}_t\}_t$-vrijeme zaustavljanja koje možda ovisi o početnoj poziciji x, a takvo da X_τ ima distribuciju π:*

$$\mathbb{P}_x(X_\tau = y) = \pi(y), \quad \forall y \in \mathcal{X}.$$

Ako X_τ ima distribuciju π neovisno o τ, tj. vrijedi:

$$\mathbb{P}_x(\tau = t, X_\tau = y) = \mathbb{P}_x(\tau = t)\pi(y), \quad \forall y \in \mathcal{X}, \forall t \in \mathbb{N}_0,$$

onda τ nazivamo ***jakim stacionarnim vremenom****.*

Na kraju, napomenimo još da ćemo u poglavljima koja slijede pretpostavljati da je $\mathcal{X}$ konačan.

Poglavlje 2

Vrijeme miješanja Markovljevog lanca

Ako za neki Markovljev lanac postoji stacionarna distribucija, postavlja se pitanje konvergira li distribucija lanca, dana sa $P^t(x, \cdot)$, prema toj stacionarnoj distribuciji π kad $t \to \infty$. U ovom poglavlju definirat ćemo odgovarajuće norme za mjerenje udaljenosti među spomenutim distribucijama te utvrditi dovoljne uvjete za konvergenciju lanca prema stacionarnoj distribuciji. Ako lanac konvergira prema stacionarnoj distribuciji, sljedeći važan korak je ocjena brzine te konvergencije. Stoga ovo poglavlje uvodi definiciju vremena miješanja, najmanjeg broja koraka t nakon kojeg je distribucija slučajne varijable X_t *blizu* stacionarnoj distribuciji π.

2.1 Udaljenost potpune varijacije

Definicija 2.1.1. *Neka su $\mu, \nu \in \mathbb{R}^{\mathcal{X}}$.* ***Normu potpune varijacije*** *definiramo kao najveću apsolutnu vrijednost dodjeljenu jednom podskupu od $\mathcal{X}$, to jest:*

$$\|\mu\|_{\mathrm{TV}} = \sup_{E \subseteq \mathcal{X}} |\mu(E)|.$$

Posebno, definiramo ***udaljenost potpune varijacije*** *koja mjeri razliku između dvije funkcije μ i ν u smislu norme potpune varijacije:*

$$\|\mu - \nu\|_{\mathrm{TV}} = \sup_{E \subseteq \mathcal{X}} |\mu(E) - \nu(E)|. \tag{2.1}$$

U oba slučaja koristimo pokratu $\mu(E) = \sum_{x \in E} \mu(x), E \subseteq \mathcal{X}$.

Napomena 2.1.2. *Norma potpune varijacije tipično se definira za realne mjere na skupu $\mathcal{X}$. Ako je $\mathcal{F}$ σ-algebra na $\mathcal{X}$ i $\mu\colon \mathcal{F} \to \mathbb{R}$ realna mjera na skupu $\mathcal{X}$, norma potpune varijacije na skupu realnih mjera na $\mathcal{X}$ definira se sa:*

$$\|\mu\|_{\mathrm{TV}} = \sup_{E \in \mathcal{F}} |\mu(E)|.$$

Za mjeru μ možemo definirati realnu funkciju $f_\mu : \mathcal{X} \to \mathbb{R}$ sa:

$$f_\mu(x) = \mu(\{x\}). \tag{2.2}$$

Definicija 2.1.1 je onda poseban slučaj izveden iz definicije norme potpune varijacije za realne mjere. Kako nam je namjera u nastavku mjeriti udaljenost među posebnim tipom funkcija, funkcijama gustoće diskretnih slučajnih varijabli (dosad nazivane distribucijama), pojam norme potpune varijacije definirali smo za realne funkcije na $\mathcal{X}$. Po potrebi ćemo identificirati μ i f_μ preko (2.2) i uvijek pisati μ, ali će iz konteksta biti jasno radi li se o mjeri ili funkciji na $\mathcal{X}$.

Propozicija 2.1.3. *Neka su $\mu, \nu \in \mathbb{R}^{\mathcal{X}}$. Za udaljenost potpune varijacije vrijedi:*

$$\|\mu - \nu\|_{\mathrm{TV}} = \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)| = \sum_{\substack{x \in \mathcal{X}, \\ \mu(x) \geq \nu(x)}} (\mu(x) - \nu(x)).$$

Dokaz. Neka je $B \coloneqq \{x \in \mathcal{X} \colon \mu(x) \geq \nu(x)\}$ i neka je $A \subseteq \mathcal{X}$. Vrijedi:

$$(\mu - \nu)(A) \leq (\mu - \nu)(A \cap B) \leq (\mu - \nu)(B).$$

Prva nejednakost vrijedi jer je $(\mu - \nu)(A \cap B^c) < 0$ po definiciji skupa B, a druga jer $A \cap B \subseteq B$. Analogno se može pokazati i sljedeće:

$$(\nu - \mu)(A) \leq (\nu - \mu)(B^c).$$

Iz prve nejednakosti vidi se da se supremum u (2.1) postiže za B, pa je $\|\mu - \nu\|_{\mathrm{TV}} = (\mu - \nu)(B)$. Slično, iz druge nejednakosti onda slijedi da je $\|\mu - \nu\|_{\mathrm{TV}} = (\nu - \mu)(B^c)$. Ako iskoristimo samo prvu opservaciju, slijedi:

$$\|\mu - \nu\|_{\mathrm{TV}} = \mu(B) - \nu(B) = \sum_{\substack{x \in \mathcal{X}, \\ \mu(x) \geq \nu(x)}} (\mu(x) - \nu(x)),$$

a ako iskoristimo obje opservacije, slijedi:

$$\|\mu - \nu\|_{\mathrm{TV}} = \frac{1}{2}(\mu(B) - \nu(B) + \nu(B^c) - \mu(B^c)) = \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)|$$

$\blacksquare$

Kad su funkcije μ i ν takve da im je slika u $[0,1]$ i $\mu(\mathcal{X}) = \nu(\mathcal{X}) = 1$, one se mogu promatrati kao vjerojatnosne funkcije gustoće. Naime, takva funkcija μ određuje diskretnu slučajnu varijablu X na $\mathcal{X}$ sa $F_X(x) = \sum_{y \le x} \mu(y)$. Egzistencija takve slučajne varijable opravdava definiciju sparivanja.

Sparivanje funkcija $\mu, \nu \colon \mathcal{X} \to [0,1]$ takvih da $\mu(\mathcal{X}) = \nu(\mathcal{X}) = 1$, par je slučajnih varijabli (X, Y) definiran na nekom vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$ takvom da μ određuje marginalnu distribuciju za X, a ν marginalnu distribuciju za Y.
U sljedećoj propoziciji sad možemo dati karakterizaciju udaljenosti potpune varijacije za dvije vjerojatnosne mjere pomoću njihovog sparivanja.

Propozicija 2.1.4. *Neka su μ i ν vjerojatnosne funkcije gustoće na $\mathcal{X}$. Vrijedi:*

$$\|\mu - \nu\|_{\mathrm{TV}} = \inf \left\{ \mathbb{P}(X \neq Y) : (X, Y) \text{ je sparivanje } \mu \text{ i } \nu \right\}.$$

*Infimum u gornjoj jednakosti se postiže, i to sparivanje za koje se infimum postiže nazivamo **optimalnim sparivanjem**.*

Na kraju iznosimo ključan rezultat ovog poglavlja - da bi lanac konvergirao prema stacionarnoj distribuciji (u normi potpune varijacije), dovoljno je da je on bude ireducibilan i aperiodičan(vidi Teorem 1.2.12). Dodatno, u tom slučaju znamo i da se ta konvergencija odvija eksponencijalnom brzinom.
Dokazi Propozicije 2.1.4 i Teorema 2.1.5 mogu se naći u [5].

Teorem 2.1.5 (Teorem konvergencije)**.** *Neka je $(X_n)_{n \in \mathbb{N}_0}$ ireducibilan i aperiodičan lanac s prijelaznom matricom P i stacionarnom distribucijom π. Postoje konstante $\alpha \in \langle 0, 1 \rangle$ i $C > 0$ takve da vrijedi:*

$$\max_{x \in \mathcal{X}} \left\| P^t(x, \cdot) - \pi \right\|_{\mathrm{TV}} \le C\alpha^t.$$

2.2 Vrijeme miješanja

Za lanac koji konvergira prema stacionarnoj distribuciji korisno je ispitati brzinu te konvergencije, odnosno, saznati nakon koliko koraka je lanac svojom distribucijom blizu stacionarnoj. U tu svrhu, prvo ćemo uvesti oznaku za udaljenost od stacionarne distribucije uz čiju pomoć ćemo definirati vrijeme miješanja - minimalan broj koraka potreban da se distribucija lanca nađe u ε-okolini stacionarne distribucije. Funkcijom $d \colon \mathbb{N}_0 \to \mathbb{R}$ definiranom sa:

$$d(t) = \max_{x \in \mathcal{X}} \left\| P^t(x, \cdot) - \pi \right\|_{\mathrm{TV}}$$

označavamo udaljenost distribucije lanca $(X_n)_{n \in \mathbb{N}_0}$ u trenutku t od stacionarne distribucije π.

Napomena 2.2.1. *Prema standardnoj definiciji iz teorije mjere, $L^p(\Omega, \mathcal{P}(\Omega), \mu)$-prostor definiramo na sljedeći način:*

$$L^p = \left\{ f\colon \Omega \to \mathbb{R} \;\middle|\; f \text{ izmjeriva}, \int |f|^p \mathrm{d}\mu < \infty \right\}.$$

Na tom prostoru se onda za $p \geq 1$ definira p-norma sa:

$$\|f\|_p = \left(\int |f|^p \mathrm{d}\mu \right)^{\frac{1}{p}}.$$

Za $p = \infty$ je $\|f\|_\infty = \inf\{M \geq 0 : |f| \leq M \text{ g.s.}\}$, a L^∞ skup izmjerivih funkcija s konačnom ∞-normom.
U posebnom slučaju kad je Ω prebrojiv i mjera μ inducirana nekom nenegativnom izmjerivom funkcijom g preko brojeće mjere $\big(\mu(E) = \sum_{e \in E} g(e)\big)$, taj se prostor označava sa $\ell^p(\Omega, \mathcal{P}(\Omega), \mu)$ ili kraće $\ell^p(\mu)$, a integral iz definicije norme jednak je sumi:

$$\int |f|^p \mathrm{d}\mu = \sum_{\omega \in \Omega} |f(\omega)|^p \mu(\{\omega\}).$$

Uzimajući u obzir (2.2), lako se vidi da je $\mathbb{R}^{\mathcal{X}} = L^p(\mathcal{X}, \mathcal{P}(\mathcal{X}), \pi)$, za svaki $p \in [1, +\infty]$, te p-normu možemo zapisati pomoću stacionarne distribucije:

$$\|f\|_p \coloneqq \begin{cases} \left[\sum_{y \in \mathcal{X}} |f(y)|^p \pi(y)\right]^{\frac{1}{p}}, & 1 \leq p < \infty \\ \max_{y \in \mathcal{X}} |f(y)|, & p = \infty. \end{cases}$$

Da bismo u p-normi mjerili udaljenost lanca od stacionarne distribucije, uvodimo sljedeću oznaku:

$$q_t(x, y) \coloneqq \frac{P^t(x, y)}{\pi(y)}.$$

Ako je lanac blizu stacionarnoj distribuciji, $q_t(x, y)$ bit će blizu 1, zato ima smisla udaljenost $d^{(p)}\colon \mathbb{N}_0 \to \mathbb{R}$ definirati sa:

$$d^{(p)}(t) \coloneqq \max_{x \in \mathcal{X}} \|q_t(x, \cdot) - 1\|_p.$$

Korisno je istaknuti neke osnovne veze između svih dosad uvedenih udaljenosti. Najčešće promatramo slučajeve kad je $p = 1, 2, \infty$, pa kako je $p \mapsto \|f\|_p$ neopadajuća te zbog Propozicije 2.1.3, za $d^{(1)}, d^{(2)}$ i $d^{(\infty)}$ vrijedi:

$$2d(t) = d^{(1)}(t) \leq d^{(2)}(t) \leq d^{(\infty)}(t). \tag{2.3}$$

Kad je lanac reverzibilan, može se pokazati da vrijedi i:

$$d^{(\infty)}(2t) = \left[d^{(2)}(t)\right]^2 = \max_{x\in\mathcal{X}} q_{2t}(x,x) - 1. \tag{2.4}$$

Pomoću funkcija d i $d^{(p)}, p \geq 1$ možemo napokon definirati vremena miješanja.

Definicija 2.2.2. *Neka je $(X_n)_{n\in\mathbb{N}_0}$ Markovljev lanac s prijelaznom matricom P i stacionarnom distribucijom π.*

i) ***Vrijeme miješanja*** *Markovljevog lanca $(X_n)_{n\in\mathbb{N}_0}$ je najmanji broj koraka potreban da distribucija lanca bude od stacionarne distribucije udaljena za najviše $\varepsilon > 0$:*

$$t_{\text{mix}}(\varepsilon) \coloneqq \min\{t \geq 0 : d(t) \leq \varepsilon\}.$$

ii) Analogno, ***ℓ^p-vrijeme miješanja*** *definiramo sa:*

$$t^{(p)}_{\text{mix}}(\varepsilon) \coloneqq \min\{t \geq 0 : d^{(p)}(t) \leq \varepsilon\}.$$

Ako lanac ne konvergira pa je za neki izbor ε udaljenost između distribucija veća od ε, uzimamo da je $\min\varnothing = \infty$. Standardni izbori ε su $\varepsilon = \frac{1}{4}$ za vrijeme miješanja, odnosno $\varepsilon = \frac{1}{2}$ za ℓ^p-vrijeme miješanja, stoga uvodimo oznake:

$$t_{\text{mix}} \coloneqq t_{\text{mix}}\left(\frac{1}{4}\right), \quad t^{(p)}_{\text{mix}} \coloneqq t^{(p)}_{\text{mix}}\left(\frac{1}{2}\right).$$

Parametar $t^{(\infty)}_{\text{mix}}$ često se naziva **uniformnim vremenom miješanja**.

Poglavlje 3

Miješanje karata

Poznat primjer koji neka svoja važna svojstva može pripisati teoriji Markovljevih lanaca i vremenu miješanja je miješanje špila karata. Kada je špil dovoljno dobro promiješan i kojom metodom je najlakše doći do tog stanja pitanja su koja su motivirala autore rada *Trailing the Dovetail Shuffle to its Lair*, [4]. Cilj ovog poglavlja je predstaviti osnovni rezultat tog rada i odgovoriti na pitanje kada možemo očekivati da su karte dobro promiješane.

3.1 Slučajna šetnja na grupi permutacija

Prije nego možemo opisati proces miješanja špila karata kao Markovljev lanac, eksplicitno ćemo navesti neke osnovne definicije i tvrdnje iz teorije grupa koje su nam potrebne za to.

Napomena 3.1.1. *Neka je G neprazan skup i $\cdot\colon G \times G \to G$ binarna operacija na G. Grupa je uređen par $(G, \cdot)$ za koji vrijede sljedeća tri svojstva:*

a) asocijativnost: $(g \cdot h) \cdot i = g \cdot (h \cdot i), \forall g, h, i \in G$,

b) egzistencija neutralnog elementa (jedinice): $(\exists \mathrm{id} \in G)\ (\forall g \in G)\ g \cdot \mathrm{id} = \mathrm{id} \cdot g = g$,

c) egzistencija inverznog elementa: $(\forall g \in G)\ (\exists g^{-1} \in G)\ g \cdot g^{-1} = g^{-1} \cdot g = \mathrm{id}$.

Za podskup $H \subset G$ kažemo da generira G ako je G najmanja grupa koja sadrži H. U tom slučaju se svaki element iz G zapisuje kao umnožak nekih elemenata iz H ili njihovih inverza. Za skup generatora H kažemo da je simetričan ako $h \in H$ implicira da je $h^{-1} \in H$.

Slučajna šetnja slijeva na konačnoj grupi G definira se kao Markovljev lanac čije su prijelazne vjerojatnosti:

$$P(g, hg) = \mu(h), \forall g, h \in G,$$

za neku distribuciju μ. Za μ kažemo da je simetrična na G ako $\mu(g) = \mu(g^{-1}), \forall g \in G$. Distribucija unatrag za μ je distribucija $\widehat{\mu}$ koju definiramo sa $\widehat{\mu}(g) = \mu(g^{-1})$ za sve $g \in G$. Distribucija unatrag određuje i Markovljev lanac unatrag s prijelaznom matricom $\widehat{P}$ (prema oznakama u Definiciji ...). Čak i kad μ nije simetrična, može se pokazati da su μ i $\widehat{\mu}$ jednako udaljene od uniformnosti, što je precizno izrečeno u Propoziciji 3.1.2. U nastavku navodimo još nekoliko važnih svojstava slučajnih šetnji na konačnim grupama. Dokazi tih tvrdnji mogu se naći u [5].

Propozicija 3.1.2. *Neka je P prijelazna matrica slučajne šetnje na konačnoj grupi G s prijelaznom distribucijom μ i neka je $\widehat{P}$ prijelazna matrica slučajne šetnje s prijelaznom distribucijom $\widehat{\mu}$. Ako je π uniformna, za svaki $t \geq 0$ vrijedi:*

$$\left\| P^t(\mathrm{id}, \cdot) - \pi \right\|_{\mathrm{TV}} = \left\| \widehat{P}^t(\mathrm{id}, \cdot) - \pi \right\|_{\mathrm{TV}}.$$

Propozicija 3.1.3. *Neka je $(X_n)_{n \in \mathbb{N}_0}$ slučajna šetnja na konačnoj grupi G s prijelaznom matricom P, te neka je U uniformna funkcija gustoće na G. Tada je U i stacionarna za $(X_n)_{n \in \mathbb{N}_0}$.*

Dokaz. Neka je μ prijelazna distribucija slučajne šetnje. Za $g \in G$ vrijedi:

$$\sum_{h \in G} U(h) P(h, g) = \frac{1}{|G|} \sum_{h \in G} P(h, g) = \frac{1}{|G|} \sum_{k \in G} P(k^{-1}g, g) = \frac{1}{|G|} \sum_{k \in G} \mu(k) = \frac{1}{|G|} = U(g),$$

gdje smo k uveli takav da $k = gh^{-1}$, odnosno $k^{-1} = hg^{-1}$. ∎

Propozicija 3.1.4. *Neka je μ distribucija na konačnoj grupi G. Slučajna šetnja na G s prijelaznom distribucijom μ je ireducibilna ako i samo ako nosač skupa G, tj. skup $S = \{g \in G : \mu(g) > 0\}$, generira cijeli G.*

Propozicija 3.1.5. *Slučajna šetnja na konačnoj grupi G s prijelaznom distribucijom μ je ireducibilna ako je μ simetrična na G.*

3.2 Miješanje karata

Dosad uvedeni pojmovi i svojstva predstavljaju dobre temelje da miješanje karata interpretiramo matematički. Proces miješanja karata zapravo je promjena poretka

karata, odnosno njihovo permutiranje. Zato promatramo grupu svih permutacija n karata i slučajnu šetnju na njoj. Grupu permutacija nazivamo simetričnom grupom i označavamo sa $\mathcal{S}_n$. Pravilo kojim miješamo karte određuje prijelaznu distribuciju između permutacija, odnosno prijelaznu matricu slučajne šetnje na grupi $\mathcal{S}_n$.
Najčešća metoda miješanja karata je takozvani *riffle shuffle*. Postupak miješanja onda podrazumijeva da se špil karata podijeli na dva dijela, a onda iz svakog od ta dva dijela nasumično ispušta karta kako bi se sve karte opet spojile u jedan špil. Početno stanje je uvijek uredno posložen špil, tj. karte su poredane po svojoj vrijednosti. Postoje tri načina kako možemo matematički opisati ovaj proces miješanja:

a) Za $M \sim B\left(n, \frac{1}{2}\right)$ podijelimo špil na dva dijela - prvih M karata i preostalih $n - M$ karata. Relativni poredak unutar dvaju dijelova mora ostati sačuvan, pa postoji $\binom{n}{M}$ jednako vjerojatnih načina da se ti dijelovi pomiješaju (samo od n pozicija odaberemo njih M za prvi dio, a ostatak popunimo drugim dijelom špila);

b) Za $M \sim B\left(n, \frac{1}{2}\right)$, špil podijelimo na prvih M i preostalih $n - M$ karata. Karte spajamo u jednu hrpu tako da ispuštamo po jednu kartu iz prvog ili drugog dijela s vjerojatnošću $\frac{c_i}{a_i+b_i}$, gdje je a_i broj preostalih karata u prvom dijelu prije ispuštanja i-te karte, b_i broj preostalih karata u drugom dijelu prije ispuštanja i-te karte, a c_i je a_i ili b_i, ovisno o tome gledamo li vjerojatnost za ispuštanje iz prvog ili drugog dijela špila redom;

c) U špilu karata svakoj karti dodijelimo indikator 0 ili 1 s jednakom vjerojatnošću. Karte označene nulom onda stavimo na vrh špila, a karte označene jedinicom na dno, poštujući relativni poredak karata u obje skupine.

Rastući niz u nekoj permutaciji σ je najveći podskup uzastopnih elemenata čiji relativni poredak je sačuvan u σ. Na primjer, permutacija (235416) ima tri rastuća niza (234), (56) i (1). Svaka od gore opisanih metoda u jednoj iteraciji stvara dvostruko više rastućih nizova nego što ih je bilo u prethodnom koraku. Prebrojavanjem permutacija karata s dva rastuća niza koje mogu nastati jednom iteracijom metodom a) ili b), može se pokazati da je vjerojatnost da permutacija σ predstavlja raspored špila dana distribucijom:

$$Q(\sigma) = \begin{cases} \frac{n+1}{2^n} & \text{ako je } \sigma = \mathrm{id}, \\ \frac{1}{2^n} & \text{ako } \sigma \text{ ima dva rastuća niza}, \\ 0 & \text{inače}. \end{cases}$$

Metoda c) naizgled je drugačija od prve dvije metode jer od jedne hrpe karata stvara dvije, tj. na neki način djeluje u obratnom smjeru. Može se pokazati da ona zbilja

opisuje šetnju unatrag za metodu a). Zbog Propozicije 3.1.2 dovoljno je istražiti što se događa iteriranjem inverznih miješanja iz metode c).
Ponavljanjem postupka generirat ćemo niz nula i jedinica na svakoj karti u špilu, a pomoću leksikografskog uređaja ćemo karte sortirati prema vrijednostima tih binarnih oznaka. Karte koje imaju iste binarne oznake ostaju u istom relativnom poretku kao u početnom stanju. To znači da su karte promiješane kada ne ostane sačuvan nijedan takav relativni poredak s početka, odnosno kad sve karte imaju međusobno različite binarne oznake.

Propozicija 3.2.1. *Neka je τ broj inverznih miješanja potreban da bi se postigle različite binarne oznake na svim kartama. Tada je τ jako stacionarno vrijeme.*

Dokaz. Prema definiciji jakog stacionarnog vremena, treba pokazati da je X_τ uvjetno na $\tau = t$ uniformno distribuirana, jer je uniformna distribucija stacionarna za slučajnu šetnju na permutacijama prema Propoziciji 3.1.3. U trenutku $\tau = t$, imamo n međusobno različitih oznaka duljine t. Takva binarna matrica veličine $n \times t$ točno određuje jednu permutaciju karata σ. Svaki od $n!$ mogućih načina da se tih n redaka složi u matricu jednako je vjerojatan, a definira neku novu permutaciju. Slijedi da je svaka permutacija jednako vjerojatna, neovisno o t. ∎

Propozicija 3.2.2. *Ako je τ jako stacionarno vrijeme uz početno stanje x, onda vrijedi:*

$$\left\|P^t(x,\cdot) - \pi\right\|_{\mathrm{TV}} \leq \mathbb{P}(\tau > t).$$

Propozicija 3.2.3. *Za riffle shuffle metodu miješanja špila od n karata, uz dovoljno velike n, vrijedi:*

$$t_{\mathrm{mix}} \leq 2 \ln\left(\frac{4n}{3}\right).$$

Dokaz. Neka je τ definiran kao u Propoziciji 3.2.1. Da bismo mogli nešto zaključiti o t_{mix}, trebamo ograničiti $\|P^t(x,\cdot) - \pi\|_{\mathrm{TV}}$ odozgo s $\frac{1}{4}$. Prema Propoziciji 3.2.2, vrijedi $\|P^t(x,\cdot) - \pi\|_{\mathrm{TV}} \leq \mathbb{P}(\tau > t)$, što znači da ćemo traženu ocjenu postići ako ograničimo $P(\tau > t) < \frac{1}{4}$, ili tomu ekvivalentno $P(\tau \leq t) > \frac{3}{4}$. Za računanje $\mathbb{P}(\tau \leq t)$, prebrojavamo moguće izbore binarnih oznaka redom po kartama, uzimajući u obzir da svaka sljedeća mora biti različita od svih prethodnih. Za prvu kartu to može biti bilo koja oznaka, za drugu onda imamo $2^t - 1$ izbora, za treću $2^t - 2$ i tako dalje do n-te karte kada imamo $2^t - (n-1)$ izbora za oznaku od svih 2^t mogućih oznaka duljine t. Tako dolazimo do sljedeće jednakosti:

$$\mathbb{P}(\tau \leq t) = \prod_{k=0}^{n-1} \frac{2^t - k}{2^t} = \prod_{k=0}^{n-1} \left(1 - \frac{k}{2^t}\right).$$

Neka je $t = 2\log_2 \frac{n}{c}$, tj. $2^t = \frac{n^2}{c^2}$. Vrijedi:

$$\begin{aligned}
\ln \prod_{k=0}^{n-1} \left(1 - \frac{k}{2^t}\right) &= \sum_{k=0}^{n-1} \ln\left(1 - \frac{c^2 k}{n^2}\right) \\
&= -\sum_{k=0}^{n-1} \left(\frac{c^2 k}{n^2} + \mathcal{O}\left(\frac{k^2}{n^4}\right)\right) \\
&= -\frac{c^2 n(n-1)}{2n^2} + \mathcal{O}\left(\frac{n^3}{n^4}\right) \\
&= -\frac{c^2}{2} + \mathcal{O}\left(\frac{1}{n}\right).
\end{aligned}$$

Kako $\frac{1}{n} \to 0$ kad $n \to \infty$, iz prethodnog možemo zaključiti:

$$\lim_{n\to\infty} \frac{\mathbb{P}(\tau \le t)}{e^{-\frac{c^2}{2}}} = 1.$$

Zato traženu ogradu $\mathbb{P}(\tau \le t) > \frac{3}{4}$ dobivamo za $c < \sqrt{2\ln\frac{4}{3}}$ koja slijedi ograničavanjem izraza iz nazivnika gornjeg limesa, tj. $e^{-\frac{c^2}{2}} > \frac{3}{4}$. Jedna povoljna vrijednost koja zadovoljava taj uvjet za c je $c = \frac{3}{4}$ što uvrštavanjem u $t = 2\log_2 \frac{n}{c}$ dokazuje tvrdnju. ∎

Klasičan špil karata ima 52 karte, pa prema Propoziciji 3.2.3 znamo da je gornja granica za vrijeme miješanja jednaka $2\ln\frac{4\cdot 52}{3} \approx 8.48$, tj. već nakon 8 iteracija miješanja smo vrlo blizu uniformnosti i možemo karte smatrati dobro promiješanima.

Poglavlje 4

Svojstvene vrijednosti prijelazne matrice Markovljevog lanca

4.1 Spektralna reprezentacija prijelazne matrice

Za prijelaznu matricu P svojstvene vektore nalazimo u vektorskom prostoru svih realnih funkcija na prostoru stanja $\mathcal{X}$ i tad svojstveni vektor nazivamo svojstvenom funkcijom matrice P. Točnije, **svojstvena funkcija** matrice P uz svojstvenu vrijednost λ je funkcija $f\colon \mathcal{X} \to \mathbb{R}$ za koju vrijedi:

$$P(x, \cdot)f(\cdot) = \lambda f(x), \forall x \in \mathcal{X}.$$

Propozicija 4.1.1. *Neka je $(X_n)_{n\in\mathbb{N}_0}$ Markovljev lanac s konačnim skupom stanja $\mathcal{X}$ i prijelaznom matricom P. Vrijedi:*

i) $|\lambda| \leq 1, \quad \forall \lambda \in \sigma(P)$,

ii) *ako je $(X_n)_{n\in\mathbb{N}_0}$ ireducibilan, $1 \in \sigma(P)$ i svojstveni potprostor uz $\lambda = 1$ generiran je sa $\mathbf{1} \in \mathbb{R}^{\mathcal{X}}$.*

Dokaz. i) Neka je $\lambda \in \sigma(P)$ i f pripadna svojstvena funkcija. Vrijedi sljedeće:

$$\|Pf\|_\infty = \max_{x\in\mathcal{X}} |P(x, y)f(y)| \leq \|f\|_\infty .$$

Kako je $Pf = \lambda f$, slijedi:

$$|\lambda| \, \|f\|_\infty = \|Pf\|_\infty \leq \|f\|_\infty ,$$

iz čega dijeljenjem s $\|f\|_\infty$ slijedi tvrdnja.

ii) Direktnom provjerom se vidi da je $1 \in \sigma(P)$ uz svojstvenu funkciju $\mathbf{1}$. Za ireducibilan lanac je stacionarna distribucija π jedini vektor za koji vrijedi $\pi P = \pi$. Transponiranjem te jednakosti slijedi da je π^τ jedina svojstvena funkcija za $\lambda = 1$ uz matricu P^t. (Jedinstvenost vrijedi do na multiplikativnu konstantu.) Još treba primijetiti da je geometrijska kratnost od $\lambda = 1$ jednaka, bilo da se gleda kao svojstvena vrijednost od P ili P^t. To vrijedi zato što je $r(P - I) = r((P - I)^\tau)$, a po teoremu o rangu i defektu (Teorem 1.1.7) su onda i defekti odgovarajućih matrica jednaki, tj. geometrijska kratnost od $\lambda = 1$ je 1 u oba slučaja zbog gore pokazane jedinstvenosti do na množenje konstantom. Zaključujemo da $\mathbf{1}$ generira cijeli svojstveni potprostor od $\lambda = 1$ uz matricu P.

■

Napomena 4.1.2. *Na spomenutom prostoru $\mathbb{R}^{\mathcal{X}}$ možemo, osim standardnog skalarnog produkta $\langle \cdot, \cdot \rangle$, definirati i skalarni produkt na sljedeći način:*

$$\langle f, g \rangle_\pi = \sum_{x \in \mathcal{X}} f(x) g(x) \pi(x) \quad \forall f, g \in \mathbb{R}^{\mathcal{X}},$$

gdje je π stacionarna distribucija promatranog lanca. Ovaj skalarni produkt inducira 2-normu na $\ell^2(\pi)$.

Ako pretpostavimo da je $(X_n)_{n \in \mathbb{N}_0}$ reverzibilan uz početnu distribuciju π, to je ekvivalentno činjenici da vrijede uvjeti detaljne ravnoteže (Teorem 1.2.18):

$$\pi(x) P(x, y) = \pi(y) P(y, x) \quad \forall x, y \in \mathcal{X}. \tag{4.1}$$

Korištenjem (4.1) *imamo:*

$$\begin{aligned} \langle Pf, g \rangle_\pi &= \sum_{x \in \mathcal{X}} (Pf)(x) g(x) \pi(x) = \sum_{x, y \in \mathcal{X}} P(x, y) f(y) g(x) \pi(x) \\ &= \sum_{x, y \in \mathcal{X}} P(y, x) f(y) g(x) \pi(y) = \sum_{y \in \mathcal{X}} f(y) (Pg)(y) \pi(y) = \langle f, Pg \rangle_\pi, \end{aligned}$$

iz čega direktno slijedi da je P hermitska matrica, a svojstvene vrijednosti hermitskih matrica su realni brojevi. Zato se u nastavku često ograničavamo na reverzibilne lance, jer su tada svojstvene vrijednosti realne, dok za lance koji nisu reverzibilni to ne mora biti slučaj.

Propozicija 4.1.3. *Neka je $(X_n)_{n \in \mathbb{N}_0}$ reverzibilan Markovljev lanac uz početnu distribuciju π i prijelaznu matricu P.*

i) Svojstvene funkcije $\{f_j\}_{j=1}^{|\mathcal{X}|}$ od P čine ortonormiranu bazu za prostor $\ell^2(\pi)$.

ii) Za elemente matrice P^t vrijedi:

$$\frac{P^t(x,y)}{\pi(y)} = \sum_{j=1}^{|\mathcal{X}|} f_j(x) f_j(y) \lambda_j^t. \tag{4.2}$$

Dokaz. i) Neka je $D_\pi = \operatorname{diag}(\pi)$ i definiramo operator A na $\left(\mathbb{R}^{\mathcal{X}}, \langle \cdot, \cdot \rangle\right)$ sa:

$$A = D_\pi^{\frac{1}{2}} P D_\pi^{-\frac{1}{2}}.$$

Kako zbog reverzibilnosti lanca za P vrijedi (4.1), slijedi da je A simetrična:

$$A(x,y) = \frac{\sqrt{\pi(x)}}{\sqrt{\pi(y)}} P(x,y) = \frac{\sqrt{\pi(x)}}{\sqrt{\pi(y)}} \cdot \frac{\pi(y)}{\pi(x)} \cdot P(y,x) = \frac{\sqrt{\pi(y)}}{\sqrt{\pi(x)}} P(y,x) = A(y,x).$$

Za simetričnu matricu A postoji ortonormirana baza (ONB) u $\left(\mathbb{R}^{\mathcal{X}}, \langle \cdot, \cdot \rangle\right)$ u kojoj se A može dijagonalizirati. Neka je to $\{\varphi_j\}_{j=1}^{|\mathcal{X}|}$ s pripadnim svojstvenim vrijednostima $\{\lambda_j\}_{j=1}^{|\mathcal{X}|}$.

Da bismo pokazali tvrdnju iz iskaza, definiramo $f_j := D_\pi^{-\frac{1}{2}} \varphi_j$. Tvrdimo da je $\{f_j\}_{j=1}^{|\mathcal{X}|}$ ONB iz $\ell^2(\pi)$ u kojoj se P može dijagonalizirati uz svojstvene vrijednosti $\{\lambda_j\}_{j=1}^{|\mathcal{X}|}$. Da su funkcije f_j svojstvene funkcije za P slijedi iz:

$$P f_j = P D_\pi^{-\frac{1}{2}} \varphi_j = D_\pi^{-\frac{1}{2}} D_\pi^{\frac{1}{2}} P D_\pi^{-\frac{1}{2}} \varphi_j = D_\pi^{-\frac{1}{2}} A \varphi_j = \lambda_j D_\pi^{-\frac{1}{2}} \varphi_j = \lambda_j f_j,$$

a da su međusobno ortogonalne uz $\langle \cdot, \cdot \rangle_\pi$ vidi se iz:

$$\begin{aligned} \delta_{ij} =& \langle \varphi_i, \varphi_j \rangle = \langle D_\pi^{\frac{1}{2}} f_i, D_\pi^{\frac{1}{2}} f_j \rangle \\ =& \sum_{x \in \mathcal{X}} D_\pi^{\frac{1}{2}} f_i(x) D_\pi^{\frac{1}{2}} f_j(x) = \sum_{x \in \mathcal{X}} \sqrt{\pi(x)} f_i(x) \sqrt{\pi(x)} f_j(x) \\ =& \sum_{x \in \mathcal{X}} f_i(x) f_j(x) \pi(x) = \langle f_i, f_j \rangle_\pi . \end{aligned}$$

ii) Definiramo:

$$\delta_y(x) = \begin{cases} 1, & y = x \\ 0, & y \neq x. \end{cases}$$

Proizvoljna funkcija iz $\mathbb{R}^{\mathcal{X}}$, pa tako i δ_y, može se zapisati u ONB $\{f_j\}_{j=1}^{|\mathcal{X}|}$:

$$\delta_y = \sum_{j=1}^{|\mathcal{X}|} \langle \delta_y, f_j \rangle_\pi f_j = \sum_{j=1}^{|\mathcal{X}|} \left(\sum_{x \in \mathcal{X}} \delta_y(x) f_j(x) \pi(x) \right) f_j = \sum_{j=1}^{|\mathcal{X}|} f_j(y) \pi(y) f_j$$

Zbog $P^t f_j = \lambda_j^t f_j$ i $P^t(x,y) = (P^t \delta_y)(x)$ imamo:

$$P^t(x,y) = \left(P^t \sum_{j=1}^{|\mathcal{X}|} f_j(y)\pi(y) f_j \right)(x) = \left(\sum_{j=1}^{|\mathcal{X}|} f_j(y)\pi(y) P^t f_j \right)(x)$$
$$= \left(\sum_{j=1}^{|\mathcal{X}|} f_j(y)\pi(y) \lambda_j^t f_j \right)(x),$$

a nakon dijeljenja s $\pi(y)$:

$$\frac{P^t(x,y)}{\pi(y)} = \sum_{j=1}^{|\mathcal{X}|} f_j(x) f_j(y) \lambda_j^t.$$

■

Napomena 4.1.4. *Tvrdnja iz Propozicije 4.1.3.ii) se može poboljšati do:*

$$\frac{P^t(x,y)}{\pi(y)} = 1 + \sum_{j=2}^{|\mathcal{X}|} f_j(x) f_j(y) \lambda_j^t. \tag{4.3}$$

Prema Propoziciji 4.1.1.ii), možemo uzeti $\lambda_1 = 1$ i $f_1 = \mathbf{1}$, zbog čega je prvi sumand jednak 1.

Za neku slučajnu varijablu $X \colon \Omega \to \mathcal{X}$ s distribucijom π ima smisla promatrati $\mathbb{E}[\varphi(X)] =: E_\pi(\varphi)$. U sljedećoj lemi pokazujemo da kad svojstvena funkcija djeluje na slučajnu varijablu s distribucijom π, njena očekivana vrijednost je 0, a za taj zaključak nam nije potrebna pretpostavka o reverzibilnosti lanca.

Lema 4.1.5. *Neka je φ svojstvena funkcija matrice P uz svojstvenu vrijednost $\lambda \neq 1$. Tada je $\mathbb{E}_\pi(\varphi) = 0$.*

Dokaz. Po definiciji stacionarne distribucije, za π vrijedi $\pi = \pi P$. Po definiciji svojstvene funkcije, za φ vrijedi $P\varphi = \lambda\varphi$. Koristeći te dvije jednakosti, slijedi:

$$E_\pi[\varphi] = \sum_{x \in \mathcal{X}} \pi(x)\varphi(x) = \pi\varphi = \pi P \varphi = \lambda \pi \varphi = \lambda E_\pi(\varphi).$$

Kako je $\lambda \neq 1$, mora biti $E_\pi(\varphi) = 0$. ■

4.2 Ograde na vrijeme miješanja

Pokazali smo u Propoziciji 4.1.1 da su svojstvene vrijednosti po apsolutnoj vrijednosti manje od 1. U nastavku ćemo pokazati kako udaljenost od ruba jedinične kugle utječe na konvergenciju lanca prema stacionarnoj distribuciji.

Definicija 4.2.1. *Neka je $(X_n)_{n\in\mathbb{N}_0}$ Markovljev lanac s prijelaznom matricom P.*

i) ***Apsolutni spektralni razmak*** *je razlika $\gamma_* := 1 - \lambda_*$, pri čemu je:*

$$\lambda_* := \max\{|\lambda| : \lambda \in \sigma(P), \lambda \neq 1\}.$$

ii) Neka je $(X_n)_{n\in\mathbb{N}_0}$ reverzibilan. Indeksiramo svojstvene vrijednosti po veličini:

$$1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_{|\mathcal{X}|} \geq -1.$$

Spektralni razmak *lanca $(X_n)_{n\in\mathbb{N}_0}$ je $\gamma := 1 - \lambda_2$.*

iii) ***Vrijeme relaksacije*** *reverzibilnog lanca $(X_n)_{n\in\mathbb{N}_0}$ sa spektralnim razmakom γ_* definirano je sa:*

$$t_{\text{rel}} := \frac{1}{\gamma_*}.$$

Važnost prethodno definiranih pojmova leži u naredna dva teorema, gdje pomoću t_{rel} i π nalazimo gornju i donju ogradu na t_{mix}.

Teorem 4.2.2. *Neka je $(X_n)_{n\in\mathbb{N}_0}$ ireducibilan i reverzibilan Markovljev lanac s prijelaznom matricom P i neka je $\pi_{\min} := \min_{x\in\mathcal{X}} \pi(x)$. Tada vrijedi:*

$$t_{\text{mix}}(\varepsilon) \leq \left\lceil t_{\text{rel}} \left(\frac{1}{2} \ln \frac{1}{\pi_{\min}} + \ln \frac{1}{2\varepsilon} \right) \right\rceil \leq t_{\text{rel}} \ln \frac{1}{\varepsilon \pi_{\min}} \tag{4.4}$$

$$t_{\text{mix}}^{(\infty)}(\varepsilon) \leq \left\lceil t_{\text{rel}} \ln \frac{1}{\varepsilon \pi_{\min}} \right\rceil. \tag{4.5}$$

Dokaz. Primjenom nejednakosti trokuta i Cauchy-Schwarzove nejednakosti na (4.3) slijedi:

$$\left| \frac{P^t(x,y)}{\pi(y)} - 1 \right| \leq \sum_{j=2}^{|\mathcal{X}|} |f_j(x) f_j(y)| \, \lambda_*^t \leq \lambda_*^t \left[\sum_{j=2}^{|\mathcal{X}|} f_j^2(x) \sum_{j=2}^{|\mathcal{X}|} f_j^2(y) \right]^{\frac{1}{2}}. \tag{4.6}$$

Dalje, zato što je $\{f_j\}_{j=1}^{|\mathcal{X}|}$ ortonormirana baza u $\ell^2(\pi)$, imamo:

$$\begin{aligned}
\pi(x) = \langle \delta_x, \delta_x \rangle_\pi &= \left\langle \sum_{j=1}^{|\mathcal{X}|} f_j(x)\pi(x) f_j, \sum_{j=1}^{|\mathcal{X}|} f_j(x)\pi(x) f_j \right\rangle_\pi \\
&= \sum_{j=1}^{|\mathcal{X}|} \langle f_j(x)\pi(x) f_j, f_j(x)\pi(x) f_j \rangle_\pi \\
&= \sum_{j=1}^{|\mathcal{X}|} \left(\sum_{y\in\mathcal{X}} f_j^2(x)\pi^2(x) f_j^2(y)\pi(y) \right) \\
&= \sum_{j=1}^{|\mathcal{X}|} f_j^2(x)\pi^2(x) \underbrace{\left(\sum_{y\in\mathcal{X}} f_j^2(y)\pi(y) \right)}_{\|f\|_\pi = 1} \\
&= \pi^2(x) \sum_{j=1}^{|\mathcal{X}|} f_j^2(x),
\end{aligned}$$

iz čega slijedi:

$$\sum_{j=2}^{|\mathcal{X}|} f_j^2(x) \le \sum_{j=1}^{|\mathcal{X}|} f_j^2(x) = \frac{1}{\pi(x)}. \tag{4.7}$$

Nejednakosti (4.6) i (4.7) zajedno daju:

$$\begin{aligned}
\left| \frac{P^t(x,y)}{\pi(y)} - 1 \right| &\le \lambda_*^t \left[\sum_{j=2}^{|\mathcal{X}|} f_j^2(x) \sum_{j=2}^{|\mathcal{X}|} f_j^2(y) \right]^{\frac{1}{2}} \le \frac{\lambda_*^t}{\sqrt{\pi(x)\pi(y)}} \\
&\le \frac{\lambda_*^t}{\pi_{\min}} = \frac{(1-\gamma_*)^t}{\pi_{\min}} \le \frac{e^{-\gamma_* t}}{\pi_{\min}}.
\end{aligned}$$

Po definiciji $t_{\text{mix}}^{(\infty)}$ iz gornje nejednakosti lako slijedi (4.5), dok za (4.4) treba još uzeti u obzir da vrijedi sljedeća posljedica nejednakosti (2.3) i (2.4):

$$d(t) \le \frac{1}{2}\sqrt{d^{(\infty)}(2t)} \le \frac{e^{-\gamma_* t}}{2\sqrt{\pi_{\min}}}.$$

$\blacksquare$

Teorem 4.2.3. *Neka je $(X_n)_{n\in\mathbb{N}_0}$ ireducibilan i aperiodičan Markovljev lanac i neka je $\lambda \neq 1$ svojstvena vrijednost za prijelaznu matricu P. Vrijedi:*

$$t_{\mathrm{mix}}(\varepsilon) \geq \left(\frac{1}{1-|\lambda|} - 1\right) \ln \frac{1}{2\varepsilon}. \tag{4.8}$$

Posebno, ako je lanac $(X_n)_{n\in\mathbb{N}_0}$ još i reverzibilan, onda imamo:

$$t_{\mathrm{mix}}(\varepsilon) \geq (t_{\mathrm{rel}} - 1) \ln \frac{1}{2\varepsilon}. \tag{4.9}$$

Dokaz. Možemo pretpostaviti da $\lambda \neq 0$ jer u slučaju $\lambda = 0$ tvrdnja slijedi trivijalno. Neka je f svojstvena funkcija uz λ, tj. vrijedi $Pf = \lambda f$. Pozivajući se na Lemu 4.1.5, znamo da je $E_\pi(f) = 0$. Uz Propoziciju 2.1.3 onda vrijedi:

$$|\lambda^t f(x)| = |P^t f(x) - E_\pi(f)| = \left| \sum_{y\in\mathcal{X}} [P^t(x,y)f(y) - \pi(y)f(y)] \right| \leq 2d(t)\|f\|_\infty.$$

Odaberemo x takav da $|f(x)| = \|f\|_\infty$ i $t_{\mathrm{mix}}(\varepsilon)$ uvrstimo na mjesto t, pa dobivamo:

$$|\lambda|^{t_{\mathrm{mix}}(\varepsilon)} \leq 2\varepsilon.$$

Izoliranjem $t_{\mathrm{mix}}(\varepsilon)$ iz gornje nejednakosti slijedi (4.8), odnosno (4.9) kad (4.8) minimiziramo po $|\lambda| \neq 1$. ∎

Teorem 4.2.4. *Neka je $(X_n)_{n\in\mathbb{N}_0}$ reverzibilan Markovljev lanac s prijelaznom matricom P čije su svojstvene vrijednosti:*

$$1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_{|\mathcal{X}|} \geq -1$$

s pripadnim svojstvenim funkcijama $\{f_j\}_{j=1}^{|\mathcal{X}|}$. Vrijedi:

$$4\left\|P^t(x,\cdot) - \pi\right\|_{\mathrm{TV}}^2 \leq \left\|\frac{P^t(x,\cdot)}{\pi(\cdot)} - 1\right\|_2^2 = \sum_{j=2}^{|\mathcal{X}|} f_j^2(x)\lambda_j^{2t}.$$

Posebno, ako je lanac tranzitivan, vrijedi:

$$4\left\|P^t(x,\cdot) - \pi\right\|_{\mathrm{TV}}^2 \leq \left\|\frac{P^t(x,\cdot)}{\pi(\cdot)} - 1\right\|_2^2 = \sum_{j=2}^{|\mathcal{X}|} \lambda_j^{2t}.$$

Dokaz. Prema Propoziciji 4.1.3 vrijedi:

$$\left\|P^t(x,\cdot)-\pi\right\|_2^2=\underbrace{\|\pi(\cdot)\|_2^2}_{=1}\left\|\frac{P^t(x,\cdot)}{\pi(\cdot)}-1\right\|_2^2=\left\|\sum_{j=2}^{|\mathcal{X}|}\lambda_j f_j(x) f_j\right\|_2^2=\sum_{j=2}^{|\mathcal{X}|}f_j(x)^2\lambda_j^{2t},$$

a zbog Propozicije 2.1.3 i činjenice da je preslikavanje $p\mapsto\|f\|_p$ neopadajuće vrijedi i:

$$\begin{aligned}4\left\|P^t(x,\cdot)-\pi\right\|_{\mathrm{TV}}^2&=\left(\sum_{y\in\mathcal{X}}\left|P^t(x,y)-\pi(y)\right|\right)^2=\left(\sum_{y\in\mathcal{X}}\pi(y)\left|\frac{P^t(x,y)}{\pi(y)}-1\right|\right)^2\\&=\left\|\frac{P^t(x,y)}{\pi(y)}-1\right\|_1^2\le\left\|\frac{P^t(x,y)}{\pi(y)}-1\right\|_2^2.\end{aligned}$$

Kombinacijom tih (ne)jednakosti slijedi prva tvrdnja.
Ako je lanac dodatno i tranzitivan, onda izraz $\left\|\frac{P^t(x,\cdot)}{\pi(\cdot)}-1\right\|_2^2$ ne ovisi o x, tj. za bilo koji $x_0\in\mathcal{X}$ vrijedi:

$$\left\|\frac{P^t(x_0,\cdot)}{\pi(\cdot)}-1\right\|_2^2=\sum_{j=2}^{|\mathcal{X}|}f_j(x)^2\lambda_j^{2t}.$$

Uprosječivanjem po x uz distribuciju π slijedi:

$$\left\|\frac{P^t(x_0,\cdot)}{\pi(\cdot)}-1\right\|_2^2=\sum_{j=2}^{|\mathcal{X}|}\underbrace{\sum_{x\in\mathcal{X}}f_j(x)^2\pi(x)}_{\|f\|_\pi^2=1}\lambda_j^{2t},$$

što u kombinaciji s prethodnom daje drugu tvrdnju iz iskaza. ∎

Poglavlje 5

Ocjene uz granično ponašanje lanaca

Neka su dane distribucija π na konačnom skupu $\mathcal{X}$ i funkcija $f\colon \mathcal{X} \to \mathbb{R}$. Cilj nam je odrediti $E_\pi(f) := \mathbb{E}[f(X)] = \sum_{x\in\mathcal{X}} f(x)\pi(x)$, gdje je X slučajna varijabla s distribucijom π. Ako nije moguće izračunati egzaktan iznos sume, možda zbog veličine skupa $\mathcal{X}$ ili zahtjevnosti izračuna $f(x)$, jedan način procjene vrijednosti $E_\pi(f)$ je uprosječiti vrijednosti funkcije f nad uzorcima iz π.
Kad je moguće generirati nezavisan i jednakodistribuiran (n.j.d.) niz $(X_n)_{n\in\mathbb{N}}$ iz π, onda je i $(f(X_n))_n$ n.j.d. niz. Jaki zakon velikih brojeva nam onda osigurava da $\frac{1}{t}\sum_{s=1}^{t} f(X_s) \to E_\pi(f)$ gotovo sigurno. Štoviše, uz Čebiševljevu nejednakost, možemo naći i veličinu uzorka n takvu da s $\varepsilon > 0$ ograničimo vjerojatnost da je greška takve procjene veća od $\eta > 0$. Sve ovo precizno je iskazano u naredne tri tvrdnje. U tim tvrdnjama koristimo i oznaku $\mathrm{Var}_\pi(f) = \mathrm{Var}(f(X))$, za $X \sim \pi$, analogno definiciji $E_\pi(f)$.

Teorem 5.1 (Jaki zakon velikih brojeva, [7])**.** *Neka je $(X_n)_{n\in\mathbb{N}}$ niz n.j.d. slučajnih varijabli. Tada niz $\left(\frac{1}{n}\sum_{j=1}^{n} X_j\right)_{n\in\mathbb{N}}$ konvergira gotovo sigurno ako i samo ako $\mathbb{E}X_1$ postoji i u tom slučaju vrijedi:*

$$(g.s.) \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} X_j = \mathbb{E}X_1.$$

Propozicija 5.2 (Čebiševljeva nejednakost, [7])**.** *Neka je X slučajna varijabla s konačnim očekivanjem i varijancom. Za proizvoljan $\varepsilon > 0$ vrijedi:*

$$\mathbb{P}(|X - \mathbb{E}X| \geq \varepsilon) \leq \frac{\mathrm{Var}X}{\varepsilon^2}.$$

Teorem 5.3. *Neka je $f\colon \mathcal{X} \to \mathbb{R}$ i $(X_n)_{n\in\mathbb{N}}$ n.j.d. niz slučajnih varijabli s vrijednostima u $\mathcal{X}$ i distribucijom π. Vrijedi:*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{s=1}^{n} f(X_s) - E_\pi(f)\right| > \eta\right) \leq \frac{\mathrm{Var}_\pi(f)}{\eta^2 n}.$$

Posebno, ako je $n \geq \frac{\mathrm{Var}_\pi(f)}{\eta^2\varepsilon}$, lijeva je strana ograđena odozgo sa ε.

Dokaz. Tvrdnja je direktna posljedica primjene Čebiševljeve nejednakosti (Teorem 5.2) na slučajnu varijablu $\frac{1}{n}\sum_{s=1}^{n} f(X_s)$ čija je varijanca $\frac{1}{n}\mathrm{Var}_\pi(f)$. ■

Ipak, moguće je da je π takva distribucija da nije jednostavno ili uopće nije moguće generirati nezavisne uzorke iz nje. Taj problem rješava posebna klasa Monte Carlo metoda simulacija, *Markov Chain Monte Carlo* ili MCMC. MCMC metode temelje se na konstrukciji Markovljevog lanca čija je stacionarna distribucija π. Prema teoriji o vremenima miješanja koju smo dosad predstavili u radu, takav lanac bi nakon t_{mix} koraka trebao imati distribuciji približno jednaku π.
Teorem 5.3 nije primjenjiv na Markovljeve lance zbog pretpostavke nezavisnosti niza $(X_n)_{n\in\mathbb{N}_0}$, ali sličan rezultat vrijedi i za reverzibilne Markovljeve lance. Dovoljno je preskočiti prvih t_{mix} koraka (tzv. *burn-in period*), a onda uzeti uzorak duljine barem $\frac{4\mathrm{Var}_\pi(f)}{\eta^2\varepsilon\gamma}$ da bismo iz Markovljevog lanca dobili dovoljno dobru procjenu za $E_\pi(f)$ u smislu Teorema 5.3. Ostatak poglavlja posvećujemo preciznom iskazu i dokazu te tvrdnje.

Lema 5.4. *Neka je $(X_n)_{n\in\mathbb{N}_0}$ reverzibilan Markovljev lanac i neka je φ svojstvena funkcija prijelazne matrice P uz svojstvenu vrijednost λ takva da vrijedi $\langle\varphi,\varphi\rangle_\pi = 1$. Za $\lambda \neq 1$ vrijedi:*

$$\mathbb{E}_\pi\left[\left(\sum_{s=0}^{t-1}\varphi(X_s)\right)^2\right] \leq \frac{2t}{1-\lambda}.$$

Ako je $f\colon \mathcal{X} \to \mathbb{R}$ proizvoljna funkcija za koju vrijedi $E_\pi(f) = 0$, onda vrijedi:

$$\mathbb{E}_\pi\left[\left(\sum_{s=0}^{t-1} f(X_s)\right)^2\right] \leq \frac{2tE_\pi(f^2)}{\gamma}.$$

Dokaz. Razvojem kvadrata sume i po linearnosti očekivanja slijedi:

$$\mathbb{E}_\pi\left[\left(\sum_{s=0}^{t-1}\varphi(X_s)\right)^2\right] = \sum_{s=0}^{t-1}\mathbb{E}_\pi\left[\varphi(X_s)^2\right] + 2\sum_{r=0}^{t-1}\sum_{s=r+1}^{t-1}\mathbb{E}_\pi\left[\varphi(X_r)\varphi(X_s)\right]. \tag{5.1}$$

Prvo ćemo izračunati $\mathbb{E}_\pi\left[\varphi(X_r)\varphi(X_s)\right]$ iz (5.1). Neka je $r < s$. Vrijedi:

$$\begin{aligned}\mathbb{E}_\pi\left[\varphi(X_r)\varphi(X_s)\right] &= \mathbb{E}_\pi\left[\mathbb{E}_\pi\left[\varphi(X_r)\varphi(X_s)|X_r\right]\right] \\ &= \mathbb{E}_\pi\left[\varphi(X_r)\mathbb{E}_\pi\left[\varphi(X_s)|X_r\right]\right] \\ &= \mathbb{E}_\pi\left[\varphi(X_r)P^{s-r}\varphi(X_r)\right],\end{aligned}$$

gdje $\mathbb{E}_\pi\left[\varphi(X_s)|X_r\right] = P^{s-r}\varphi(X_r)$ vrijedi zbog:

$$\begin{aligned}\mathbb{E}_\pi\left[\varphi(X_s)|X_r = x\right] &= \sum_{y\in\mathcal{X}} \varphi(y)\mathbb{P}_\pi(X_s = y|X_r = x) \\ &= \sum_{y\in\mathcal{X}} \varphi(y)P^{s-r}(x,y) \\ &= P^{s-r}(x,\cdot)\varphi(\cdot) = (P^{s-r}\varphi)(x).\end{aligned}$$

Budući da je $P\varphi = \lambda\varphi$ i $\langle\varphi,\varphi\rangle_\pi = E_\pi(\varphi^2) = 1$, slijedi:

$$\mathbb{E}_\pi\left[\varphi(X_r)\varphi(X_s)\right] = \lambda^{s-r}\mathbb{E}_\pi\left[\varphi(X_r)^2\right] = \lambda^{s-r}E_\pi(\varphi^2) = \lambda^{s-r}.$$

Dalje, pojednostavljujemo dvostruku sumu iz (5.1):

$$\begin{aligned}\sum_{r=0}^{t-1}\sum_{s=r+1}^{t-1}\mathbb{E}_\pi\left[\varphi(X_r)\varphi(X_s)\right] &= \sum_{r=0}^{t-1}\sum_{s=r+1}^{t-1}\lambda^{s-r} = \sum_{r=0}^{t-1}\sum_{s=1}^{t-1-r}\lambda^{s} \\ &= \sum_{r=0}^{t-1}\left(\frac{1-\lambda^{t-r}}{1-\lambda} - 1\right) = \sum_{r=0}^{t-1}\left(\frac{\lambda-\lambda^{t-r}}{1-\lambda}\right) \\ &= \frac{t\lambda}{1-\lambda} - \frac{\lambda^t}{1-\lambda}\sum_{r=0}^{t-1}\lambda^{-r} = \frac{t\lambda}{1-\lambda} - \frac{\lambda^t}{1-\lambda}\cdot\frac{1-\lambda^{-t}}{1-\lambda^{-1}} \\ &= \frac{t\lambda}{1-\lambda} - \frac{\lambda}{1-\lambda}\cdot\frac{1-\lambda^t}{1-\lambda}.\end{aligned}$$

Uzimajući u obzir dosad izvedeno te da je $\mathbb{E}_\pi\left[\varphi^2(X_s)\right] = E_\pi(\varphi) = 1$, (5.1) možemo raspisati dalje:

$$\begin{aligned}\mathbb{E}_\pi\left[\left(\sum_{s=0}^{t-1}\varphi(X_s)\right)^2\right] &= \sum_{s=0}^{t-1}\mathbb{E}_\pi\left[\varphi(X_s)^2\right] + 2\sum_{r=0}^{t-1}\sum_{s=r+1}^{t-1}\mathbb{E}_\pi\left[\varphi(X_r)\varphi(X_s)\right], \\ &= t + \frac{2t\lambda - 2\lambda g(t)}{1-\lambda} = \frac{t(1+\lambda) - 2\lambda g(t)}{1-\lambda}\end{aligned}$$

gdje je $g(\lambda) \coloneqq \frac{1-\lambda^t}{1-\lambda}$. Za $\lambda \in \langle -1, 1\rangle$ je $g(\lambda) \geq 0$. Ovisno o λ gornje ograde na brojnik iz prethodnog izraza su:

$$\begin{aligned} \lambda \in \langle 0,1\rangle &\implies t(1+\lambda) - 2\lambda g(\lambda) \leq t(1+\lambda) \leq 2t \\ \lambda \in [-1,0], t \geq 2 &\implies g(\lambda) \leq 1, \\ &\qquad t(1+\lambda) - 2\lambda g(\lambda) \leq t(1+\lambda) - t\lambda \leq 2t, \end{aligned}$$

što dokazuje prvu tvrdnju.

Za drugu tvrdnju, neka je $f\colon \mathcal{X} \to \mathbb{R}$ takva da $E_\pi(f) = 0$. Neka je $\{f_j\}_{j=1}^{|\mathcal{X}|}$ ONB svojstvenih funkcija od P. Funkcija f zapisuje se u toj ONB kao $f = \sum_{j=1}^{|\mathcal{X}|} a_j f_j$, gdje su $a_j = \langle f, f_j\rangle_\pi$. Treba primijetiti da je:

$$a_1 = \langle f, f_1\rangle_\pi = \langle f, \mathbf{1}\rangle_\pi = E_\pi(f) = 0,$$

a prema Parsevalovoj jednakosti (Propozicija 1.1.5.ii)), vrijedi još i:

$$E_\pi(f) = \langle f, f\rangle_\pi = \sum_{j=1}^{|\mathcal{X}|} a_j^2.$$

Ako definiramo $G_j \coloneqq \sum_{s=0}^{t-1} f_j(X_s)$, uz $a_1 = 0$ vrijedi da je:

$$\sum_{s=0}^{t-1} f(X_s) = \sum_{s=0}^{t-1} \left(\sum_{j=1}^{|\mathcal{X}|} a_j f_j(X_s) \right) = \sum_{j=2}^{|\mathcal{X}|} a_j \sum_{s=0}^{t-1} f_j(X_s) = \sum_{j=2}^{|\mathcal{X}|} a_j G_j.$$

Nadalje, za $r \leq s$ i $j \neq k$ vrijedi:

$$\begin{aligned} \mathbb{E}_\pi \left[f_j(X_s) f_k(X_r) \right] =& \mathbb{E}_\pi \left[f_k(X_r) \mathbb{E}_\pi \left[f_j(X_s) | X_r \right] \right] \\ =& \mathbb{E}_\pi \left[f_k(X_r) P^{s-r} f_j(X_r) \right] \\ =& \lambda^{s-r} \mathbb{E}_\pi \left[f_k(X_r) f_j(X_r) \right] = 0. \end{aligned}$$

Iz toga slijedi i da je $\mathbb{E}_\pi [G_j G_k] = 0$ za $j \neq k$, pa u razvoju sume kvadrata samo su kvadratni članovi različiti od nule:

$$\mathbb{E}_\pi \left[\left(\sum_{s=0}^{t-1} f(X_s) \right)^2 \right] = \mathbb{E}_\pi \left[\left(\sum_{j=2}^{|\mathcal{X}|} a_j G_j \right)^2 \right] = \sum_{j=2}^{|\mathcal{X}|} \mathbb{E}_\pi \left[a_j^2 G_j^2 \right] = \sum_{j=2}^{|\mathcal{X}|} a_j^2 \mathbb{E}_\pi \left[G_j^2 \right].$$

Još treba primijetiti da je G_j definiran kao izraz u prvoj tvrdnji za $\varphi = f_j$, zato prema prethodno dokazanome možemo zaključiti:

$$\sum_{j=2}^{|\mathcal{X}|} a_j^2 \mathbb{E}_\pi \left[G_j^2 \right] = \sum_{j=2}^{|\mathcal{X}|} \frac{2t a_j^2}{1-\lambda_j} \leq 2t \sum_{j=1}^{|\mathcal{X}|} \frac{a_j^2}{1-\lambda_2} = \frac{2t E_\pi(f^2)}{\gamma}.$$

■

Teorem 5.5. *Neka je $(X_n)_{n\in\mathbb{N}_0}$ reverzibilan Markovljev lanac. Ako je $r \geq t_{\text{mix}}\left(\frac{\varepsilon}{2}\right)$ i $t \geq \frac{4\text{Var}_\pi(f)}{\eta^2\varepsilon\gamma}$, onda za bilo koje početno stanje $x \in \mathcal{X}$ vrijedi:*

$$\mathbb{P}_x\left(\left|\frac{1}{t}\sum_{s=0}^{t-1} f(X_{r+s}) - E_\pi(f)\right| \geq \eta\right) \leq \varepsilon.$$

Dokaz. Bez smanjenja općenitosti, neka je f takva da $E_\pi(f) = 0$. U suprotnom samo promatramo $f - E_\pi(f)$. Neka je μ_r optimalno sparivanje za $P^r(x,\cdot)$ i π. Prema Propoziciji 2.1.4 za μ_r vrijedi:

$$\sum_{y\neq x} \mu_r(x,y) = \|P^r(x,\cdot) - \pi\|_{\text{TV}}\,. \tag{5.2}$$

Definiramo proces $(Y_t, Z_t)_{t\in\mathbb{N}_0}$ tako da $(Y_0, Z_0) \sim \mu_r$, a onda uz dano (Y_0, Z_0) su $(Y_t)_{t\in\mathbb{N}}$ i $(Z_t)_{t\in\mathbb{N}}$ dva nezavisna lanca koji se kreću po $\mathcal{X}$ prema prijelaznoj matrici P, dok se prvi put ne sretnu u istom stanju, nakon čega ostaju zauvijek međusobno isti, a stanja i dalje mijenjaju prema matrici P. Drugim riječima, definirali smo Markovljev lanac $(Y_t, Z_t)_{t\in\mathbb{N}_0}$ s početnom distribucijom μ_r i prijelaznom matricom Q koja je dana sa:

$$Q((y,z),(u,v)) = \begin{cases} P(y,u) & \text{ako je } y = z \text{ i } u = v, \\ P(y,u)P(z,v) & \text{ako je } y \neq z, \\ 0 & \text{inače.} \end{cases}$$

Prema toj konstrukciji, proces $(Y_s)_{s\geq 0}$ je Markovljev lanac s prijelaznom matricom P i početnom distribucijom $P^r(x,\cdot)$, pa možemo zaključiti da je jednako distribuiran kao $(X_{r+s})_{s\geq 0}$. Također, $(Z_s)_{s\geq 0}$ je Markovljev lanac s prijelaznom matricom π i početnom distribucijom π.
Kako je $(Y_0, Z_0) \sim \mu_r$, prema (5.2) vrijedi:

$$\mathbb{P}(Y_0 \neq Z_0) = \|P^r(x,\cdot) - \pi\|_{\text{TV}}\,. \tag{5.3}$$

Zbog $Y_s \sim X_{r+s}$ vrijedi:

$$\mathbb{P}_x\left(\left|\frac{1}{t}\sum_{s=0}^{t-1} f(X_{r+s}) - E_\pi(f)\right| \geq \eta\right) = \mathbb{P}\left(\left|\frac{1}{t}\sum_{s=0}^{t-1} f(Y_s) - E_\pi(f)\right| \geq \eta\right),$$

gdje više ne pišemo $\mathbb{P}_x$ već samo $\mathbb{P}$ jer znamo $Y_0 \sim P^r(x,\cdot)$. Ovisno o tome je li $Y_0 = Z_0$ ili ne, slijedi:

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{1}{t}\sum_{s=0}^{t-1} f(Y_s) - E_\pi(f)\right| \geq \eta\right) &= \underbrace{\mathbb{P}(Y_0 = Z_0)}_{\leq 1}\,\mathbb{P}\left(\left|\frac{1}{t}\sum_{s=0}^{t-1} f(Z_s) - E_\pi(f)\right| \geq \eta\right) \\
&\quad + \mathbb{P}(Y_0 \neq Z_0)\,\underbrace{\mathbb{P}\left(\left|\frac{1}{t}\sum_{s=0}^{t-1} f(Y_s) - E_\pi(f)\right| \geq \eta \,\middle|\, Y_0 \neq Z_0\right)}_{\leq 1} \\
&\leq \mathbb{P}(Y_0 \neq Z_0) + \mathbb{P}\left(\left|\frac{1}{t}\sum_{s=0}^{t-1} f(Z_s) - E_\pi(f)\right| \geq \eta\right).
\end{aligned}$$

Po definiciji $t_{\mathrm{mix}}(\varepsilon)$ i (5.3), za $r \geq t_{\mathrm{mix}}\left(\frac{\varepsilon}{2}\right)$ vrijedi da je $\mathbb{P}(Y_0 \neq Z_0) \leq \frac{\varepsilon}{2}$. Uz pomoć Leme 5.4 dobivamo još da vrijedi sljedeća ograda:

$$\begin{aligned}
\mathrm{Var}\left(\frac{1}{t}\sum_{s=0}^{t-1} f(Z_s)\right) &= \frac{1}{t^2}\mathrm{Var}_\pi\left(\sum_{s=0}^{t-1} f(Z_s)\right) \\
&= \frac{1}{t^2}\left(\mathbb{E}_\pi\left[\left(\sum_{s=0}^{t-1} f(Z_s)\right)^2\right] - \underbrace{\mathbb{E}_\pi\left[\left(\sum_{s=0}^{t-1} f(Z_s)\right)\right]^2}_{=0,\text{ jer } E_\pi(f)=0}\right) \\
&\overset{\text{Lema 5.4}}{\leq} \frac{1}{t^2}\cdot\frac{2t\mathbb{E}_\pi[f^2]}{\gamma} = \frac{2\mathrm{Var}_\pi(f)}{t\gamma}.
\end{aligned}$$

Čebiševljeva najednakost iz Teorema 5.2 i gornja ograda na $\mathrm{Var}\left(\frac{1}{t}\sum_{s=0}^{t-1} f(Z_s)\right)$ daju:

$$\mathbb{P}\left(\left|\frac{1}{t}\sum_{s=0}^{t-1} f(Z_s) - E_\pi(f)\right| \geq \eta\right) \leq \frac{2\mathrm{Var}_\pi(f)}{\eta^2 t\gamma}.$$

Iz $\frac{2\mathrm{Var}_\pi(f)}{\eta^2 t\gamma} \leq \frac{\varepsilon}{2}$ dobivamo da za $t \geq \frac{4\mathrm{Var}_\pi(f)}{\eta^2\varepsilon\gamma}$ je $\mathbb{P}\left(\left|\frac{1}{t}\sum_{s=0}^{t-1} f(Z_s) - E_\pi(f)\right| \geq \eta\right) < \frac{\varepsilon}{2}$, što dokazuje tvrdnju teorema. ■

Bibliografija

[1] David Aldous i Persi Diaconis, *Shuffling Cards and Stopping Times*, The American Mathematical Monthly **93** (1986), 333–348.

[2] Damir Bakić, *Linearna algebra*, `https://web.math.pmf.unizg.hr/~bakic/la/linearna_algebra_sk_7.pdf`, pristupljeno u lipnju 2023.

[3] ———, *Normirani prostori*, `https://web.math.pmf.unizg.hr/nastava/np/predavanja/np-2122.pdf`, pristupljeno u lipnju 2023.

[4] Dave Bayer i Persi Diaconis, *Trailing the Dovetail Shuffle to its Lair*, The Annals of Applied Probability **2** (1992), 294–313.

[5] David A. Levin, Yuval Peres i Elizabeth L. Wilmer, *Markov chains and Mixing times*, American Mathematical Society, 2017.

[6] Rudi Mrazović, *Mjera i integral*, `https://www.dropbox.com/s/fgpz096bvduutn7/MII-skripta.pdf?dl=0&utm_source=Mjera+i+integral+-+predavanja`, pristupljeno u travnju 2023.

[7] Nikola Sarapa, *Teorija vjerojatnosti*, Školska knjiga, Zagreb, 2002.

[8] Zoran Vondraček, *Markovljevi lanci*, `https://web.math.pmf.unizg.hr/~vondra/ml12-predavanja.html`, pristupljeno u lipnju 2023.

Sažetak

Cilj ovoga rada je proučiti granično ponašanje Markovljevih lanaca. Na početku dajemo kratak pregled osnovnih definicija i teorema iz područja Markovljevih lanaca. Teorem o konvergenciji ireducibilnog aperiodičnog lanca ka stacionarnoj distribuciji motivira nas na proučavanje brzine te konvergencije. Zato uvodimo normu potpune varijacije kao mjeru udaljenosti između dvije distribucije, te vrijeme miješanja kao broj koraka potreban da se lanac svojom distribucijom dovoljno približi stacionarnoj distribuciji.

Promatrajući prijelaznu matricu reverzibilnog lanca kao linearan operator u odgovarajućem unitarnom prostoru, pokazujemo da je ta matrica hermitska te konstruiramo jednu ortonormiranu bazu. Također, pomoću svojstvenih vrijednosti prijelazne matrice definiramo pojmove (apsolutnog) spektralnog razmaka i vremena relaksacije, koji u konačnici određuju gornju i donju ogradu na vrijeme miješanja.

Također, proučit ćemo dvije važne primjene. Prva primjena je optimalno miješanje karata, pitanje koje je potaknulo razvoj teorije vremena miješanja. U drugom primjeru poopćujemo ideju jakog zakona velikih brojeva i Čebiševljeve nejednakosti na Markovljeve lance, što je dio teorijske podloge na kojoj se zasnivaju MCMC metode simuliranja.

Summary

The aim of this paper is to study the limiting behavior of Markov chains. We begin by introducing the fundamental definitions and theorems of Markov chain theory. A theorem which states that an irreducible and aperiodic Markov chain approaches its stationary distribution motivates us to study the speed of that convergence. Therefore, we define the total variation norm as a measure of distance between two probability distributions, and mixing times as the number of steps needed for a Markov chain to come close to the stationary distribution.

Applying the theory of linear operators on a reversible transition matrix within a appropriately defined unitary space allows us to prove that the matrix is self-adjoint and so we construct an orthonormal basis which will be useful in upcoming proofs. Using eigenvalues of the transition matrix, we define the (absolute) spectral gap and relaxation time of a Markov chain and show how those values are used to establish upper and lower bounds on mixing time.

We also discuss two important consequences of the mixing time theory. The first example is a problem that sparked the development of the mixing time theory in the first place – the optimal number of shuffles for a deck of cards. The second application is a generalization of the law of large numbers and Chebyshev inequality to Markov chains, which is strongly related to the theory MCMC methods are built upon.

Životopis

Anna-Lena Batković rođena je 9.2.1999. u Neumünsteru u Njemačkoj. Po završetku prirodoslovno-matemtičke gimnazije u Srednjoj školi Tina Ujevića u Kutini 2017. godine, upisuje preddiplomski studij matematike na Prirodoslovno-matematičkom fakultetu Sveučilišta u Zagrebu. Školovanje nastavlja upisom diplomskog studija Matematička statistika na istom fakultetu 2020. godine.