Poglavlje 12

Markovljeva slučajna polja i Monte Carlo metoda

Markovljevo svojstvo slučajnog niza $X = (X_n : n \geq 0)$ ima kao posljedicu da je za sve $n \geq 1$, slučajna varijabla X_n uvjetno nezavisna s familijom $(X_k : k \notin \{n-1, n, n+1\})$ uz dane (X_{n-1}, X_{n+1}). Zaista,

$$
\begin{aligned}
&\mathbb{P}(X_n = i_n, X_k = i_k, 1 \leq k \leq n-2, X_l = i_l, n+2 \leq l \leq m \,|\, X_{n-1} = i_{n-1}, X_{n+1} = i_{n+1}) \\
&= \frac{\mathbb{P}(X_k = i_k, 1 \leq k \leq m)}{\mathbb{P}(X_{n-1} = i_{n-1}, X_{n+1} = i_{n+1})} \\
&= \frac{\mathbb{P}(X_k = i_k, 1 \leq k \leq n-1)\mathbb{P}(X_l = i_l, n \leq l \leq m \,|\, X_{n-1} = i_{n-1})}{\mathbb{P}(X_{n-1} = i_{n-1}, X_{n+1} = i_{n+1})} \\
&= \frac{\mathbb{P}(X_k = i_k, 1 \leq k \leq n-1)\mathbb{P}(X_l = i_l, n+1 \leq l \leq m \,|\, X_{n-1} = i_{n-1})}{\mathbb{P}(X_{n-1} = i_{n-1}, X_{n+1} = i_{n+1})} \cdot \\
&\quad \cdot \frac{\mathbb{P}(X_l = i_l, n \leq l \leq m \,|\, X_{n-1} = i_{n-1})}{\mathbb{P}(X_l = i_l, n+1 \leq l \leq m \,|\, X_{n-1} = i_{n-1})} \\
&= \frac{\mathbb{P}(X_k = i_k, 1 \leq k \leq n-1, X_l = i_l, n+1 \leq l \leq m)}{\mathbb{P}(X_{n-1} = i_{n-1}, X_{n+1} = i_{n+1})} \cdot \\
&\quad \cdot \frac{\mathbb{P}(X_{n-1} = i_{n-1}, X_n = i_n, X_l = i_l, n+1 \leq l \leq m)}{\mathbb{P}(X_{n-1} = i_{n-1}, X_l = i_l, n+1 \leq l \leq m)} \\
&= \mathbb{P}(X_k = i_k, 1 \leq k \leq n-2, X_l = i_l, n+2 \leq l \leq m \,|\, X_{n-1} = i_{n-1}, X_{n+1} = i_{n+1}) \cdot \\
&\quad \cdot \frac{\mathbb{P}(X_{n-1} = i_{n-1}, X_n = i_n, X_{n+1} = i_{n+1})}{\mathbb{P}(X_{n-1} = i_{n-1}, X_{n+1} = i_{n+1})} \\
&= \mathbb{P}(X_k = i_k, 1 \leq k \leq n-2, X_l = i_l, n+2 \leq l \leq m | \, X_{n-1} = i_{n-1}, X_{n+1} = i_{n+1}) \cdot \\
&\quad \cdot \mathbb{P}(X_{n-1} = i_{n-1}, X_{n+1} = i_{n+1} \,|\, X_n = i_n) \,.
\end{aligned}
$$

Do sada smo na niz $(X_n : n \geq 0)$ gledali kao na niz slučajnih varijabli indeksiranih vremenskim parametrima. Promotrimo sada geometrijsku sliku: $n \geq 0$ smatramo mjestom (pozicijom) uz koje nam je vezana slučajna varijabla X_n. Dakle, istovremeno nam je dana familija slučajnih varijabli $(X_n : n \geq 0)$ indeksirana mjestima $(n \geq 0)$. Svako mjesto $n \geq 1$ ima dva susjeda, $n-1$ i $n+1$, koji definiraju okolinu mjesta n. Sada Markovljevo svojstvo možemo iskazati na sljedeći način: za sve $n \geq 1$, vrijednost slučajne varijable na mjestu n uvjetno je nezavisna od vrijednosti slučajnih varijabli na mjestima $k \notin \{n-1, n, n+1\}$ uz dane vrijednosti slučajnih varijabli na okolini (mjesta n) $\{n-1, n+1\}$.

Definicija 12.1 *Neka je V konačan skup čije elemente v zovemo* mjestima*, te neka je Λ konačan skup kojeg zovemo* fazni prostor*. * Slučajno polje *na V s fazama u Λ je familija $X = (X(v) : v \in V)$ slučajnih varijabli $X(v)$ s vrijednostima u Λ.*

Skup $S = \Lambda^V$ (svih funkcija s V u Λ) zovemo *konfiguracijski prostor*. Slučajno polje možemo promatrati kao slučajnu varijablu s vrijednostima u konfiguracijskom prostoru S. Na primjer, ako je $V = \{1, 2, \ldots, N\}$, $N \in \mathbb{N}$, tada je slučajno polje N-dimenzionalni slučajni vektor čije komponente primaju vrijednosti u Λ (zaista, Λ^V možemo identificirati s Kartezijevim produktom Λ^N). Zanimljiviji slučaj mjesta je skup $V = \{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\}$ (ovdje implicitno pretpostavljamo geometriju prostora mjesta).

Za danu konfiguraciju $x = (x(v) : v \in V)$ i $A \subset V$ definiramo $x(A) = (x(v) : v \in A)$ i zovemo restrikcija od x na A. Označimo li komplement od A sa $V \setminus A$, tada pišemo $x = (x(A), x(V \setminus A))$. Specijalno, za $v \in V$, $x = (x(v), x(V \setminus v))$.

Zanimat će nas slučajna polja koja imaju Markovljevo svojstvo karakterizirano lokalnim interakcijama. Zbog toga moramo znati koja mjesta u V su blizu jedno drugoga, odnosno preciznije, potreban je pojam okolina.

Definicija 12.2 Sustav okolina *na V je familija $E = (\mathcal{N}_v : v \in V)$ podskupova od V takva da za sve $v \in V$*

(a) $v \notin \mathcal{N}_v$,

(b) $w \in \mathcal{N}_v \ \Rightarrow \ v \in \mathcal{N}_w$.

Podskup $\mathcal{N}_v$ zovemo okolinom *mjesta v. Uređen par (V, E) zovemo* graf*. * Granica *skupa $A \subset V$ je po definiciji skup $\partial A = (\cup_{v \in A} \mathcal{N}_v) \setminus A$.*

Označimo $\widetilde{\mathcal{N}}_v = \mathcal{N}_v \cup \{v\}$.

Definicija 12.3 *Slučajno polje X naziva se* Markovljevo slučajno polje *s obzirom na sustav okolina $E = (\mathcal{N}_v : v \in V)$ ako su za svako mjesto $v \in V$ familije slučajnih varijabli $X(v)$ i $X(V \setminus \mathcal{N}_v)$ uvjetno nezavisne uz dane $X(\mathcal{N}_v)$, t.j.,*

$$
\begin{aligned}
&\mathbb{P}(X(v) = x(v), X(V \setminus \widetilde{\mathcal{N}}_v) = x(V \setminus \widetilde{\mathcal{N}}_v) \,|\, X(\mathcal{N}_v) = x(\mathcal{N}_v)) \\
&= \ \mathbb{P}(X(v) = x(v) \,|\, X(\mathcal{N}_v) = x(\mathcal{N}_v))\, \mathbb{P}(X(V \setminus \widetilde{\mathcal{N}}_v) = x(V \setminus \widetilde{\mathcal{N}}_v) \,|\, X(\mathcal{N}_v) = x(\mathcal{N}_v))
\end{aligned}
$$

Gornja jednakost je zbog Leme 13.1 ekvivalentna sa

$$\mathbb{P}(X(v) = x(v) \,|\, X(V \setminus v) = x(V \setminus v)) = \mathbb{P}(X(v) = x(v) \,|\, X(\mathcal{N}_v) = x(\mathcal{N}_v)) \tag{12.1}$$

(uzmite $B = \{X(v) = x(v)\}$, $C = \{X(V \setminus \widetilde{\mathcal{N}}_v) = x(V \setminus \widetilde{\mathcal{N}}_v)\}$ i $A = \{X(\mathcal{N}_v) = x(\mathcal{N}_v)\}$). Ta jednakost kaže da na distribuciju slučajne varijable na mjestu v neposredno utječu samo slučajne varijable na okolnim mjestima.

Uočimo da Markovljevo svojstvo ovisi o sustavu okolina $E = (\mathcal{N}_v : v \in V)$. Na primjer, svako slučajno polje je Markovljevo uz okoline $\mathcal{N}_v = V \setminus v$. Zanimljiva Markovljeva polja su ona sa malim okolinama.

Definicija 12.4 Lokalna karakteristika *Markovljevog polja na mjestu v je funkcija* $\pi^v : \Lambda^V \to [0,1]$ *definirana s*

$$\pi^v(x) = \mathbb{P}(X(v) = x(v) \,|\, X(\mathcal{N}_v) = x(\mathcal{N}_v)) \,. \tag{12.2}$$

Familija $(\pi^v : v \in V)$ zove se lokalna specifikacija *Markovljevog slučajnog polja.*

Promatramo slučajno polje koje se mijenja kroz vrijeme na slučajan način. Dobivamo slučajan proces $(X_n : n \geq 0)$ kod kojeg je $X_n = (X_n(v) : v \in V)$, te $X_n(v) \in \Lambda$. U vremenu $n \geq 0$ je X_n slučajno polje na V s fazama u Λ, ili ekvivalentno, slučajni element s vrijednostima u konfiguracijskom prostoru $S = \Lambda^V$. Slučajni proces $(X_n : n \geq 0)$ zovemo *dinamičkim slučajnim poljem.*

Pretpostavimo da je $(X_n : n \geq 0)$ Markovljev lanac s prostorom stanja $S = \Lambda^V$ i stacionarnom distribucijom π. Uočimo da je π vjerojatnosna distribucija na konfiguracijskom prostoru $S = \Lambda^V$. Ukoliko je $(X_n : n \geq 0)$ ireducibilan i aperiodički, tada po Teoremu 8.9 vrijedi

$$\lim_{n\to\infty} \mathbb{P}(X_n = x) = \pi(x) \,, \quad x \in \Lambda^V \,.$$

Drugim riječima, distribucija slučajnog polja X_n je za velike n približno jednaka stacionarnoj distribuciji π.

Pretpostavimo da je dana vjerojatnosna distribucija π na konfiguracijskom prostoru $S = \Lambda^V$. Cilj nam je simulirati tu distribuciju. U nekim situacijama jedina moguća metoda je takozvana MCMC metoda (Markov chain Monte Carlo). Ideja metode satoji se u tome da se konstruira ireducibilan aperiodički Markovljev lanac na prostoru $S = \Lambda^V$ čija je stacionarna distribucija upravo π. Po teoremu o graničnoj distribuciji vrijedit će gornja jednakost, pa će za dovoljno velike n slučajno polje X_n imati približno distribuciju π. Drugim riječima, π možemo simulirati vrijednostima Markovljevog lanca X_n za dovoljno velike n. Pitanje na koje preostaje odgovoriti je kako odabrati Markovljev lanac sa danom stacionarnom distribucijom. Naime, postoje razni Markovljevi lanci (t.j., razne prijelazne matrice) sa istom stacionarnom distribucijom. Cilj je odabrati onaj kod kojeg će simulacija lanca biti brza i jeftina. U tom kontekstu se prijelazni mehanizam zove *simulacijski algoritam.*

Prvi takav algoritam koji ćemo promatrati je *Gibssov algoritam.* Pretpostavimo da je dana vjerojatnosna distribucija $(q_v : v \in V)$. Prijelaz iz $X_n = x$ u $X_{n+1} = y$, $x, y \in S = \Lambda^V$, dan je na sljedeći način. Novo stanje y dobiva se iz starog stanja x mijenjanjem (ili ne) vrijednosti faze samo na jednom mjestu. Mjesto v koje će se promijeniti (ili ne) u vremenu n odabire se nezavisno od prošlosti s vjerojatnosti q_v. Nakon što je odabrano mjesto v, trenutna konfiguracija x promijeni se u y tako da bude $y(V \setminus v) = x(V \setminus v)$, a nova faza $y(v)$ na mjestu v odabire se s vjerojatnosti $\pi(y(v) \,|\, x(V \setminus v))$ (uvjetna distribucija od π uz dano $x(V \setminus v)$). Dakle, konfiguracija x promijenjena je u konfiguraciju $(y(v), x(V \setminus v))$ s vjerojatnosti $\pi(y(v) \,|\, x(V \setminus v))$, lokalna specifikacija na mjestu v. Slijedi da su prijelazne vjerojatnosti dane sa

$$\mathbb{P}(X_{n+1} = y \,|\, X_n = x) = q_v \pi(y(v) \,|\, x(V \setminus v)) 1_{\{y(V \setminus v) = x(V \setminus v)\}} . \tag{12.3}$$

Jednostavno se vidi da je dobiveni lanac X ireducibilan i aperiodički. Provjerit ćemo da vrijedi jednadžba detaljne ravnoteže, t.j.,

$$\pi(x)\mathbb{P}(X_{n+1} = y \,|\, X_n = x) = \pi(y)\mathbb{P}(X_{n+1} = x \,|\, X_n = y)$$

za sve $x, y \in S = \Lambda^V$. Po (12.3) gornja jednakost ekvivalentna je s

$$\pi(x) q_v \pi(y(v) \,|\, x(V \setminus v)) = \pi(y) q_v \pi(x(v) \,|\, x(V \setminus v)) , \tag{12.4}$$

za sve $v \in S$. Lijeva strana gore jednaka je

$$\pi(x) q_v \frac{\pi(y(v), x(V \setminus v))}{\pi(x(V \setminus v))} ,$$

gdje $\pi(x(V \setminus v)$ označava π mjeru svih konfiguracija koje na mjestima iz $V \setminus v$ imaju vrijednosti dane s x. Desna strana u (12.4) jednaka je

$$\pi(y(v), y(V \setminus v)) q_v \frac{\pi(x(v), x(V \setminus v))}{\pi(x(V \setminus v))} = \pi(y(v), x(V \setminus v)) q_v \frac{\pi(x)}{\pi(x(V \setminus v))} .$$

Dakle vrijedi jednakost u (12.4). To znači da je π stacionarna distribucija za konstruirane prijelazne vjerojatnosti.

Primjer 12.5 (Gibbsov algoritam za slučajne vektore) Neka je $\pi = (\pi(x(1), \dots, x(N))$ vjerojatnosna distribucija na $S = \Lambda^N$, $N \in \mathbb{N}$. Gibbsov algoritam može se primjeniti za simuliranje N-dimenzionalne distribucije π. Prvi korak simulacije sastoji se u odabiru koordinate $i \in \{1, 2, \dots, N\}$, a drugi u odabiru vrijednosti $y(i)$ odabrane koordinate uvjetno na trenutne vrijednosti $x(1), \dots, x(i-1), x(i+1), \dots, x(N)$ ostalih koordinata s vjerojatnosti

$$\pi(y(i) \,|\, x(1), \dots, x(i-1), x(i+1), \dots, x(N)) .$$

Sljedeći algoritam koji promatramo je *Hastingsov algoritam*. Pretpostavimo, za početak, da je S proizvoljan konačan skup čije elemente označavamo kao $i, j, \dots$. Tražimo prijelaznu matricu $P = (p_{ij} : i, j \in S)$ kojoj je dana vjerojatnosna distribucija π na S stacionarna distribucija. Preciznije, tražimo P tako da su π i P u detaljnoj ravnoteži:

$$\pi_i p_{ij} = \pi_j p_{ji} \, .$$

Rješenje tražimo u obliku

$$p_{ij} = q_{ij}\alpha_{ij} \, , \quad j \neq i \, , \tag{12.5}$$

gdje je $Q = (q_{ij} : i, j \in S)$ proizvoljna ireducibilna prijelazna matrica na S, a $\alpha_{ij} \in [0, 1]$ *vjerojatnosti prihvaćanja*. Ako je trenutno stanje i, sljedeće moguće stanje j bira se s vjerojatnosti q_{ij}. Kada je $j \neq i$, to stanje j prihvaća se s vjerojatnosti α_{ij}. U suprotnom, $j = i$. Preostaje odrediti vjerojatnosti α_{ij}. Vrlo opći oblik koji pokriva većinu poznatih MCMC algoritama dan je sa

$$\alpha_{ij} = \frac{s_{ij}}{1 + t_{ij}} \, , \tag{12.6}$$

gdje je $\Sigma = (s_{ij} : i, j \in S)$ simetrična matrica i

$$t_{ij} = \frac{\pi_i q_{ij}}{\pi_j q_{ji}} \, . \tag{12.7}$$

Pretpostavka je da je Σ odabrana tako da vrijedi $\alpha_{ij} \in [0, 1]$. Provjerimo jednadžbu detaljne ravnoteže:

$$\pi_i p_{ij} = \pi_i q_{ij} \frac{s_{ij}}{1 + \frac{\pi_i q_{ij}}{\pi_j q_{ji}}} = s_{ij} \frac{\pi_i q_{ij} \pi_j q_{ji}}{\pi_i q_{ij} + \pi_j q_{ji}} = \pi_j p_{ji} \, .$$

Primjer 12.6 (*Metropolis algoritam*) Izbor

$$s_{ij} = 1 + \min(t_{ij}, t_{ji})$$

odgovara Metropolis algortimu. Jednostavni račun daje

$$\alpha_{ij} = \min\left(1, \frac{\pi_i p_{ij}}{\pi_j p_{ji}}\right) \, ,$$

otkud

$$p_{ij} = \min\left(q_{ij}, \frac{\pi_j}{\pi_i} q_{ji}\right) \, .$$

Vratimo se na slučaj konfiguracijskog prostora $S = \Lambda^V$. Stanja su konfiguracije $x = (x(v) : v \in V)$. U ovoj situaciji pogodno je matricu Q birati tako da je $Q(x, y) \neq 0$ samo u slučaju da se konfiguracije x i y razlikuju samo na jednom mjestu: postoji $v \in V$ tako da je $x(V \setminus v) = y(V \setminus v)$. Pripadajuća prijelazna matrica P definirana s (12.5) imat će isto svojstvo, odnosno odgovarajući Markovljev lanac promjenit će fazu samo na jednom mjestu.