Vježbe - Slučajni procesi
IX. dio

Ergodski teorem

Teorem 1 (Jaki zakon velikih brojeva).

Neka je $\{Y_n,\ n \geq 1\}$ *niz nezavisnih jednako distribuiranih slučajnih varijabli na vjerojatnosnom prostoru* $(\Omega, \mathcal{F}, P)$ *takvih da je* $E|Y_1| < \infty$ *i stavimo* $EY_1 = \mu$. *Tada je*

$$P\left(\lim_{n\to\infty} \frac{Y_1 + \ldots + Y_n}{n} = \mu\right) = 1,$$

tj.

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{g.s.} \mu.$$

Teorem 2 (Ergodski teorem).

Neka je $\{X_n,\ n \in \mathbb{N}_0\}$ ireducibilan i pozitivno povratan Markovljev lanac s jedinstvenom stacionarnom distribucijom π te f nenegativna ili ograničena realna funkcija definirana na S. Tada vrijedi:

$$P\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(X_i) = \sum_{k\in S} f(k)\pi_k\right) = 1,$$

tj.

$$\frac{1}{n}\sum_{i=0}^{n-1} f(X_i) \xrightarrow{g.s.} \sum_{k\in S} f(k)\pi_k.$$

- Definirajmo

$$N_i(n) = \sum_{k=0}^{n-1} 1_{\{X_k = i\}}, \quad n \in \mathbb{N},$$

vrijeme koje lanac provede u stanju i prije trenutka n.
Specijalno, za $f = 1_{\{j\}}$ iz ergodskog teorema dobivamo

Korolar 1.

Za ireducibilan i pozitivno povratan Markovljev lanac sa stacionarnom distribucijom π vrijedi

$$P\left(\lim_{n\to\infty} \frac{N_j(n)}{n} = \pi_j\right) = 1, \quad \forall j \in S.$$

Korolar 2.

Neka $f : S \to \mathbb{R}$ ograničena funkcija. Tada za svaku početnu distribuciju lanca X vrijedi

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} E[f(X_k)] = \sum_{j\in S} f(j)\pi_j.$$

Specijalno, za $f = 1_{\{j\}}$ te početno stanje i dobivamo

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} p_{ij}^{(k)} = \pi_j.$$

Zadaci

Zadatak 1.

Neka je Markovljev lanac sa skupom stanja $S = \{0, 1, 2\}$ *i matricom prijelaznih vjerojatnosti*

$$\Pi = \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.2 & 0.5 & 0.3 \\ 0.1 & 0.7 & 0.2 \end{bmatrix}$$

model za predviđanje sutrašnjih vremenskih prilika na temelju poznatih informacija o današnjem vremenu. Pri tome je s 0 *označeno oblačno vrijeme, s* 1 *kišno vrijeme, a s* 2 *sunčano vrijeme. Odredite dugoročno prosječan broj sunčanih dana (tj. prosječan broj sunčanih dana kroz dulji vremenski period).*

Zadatak 2 (Bonus-malus sustav automobilskog osiguranja).

U bonus-malus sustavu automobilskog osiguranja visina osiguranikove premije proporcionalna je broju potraživanja od osiguravajućeg društva (odnosno broju šteta) u prethodno osiguranoj godini. Promotrimo pojednostavljen sustav bonus-malus osiguranja, tj. pretpostavimo da razlikujemo samo četiri skupine rizika: $0, 1, 2, 3$ pri čemu je 0 skupina najnižeg rizika, a 3 skupina najvišeg rizika. Osiguranik prelazi u skupinu nižeg rizika samo ako nije imao nikakvih potraživanja od osiguravajućeg društva u prethodnoj godini. Ako je u prethodnoj godini imao k prijavljenih šteta, onda prelazi u skupinu k višeg rizika, do najviše 3. Npr. ako je bio u skupini 0 i prijavio dvije štete, u idućoj godini će biti svrstan u skupinu 2.

Pretpostavimo da broj šteta u jednoj godini možemo modelirati nezavisno od prethodnih Poissonovom slučajnom varijablom s parametrom $\lambda = \frac{1}{2}$.

a) *Objasnite zašto prelaske osiguranika iz jedne skupine rizika u drugu možemo modelirati Markovljevim lancem. Odredite skup stanja i matricu prijelaznih vjerojatnosti tog Markovljevog lanca. Je li on ireducibilan i aperiodičan?*

b) *Odredite graničnu distribuciju ovog Markovljevig lanca.*

c) *Neka su visine premija definirane funkcijom*

$$f(k) = \begin{cases} 200 & , \quad k = 0 \\ 200 + 50 \cdot 2^k & , \quad k \in \{1, 2, 3\} \end{cases},$$

gdje je k oznaka skupine rizika. Odredite dugoročno prosječnu visinu premije osiguranja jedne police.

Zadatak 3.

Pretpostavimo da promatramo skladište ograničenog kapaciteta, tj. neka u promatrano skladište stane najviše c istovrsnih proizvoda. Potražnja za proizvodima u skladištu u n-tom danu modelirana je slučajnom varijablom D_n *za koju vrijedi*

$$P(D_n \geq i) = \frac{1}{2^i}, \quad i \in \mathbb{N}_0.$$

Ako na kraju n-tog dana u skladištu ostane m ili manje proizvoda skladište se nadopunjuje do punog kapaciteta, a u suprotnom (ako ostane više od m proizvoda) se ne nadopunjuje. Zbog jednostavnosti pretpostavimo da je $c = 3$ *i promotrimo slučajeve* $m = 0$, $m = 1$ *i* $m = 2$.

a) *Definirajte slučajnu varijablu kojom modeliramo broj proizvoda u skladištu na kraju* $(n+1)$*-og dana.*
b) *Za sva tri slučaja odredite skup stanja i matricu prijelaznih vjerojatnosti Markovljevog lanca* $\{X_n, n \in \mathbb{N}_0\}$. *Provjerite jesu li u sva tri slučaja Markovljevi lanci ireducibilni i aperiodični.*
c) *Odredite stacionarne distribucije ovih triju Markovljevih lanaca.*
d) *Definirajte funkciju koja opisuje očekivanu nerealiziranu potražnju (tj. broj proizvoda koji su traženi ali nisu mogli biti isporučeni jer ih, zbog ograničenog kapaciteta skladišta, nije bilo)* $(n+1)$*-og dana. Odredite dugoročno očekivanu nerealiziranu potražnju.*

Zadatak 4.

Rukometaš Fabijan igra u napadu i broj postignutih golova na jednoj utakmici ovisi o njegovim ozljedama. Ako je

0	-	*zdrav, onda daje* 6 *golova,*
1	-	*lakše ozlijeđen, onda daje* 4 *gola,*
2	-	*teže ozlijeđen, onda daje* 1 *gol.*

Prijelazi njegove spremnosti izmedu dvije utakmice modelirani su Markovljevim lancem sa skupom stanja $S = \{0, 1, 2\}$ *i matricom prijelaznih vjerojatnosti*

$$\Pi = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.4 & 0.4 \end{bmatrix}.$$

a) *Kolika je dugoročna proporcija vremena u kojem je Fabijan zdrav?*
b) *Koliki je očekivani broj utakmica između dvije teže ozljede?*
c) *Koliki je dugoročno prosječan broj golova koje Fabijan da po utakmici?*

Primjer: Google PageRank

- Pretpostavimo da na svijetu postoji N web stranica. Postavimo model kretanja web stranicama kao Markovljev lanac u kojem je svako stanje jedna web stranica. Htjeli bi napraviti algoritam za rangiranje web stranica.

1. Svaka stranica je jedno stanje $i = 1, \dots, N$.
2. Dodajmo izmišljeno stanje - *restart stranica* - i označimo je kao stanje 0.
3. Definirajmo prijelazne vjerojatnosti:

(a) Neka je $0 < d < 1$ konstanta takva da je $1 - d$ vjerojatnost da surfajući stranicama prestanemo pratiti linkove i restartiramo navigiranje (npr. sami upišemo stranicu ili bookmark). Svako stanje i, $0 \leq i \leq N$ ima vjerojatnost $1 - d$ prelaska u stanje 0, tj. $p_{i0} = 1 - d$. Dakle, na kojoj god stranici se nalazimo postoji fiksna vjerojatnost da ćemo restartirati surfanje umjesto ići po linkovima.

(b) Vjerojatnost prelaska iz stanja 0 u neko stanje $j \neq 0$ je $p_{0,j} = \frac{d}{N}$, tj. iz restartirajućeg stanja vjerojatnost odlaska na bilo koju stvarnu stranicu je jednaka.

(c) Vjerojatnost prelaska iz stanja $i \neq 0$ u stanje $j \neq 0$ je

$$p_{ij} = \frac{d}{C(i)} 1_{\{i \text{ ima link na } j\}},$$

gdje je $C(i)$ broj linkova sa stranice i. Znači, za svaku stranicu osim vjerojatnosti odlaska u restart $1 - d$, ostatak vjerojatnosti d je uniformno raspoređena među svim stranicama na koje se može putem linka otići s i. Ako i ne sadrži link na j tada je $p_{ij} = 0$.

- Intuitivno, htjeli bi rangirati stranice tako da ona na kojoj se dugoročno provede više vremena ima što bolji rang.
- Je li lanac ireducibilan? U svakom trenutku možemo restartirati surfanje i otići na svaku stranicu.
- Je li pozitivno povratan? Konačan skup stanja, ireducibilan pa i pozitivno povratan.
- Jedinstvena stacionarna distribucija po ergodskom teoremu daje prosječno vrijeme boravka za svako stanje:

$$\pi_0 = \sum_{i=0}^{N} \pi_i p_{i0} = \sum_{i=0}^{N} \pi_i (1-d) = 1-d$$

$$\begin{aligned} \pi_j &= \pi_0 p_{0j} + \sum_{i=1}^{N} \pi_i p_{ij} \\ &= (1-d)\frac{d}{N} + \sum_{\{i : i \text{ ima link na } j\}} \pi_i \frac{d}{C(i)} \end{aligned}$$

- Da bi dobili originalni Google PageRank pomnožimo sve s $\frac{N}{d}$:

$$\frac{N}{d}\pi_j = (1-d) + \sum_{\{i:i \text{ ima link na } j\}} \left(\frac{N}{d}\pi_i\right)\frac{d}{C(i)}$$

- Definiramo PageRank kao $PR(j) = \frac{N}{d}\pi_j$ i dobijemo

$$PR(j) = (1-d) + \sum_{\{i:i \text{ ima link na } j\}} PR(i)\frac{d}{C(i)},$$

  što je izvorni Google PageRank.
- Ergodski teorem zapravo kaže da ako surfamo webom po ovom Markovljevom lancu tada vjerojatnost posjeta nekoj stranici u nekom trenutku konvergira, i ta vjerojatnost je ista bez obzira odakle krenemo sa surfanjem (početna distribucija)
- Za PageRank se prosječno provedeno vrijeme skalira s $\frac{N}{d}$, što je nebitno jer poredak ostaje isti.
- Neke procjene za d su oko 0.85.