Skriveni Markovljevi lanci

Stjepanović, Sanja

Sveučilište J.J. Stossmayera u Osijeku
Odjel za matematiku

Sanja Stjepanović

Skriveni Markovljevi lanci

Diplomski rad

Osijek, 2020

Sveučilište J.J. Stossmayera u Osijeku
Odjel za matematiku

Sanja Stjepanović

Skriveni Markovljevi lanci

Diplomski rad

Mentor:
izv. prof. dr. sc. Nenad Šuvak

Osijek, 2020

Sadržaj

Uvod **1**

1 Slučajni procesi **2**
1.1 Definicija i osnovni pojmovi 2
1.2 Markovljevi lanci 3

2 Skriveni Markovljevi lanci **6**
2.1 Definicija i osnovni pojmovi 7
2.2 Primjena skrivenih Markovljevih modela 10

3 Osnovni problemi skrivenih Markovljevih lanaca **12**
3.1 Problem izglađivanja 15
3.2 Problem dekodiranja 21
3.3 Problem treniranja 24

Sažetak **32**

Abstract **32**

Literatura **33**

Životopis **34**

Uvod

Glavni zadatak ovog rada je upoznati nas sa skrivenim Markovljevim modelima, specifično skrivenim Markovljevim lancima, i osnovnim problemima koji se pomoću njih rješavaju. Skriveni Markovljev model je široko primijenjen alat za matematičko modeliranje te je tako postao sastavni dio u područjima znanosti poput bioinformatike (*pattern recognition*), komunikacijske teorije (*speech recognition*, *handwriting recognition*) i financijske matematike.

U prvom poglavlju ćemo se prisjetiti osnovne terminologije vezane uz slučajne procese kao i definicije Markovljevog lanca. Na temelju toga ćemo uvesti pojam i osnovna svojstva skrivenog Markovljevog modela te na nekoliko primjera ilustrirati njegovu raznovrsnu primjenu u drugom poglavlju. Zatim ćemo u trećem poglavlju navesti probleme skrivenih Markovljevih modela i izvesti njihova rješenja zajedno sa algoritmima i prikazom njihove efikasnosti na konkretnim primjerima.

1 Slučajni procesi

Intuitivno možemo zaključiti kako pojedine prirodne ili društvene pojave poput temperature zraka na određenoj lokaciji ili cijena dionice neke kompanije, ima smisla modelirati slučajnom varijablom u vremenskom trenutku t. Tada promjene i fluktuacije tih obilježja možemo promatrati kroz vrijeme kako bi došli do određenih informacija, a takvim pristupom se upravo koristimo slučajnim procesima.

U ovom poglavlju ćemo se najprije upoznati sa osnovnim pojmovima i terminima koje ćemo upotrebljavati u našem radu i tako stvoriti bazu za korištenje naprednijih alata za modeliranje kao što su to skriveni Markovljevi lanci.

1.1 Definicija i osnovni pojmovi

Slučajni procesi se koriste za modeliranje pojava koje se naizgled nasumično mijenjaju s vremenom. Nakon što smo se upoznali s motivacijom za takav matematički objekt možemo navesti njegovu definiciju.

Definicija 1.1. *Slučajni proces je familija slučajnih varijabli* $(X_t, t \in T)$ *na istom vjerojatnosnom prostoru* $(\Omega, \mathcal{F}, P)$*, gdje je* $T \subseteq \mathbb{R}$*.*

Iz definicije možemo zaključiti kako slučajne procese karakteriziraju dva važna skupa:

- **Skup stanja S** koji označava skup svih mogućih realizacija slučajne varijable X_t za sve $t \in T$. U slučaju da je S diskretan skup pripadni slučajni proces ćemo nazivati **lancem**, a u suprotnom će se raditi o slučajnom procesu s neprekidnim skupom stanja.
- **Skup indeksa T** čije elemente najčešće interpretiramo u terminima vremenskih trenutaka. On može biti neprekidan ili diskretan pa ćemo govoriti o slučajnim procesima u neprekidnom ili diskretnom vremenu, redom.

Naravno, biti će potrebno promatrati familiju marginalnih distribucija slučajnog procesa $(X_t, t \geq 0)$ kao npr. jednodimenzionalne marginalne distribucije $F_{X_t}(x) = P(X_t \leq x)$. U nastavku rada ćemo izvoditi formule za takve konačnedimenzionalne distribucije određenih procesa.

Napomena 1.1. Važno je napomenuti kako nije dovoljno poznavati samo jednodimenzionalne marginalne distribucije slučajnog procesa za donošenje zaključaka, već je za potpunu karakterizaciju procesa potrebno znati sve njegove konačnodimenzionalne distribucije, tj. distibucije slučajnih vektora $(X_{t_1}, \ldots, X_{t_n})$ za $n \in \mathbb{N}$ i bilo koje $t_1, \ldots, t_n \in T$.

1.2 Markovljevi lanci

Uzmimo proizvoljan slučajni proces u diskretnom vremenu s diskretnim skupom stanja, tj. lanac $\mathbb{X} = (X_n, n \in \mathbb{N}_0)$ s vrijednostima u S. Kažemo da slučajni proces $\mathbb{X}$ zadovoljava **Markovljevo svojstvo** ako vrijedi:

$$P(X_{n+1} = x_{n+1}|X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_1 = x_1, X_0 = x_0) = P(X_{n+1} = x_{n+1}|X_n = x_n)$$

za svaki $n \in \mathbb{N}_0$ i za sve $x_0, \ldots, x_{n-1}, x_n \in S$ za koje su ove uvjetne vjerojatnosti dobro definirane. Ukoliko interpretiramo trenutak $n + 1$ kao neposrednu budućnost, n kao sadašnjost i $n - 1, \ldots, 1, 0$ kao prošlost našeg slučajnog procesa $\mathbb{X}$, možemo objasniti Markovljevo svojstvo kao nezavisnost neposredne buduće vrijednosti procesa o prošlosti ako je dana njegova sadašnja vrijednost. Iz Markovljevog svojstva prirodno slijedi i definicija Markovljevog lanca.

Definicija 1.2. *Neka je S diskretan skup. Slučajan proces $\mathbb{X} = (X_n, n \in \mathbb{N}_0)$ na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, P)$ s vrijednostima u S je **Markovljev lanac** ako vrijedi*

$$P(X_{n+1} = j|X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0) = P(X_{n+1} = j|X_n = i)$$

za svaki $n \in \mathbb{N}_0$ i za sve $i_0, \ldots, i_{n-1}, i, j \in S$ za koje su obje gornje uvjetne vjerojatnosti dobro definirane.

Markovljevi lanci su visoko zastupljeni u stohastičkom modeliranju i to s dobrim razlogom. S jedne strane, mnogi modeli prirodno zadovoljavaju Markovljevo svojstvo. Na primjer, osnovni zakoni fizike utvrđuju kako se gibanje čestice u (malom) vremenskom intervalu određuje samo pomoću sadašnjeg položaja i brzine, a ne pomoću prošlih kretnji. S druge strane, jednostavna struktura Markovljevih lanaca nam znatno pojednostavljuje razvoj matematičkih tehnika i algoritama koji bi bili nemogući za napraviti bez postojanja Markovljevog svojstva. Stoga je u praksi vrlo poželjno graditi stohastičke modele koji ga posjeduju.

Uz Markovljeve lance su usko vezane određene vjerojatnosti koje, ako su zadane, u potpunosti mogu odrediti Markovljev model. Za stanja $i, j \in S$, $0 \leq s \leq t$, definiramo funkciju prijelaznih vjerojatnosti Markovljevog lanca

$$p(i, s; t, j) = P(X_t = j|X_s = i).$$

Pomoću takve funkcije računamo koja je vjerojatnost da u trenutku t Markovljev lanac ima vrijednost j ako znamo da je u trenutku s imao vrijednost i. Specijalno, ako se radi o 1-koračnim prijelaznim vjerojatnostima računamo koja je vjerojatnost da je Markovljev lanac prešao sa vrijednosti i u trenutku n u vrijednost j u trenutku $n + 1$, tj.

$$p(i, n; n + 1, j) = P(X_{n+1} = j|X_n = i).$$

Nadalje, uvodimo distribuciju slučajne varijable X_0 kojom modeliramo početno stanje Markovljevog lanca $\lambda = [\lambda_i]_{i\in S}$, gdje je $P(X_0 = i) = \lambda_i,\ i \in S$. Poznavanjem početne distribucije Markovljevog lanca i svih njegovih 1-koračnih prijelaznih vjerojatnosti je naš model sasvim određen. O tome nam govori sljedeći teorem.

Teorem 1.1. *Markovljev lanac je u potpunosti određen poznavanjem distribucije λ slučajne varijable X_0 i funkcije prijelaznih vjerojatnosti u 1 koraku, tj. funkcije*

$$p(i, n; n+1, j) = P(X_{n+1} = j | X_n = i).$$

Dokaz: Prisjetimo se *Napomene 1.1* iz prošlog poglavlja koja kaže kako poznavanje svih konačnodimenzionalnih distribucija znači poznavanje čitavog procesa, tj. Markovljev lanac će biti potpuno određen ako znamo njegove konačnodimenzionalne distribucije. Kako bi dokazali teorem moramo pokazati da se konačnodimenzionalne distribucije mogu raspisati pomoću λ i prijelaznih vjerojatnosti u 1 koraku. Neka je $n \in \mathbb{N}$ proizvoljan:

$$\begin{aligned}
P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) &= P(X_n = i_n | X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \\
&\quad \cdot P(X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \\
&\overset{MS}{=} P(X_n = i_n | X_{n-1} = i_{n-1}) \\
&\quad \cdot P(X_{n-1} = i_{n-1} | X_{n-2} = i_{n-2}, \ldots, X_0 = i_0) \\
&\quad \cdot P(X_{n-2} = i_{n-2}, \ldots, X_0 = i_0) \\
&= \cdots = p(i_{n-1}, n-1; n, i_n) \cdot p(i_{n-2}, n-2; n-1, i_{n-1}) \\
&\quad \cdot \ldots P(X_1 = i_1 | X_0 = i_0) P(X_0 = i_0) \\
&= p(i_{n-1}, n-1; n, i_n) \ldots p(i_1, 1; 0, i_0) \lambda_0.
\end{aligned}$$

□

Posebno, ako za funkciju prijelaznih vjerojatnosti vrijedi

$$p(i, n; n+1, j) = p(i, m; m+1, j), \quad \forall n, m \in \mathbb{N},$$

onda kažemo da je Markovljev lanac $\mathbb{X} = (X_n, n \in \mathbb{N}_0)$ **homogen**. U tom slučaju, funkcija prijelaznih vjerojatnosti ne ovisi o određenom trenutku nego samo o stanju iz kojeg polazi Markovljev lanac i i stanju u koji on dolazi j, pa ju možemo zapisati na jednostavniji način:

$$p(i, n; n+1, j) = p_{ij}, \quad i, j \in S.$$

Ovakva oznaka dopušta elegantan zapis prijelaznih vjerojatnosti u obliku matrice

$$\Pi = \begin{bmatrix} p_{11} & p_{12} & \cdots \\ p_{21} & p_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

koju ćemo zvati matricom 1-koračnih prijelaznih vjerojatnosti homogenog Markovljevog lanca. Takva matrica prijelaznih 1-koračnih vjerojatnosti $\Pi = [p_{ij}]_{i,j \in S}$ ima sljedeća svojstva:

1. $p_{ij} = P(X_{n+1} = j | X_n = i) \geq 0, \quad \forall i, j \in S, \forall n \in \mathbb{N}_0.$

2. $\sum_{j \in S} p_{ij} = \sum_{j \in S} P(X_{n+1} = j | X_n = i) = \sum_{j \in S} P_i(X_{n+1} = j).$

   Ako na takav način definiramo novu vjerojatnost P_i na manjem skupu $\Omega' = \{\omega \in \Omega : X_n(\omega) = i\} \subseteq \Omega$ imamo sljedeće

$$P_i \left(\bigcup_{j \in S} \{X_{n+1} = j\} \right) = P_i(\Omega') = 1.$$

   Drugim riječima, suma vjerojatnosti iz istog retka matrice Π je jednaka 1.

Ova dva svojstva čine matricu Π stohastičkom matricom, a homogen Markovljev lanac određen matricom 1-koračnih prijelaznih vjerojatnosti Π i početnom distribucijom λ ćemo označavati kao (Π, λ)-Markovljev lanac. Iz dokaza *Teorema 1.1* smo dobili formulu za izračun konačnodimenzionalnih distribucija Markovljevog lanca (Π, λ) koja glasi

$$P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \lambda_{i_0} \prod_{k=1}^{n} p_{i_k i_{k-1}}.$$

Specijalno, sada možemo odrediti jednodimenzionalnu marginalnu distribuciju od X_n za $n \in \mathbb{N}_0$ na temelju poznate početne distribucije $\lambda = [\lambda_i]_{i \in S}$ i matrice $\Pi = [p_{ij}]_{i,j \in S}$. Kako bismo došli do općenite formule, krenimo od početog trenutka $n = 0$:

$$n = 0: \; P(X_0 = i) = \lambda_i$$

$$\begin{aligned} n = 1: \; P(X_1 = i) &= P(X_1 = i, X_0 \in S) = \sum_{i_o \in S} P(X_1 = i, X_0 = i_0) \\ &= \sum_{i_0 \in S} P(X_1 = i | X_0 = i_0) P(X_0 = i_0) = \sum_{i_0 \in S} \lambda_{i_0} p_{i_0 i} = \sum_{i_0 \in S} \lambda_{i_0} (\Pi)_{i_0 i} = (\lambda \Pi)_i \,, \quad i \in S \end{aligned}$$

$$\begin{aligned} n = 2: \; P(X_2 = i) &= P(X_2 = i, X_1 \in S, X_0 \in S) = \sum_{i_1 \in S} \sum_{i_0 \in S} P(X_2 = i, X_1 = i_1, X_0 = i_0) \\ &= \sum_{i_1 \in S} \sum_{i_0 \in S} P(X_2 = i | X_1 = i_1, X_0 = i_0) P(X_1 = i_1, X_0 = i_0) \\ &\overset{MS}{=} \sum_{i_1 \in S} \sum_{i_0 \in S} P(X_2 = i | X_1 = i_1) P(X_1 = i_1 | X_0 = i_0) P(X_0 = i_0) \\ &= \sum_{i_0 \in S} \lambda_{i_0} \left(\sum_{i_1 \in S} p_{i_0 i_1} p_{i_1 i} \right) = \sum_{i_0 \in S} \lambda_{i_0} \left(\Pi^2\right)_{i_0 i} = \left(\lambda \Pi^2\right)_i \,, \quad i \in S. \end{aligned}$$

Možemo vidjeti da općenito onda vrijedi:

$$P(X_n = i) = (\lambda \Pi^n)_i \,, \quad i \in S.$$

Još što možemo odrediti pomoću matrice prijelaznih 1-koračnih vjerojatnosti jest prijelazna vjerojatnost u više koraka, npr. u m koraka.

$$
\begin{aligned}
P(X_{n+m}=j|X_n=i) = P(X_m=j|X_0=i) &= \frac{P(X_m=j, X_0=i)}{P(X_0=i)} \\
&= \frac{\sum_{i_1\in S}\cdots\sum_{i_{m-1}\in S}\lambda_i p_{ii_1}\dots p_{i_{m-1}j}}{\lambda_i} \\
&= \sum_{i_1\in S}\cdots\sum_{i_{m-1}\in S} p_{ii_1}\dots p_{i_{m-1}j} = (\Pi^m)_{ij}\,.
\end{aligned}
$$

Međutim, u jednom i u drugom slučaju se javlja problem kompleksnosti izračuna matrice $\Pi^n, n\in\mathbb{N}_0$, a više o rješavanju takvog problema možete pronaći u [4].

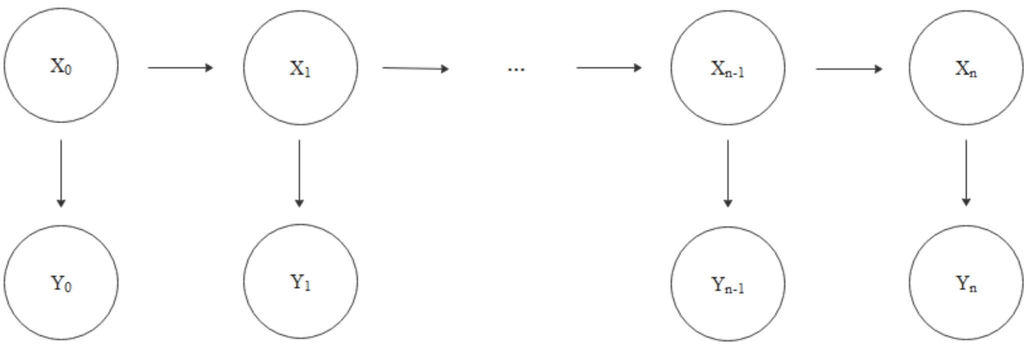
2 Skriveni Markovljevi lanci

Iako Markovljevi lanci imaju široku primjenu, nije sasvim realistično pretpostaviti da pri modeliranju možemo direktno promatrati promjene stanja Markovljevog lanca, već se češće nalazimo u situaciji da možemo promatrati mijenjanje određenih čimbenika koji su pod utjecajem kretanja procesa koji nas zanima. Ovakva problematika je upravo motivirala uvođenje skrivenih Markovljevih modela jer oni uzimaju u obzir pozadinski proces čija nas kretanja zanimaju ali su nam skrivena.

Općenito se na skriveni Markovljev model može gledati kao na Markovljev proces koji je podijeljen na dvije komponente:

- **Skriveni proces** (*hidden process*) kojeg čini niz slučajnih varijabli $(X_t, t\in T)$ sa vrijednostima u S čije promjene stanja su nama skrivene.
- **Opservacijski proces** (*observable process*) kojeg čini niz slučajnih varijabli $(Y_t, t\in T)$ sa vrijednostima u O čije vrijednosti i promjene možemo promatrati.

Odnosno, pod skrivenim Markovljevim procesom smatramo uređeni par procesa $((X_t, Y_t), t\in T)$ na prostoru stanja $S\times O$, gdje pretpostavljamo da imamo sredstva za promatranje Y_t, ali ne i X_t.



Slika 1: Shematski prikaz skrivenog Markovljevog procesa

U svrhu boljeg razumijevanja ovakvog koncepta možemo uzeti primjer iz komunikacijske teorije gdje bi X_t predstavljao slučajni signal koji se prenosi putem komunikacijskog kanala. Kako u kanalu dolazi do smetnji, prijemnik dobiva oštećenu verziju izvornog signala Y_t, a cilj mu je onda rekonstruirati što je moguće bolje izvorni signal. S druge strane, možda nam je proces Y_t u konačnici zanimljiv, dok X_t samo predstavlja utjecaj nekih neupadljivih vanjskih čimbenika na Y_t. Na primjer, Y_t može predstavljati tržišnu cijenu dionica, a X_t nezamijećeni ekonomski faktor koji utječe na fluktuacije cijene tih dionica. Mi bismo htjeli modelirati promatrane oscilacije cijena dionica, ne proces neprimjetnih faktora, ali njihovim uključivanjem bi mogli biti u mogućnosti izgraditi model koji vjernije odražava statistička svojstva promatranih cijena dionica.

U ovom poglavlju ćemo definirati što je točno skriveni Markovljev model i upoznati se sa njegovim osnovnim svojstvima. Također, navest ćemo nekoliko primjera iz raznih znanstvenih područja gdje se skriveni Markovljevi procesi primjenjuju u rješavanju problema.

2.1 Definicija i osnovni pojmovi

U ostatku rada ćemo uvijek pretpostaviti da su nam skup indeksa T i skupovi stanja S i O diskretni, tj. da ćemo raditi isključivo sa Markovljevim i skrivenim Markovljevim lancima. Svi rezultati koje budemo izveli u ovom poglavlju se mogu proširiti na procese sa neprekidnim skupom stanja i u neprekidnom vremenu, ali radi jednostavnosti i implementacije kasnijih algoritama ćemo se u ovom radu suziti na diskretne skupove.

Na samom početku se trebamo osvrnuti na postupak modeliranja, tj. na pretpostavke prema kojima smatramo da se skriven Markovljev lanac odvija i svojstva koja iz njih proizlaze.

- Prvo pravilo je da skriveni lanac $(X_n, n \in \mathbb{N}_0)$ prelazi iz trenutnog stanja u sljedeće stanje prema nekoj raspodjeli vjerojatnosti koja ovisi samo o trenutnom stanju, tj.

  $$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0) = P(X_{n+1} = j | X_n = i)$$

  za svaki $n \in \mathbb{N}_0$ i za sve $i_0, \ldots, i_{n-1}, i, j \in S$ za koje su ove uvjetne vjerojatnosti dobro definirane. Odnosno, tražit ćemo da proces $(X_n, n \in \mathbb{N}_0)$ bude Markovljev lanac.

- Drugo pravilo je da skriveni lanac pri svakom prijelazu u novo stanje emitira opservacijsko stanje Y_n čija raspodjela ovisi samo o trenutnom skrivenom stanju X_n, tj.

  $$P(Y_n = j | X_n = i, Y_{n-1} = j_{n-1}, X_{n-1} = i_{n-1}, \ldots, Y_0 = j_0, X_0 = i_0) = P(Y_n = j | X_n = i).$$

  Odnosno, trenutna opservacija je neovisna o prijašnjim opservacijama i prošlim skrivenim stanjima.

Napomena 2.1. Iako tražimo da $(X_n, n \in \mathbb{N}_0)$ bude Markovljev lanac, opservacijski lanac $(Y_n, n \in \mathbb{N}_0)$ sam po sebi ne mora biti Markovljev, čak u većini slučajeva on to i nije. Zbog toga se skriveni Markovljevi lanci mogu koristiti za modeliranje ponašanja koji ne zadovoljavaju Markovljevo svojstvo (npr. cijene dionica), zadržavajući pri tome većinu matematičkih rezultata koji proizlaze iz tog svojstva.

Sada kada znamo pravila po kojima se odvija naš proces, analogno kao kod Markovljevih lanca, možemo navesti pojmove koji određuju skriveni Markovljev model. Skriveni Markovljevi modeli definiraju se pomoću sljedećih pet parametara:

1. N predstavlja kardinalni broj skupa stanja S za skriveni lanac $(X_n, n \in \mathbb{N}_0)$, pa se njegov skup stanja može zapisati na sljedeći način

$$S = \{i_1, i_2, \dots, i_{N-1}, i_N\}.$$

2. M predstavlja kardinalni broj skupa stanja O za opservacijski lanac, tj. za $(Y_n, n \in \mathbb{N}_0)$ pa se skup stanja O može zapisati na sljedeći način

$$O = \{j_1, j_2, \dots, j_{M-1}, j_M\}.$$

3. Π je $N \times N$ matrica prijelaznih 1-koračnih vjerojatnosti skrivenog lanca $(X_n, n \in \mathbb{N}_0)$ definirana na sljedeći način

$$\Pi = [p_{ij}]_{i,j \in S}, \quad p_{ij} = P(X_{n+1} = j | X_n = i).$$

4. Φ je $N \times M$ matrica emitiranih vjerojatnosti, odnosno vjerojatnosti da neko skriveno stanje $i \in S$ generira određeno opservacijsko stanje $j \in O$.

$$\Phi = [q_{ij}]_{i \in S, j \in O}, \quad q_{ij} = P(Y_n = j | X_n = i).$$

Primijetimo da je Φ također stohastička matrica poput matrice prijelaznih 1-koračnih vjerojatnosti Π jer zadovoljava sljedeća svojstva:

(a) $q_{ij} = P(Y_n = j | X_n = i) \geq 0, \quad \forall i \in S, \forall j \in O, \forall n \in \mathbb{N}_0.$

(b) $\displaystyle\sum_{j \in O} q_{ij} = \sum_{j \in O} P(Y_n = j | X_n = i) = \sum_{j \in O} P_i(Y_n = j).$

Ako na takav način definiramo novu vjerojatnost P_i na manjem skupu $\Omega' = \{\omega \in \Omega : X_n(\omega) = i\} \subseteq \Omega$ imamo sljedeće

$$P_i\left(\bigcup_{j \in O} \{Y_n = j\}\right) = P_i(\Omega') = 1.$$

5. λ označava distribuciju slučajne varijable X_0 kojom modeliramo početno stanje skrivenog procesa

$$\lambda = [\lambda_i]_{i \in S}, \quad \lambda_i = P(X_0 = i).$$

Poznavanje gornjih parametara nam omogućuje uvođenje oznake za skriveni Markovljev model kao (Π, Φ, λ) i formuliranje formalne definicije skrivenog Markovljevog lanca.

Definicija 2.1. *Neka su S i O diskretni skupovi. Slučajni proces $((X_n, Y_n), n \in \mathbb{N}_0)$ na vjerojatnosnom prostoru (Ω, F, P) sa vrijednostima u $S \times O$ se naziva skriveni Markovljev lanac ako postoji matrica prijelaznih vjerojatnosti $\Pi = [p_{ij}]_{i,j \in S}$ i matrica emitiranih vjerojatnosti $\Phi = [q_{ij}]_{i \in S, j \in O}$ takve da vrijedi*

$$\begin{aligned} P(X_{n+1} = i, Y_{n+1} = j | X_n = i_n, Y_n = j_n, \ldots, X_0 = i_0, Y_0 = j_0) \\ &= P(X_{n+1} = i | X_n = i_n) \cdot P(Y_{n+1} = j | X_{n+1} = i) \\ &= p_{i_n i} \cdot q_{ij}. \end{aligned} \tag{2.1}$$

za svaki $n \in \mathbb{N}_0$ i za sve $i, i_n, \ldots, i_0 \in S$ i $j, j_n, \ldots, j_0 \in O$ za koje su gornje uvjetne vjerojatnosti dobro definirane.

Uočimo da uvjet iz definicije, koji je ekvivalent Markovljevog svojstva za skriveni Markovljev lanac, direktno slijedi iz pretpostavki skrivenog Markovljevog procesa, tj. iz činjenice da $(X_n, n \in \mathbb{N}_0)$ zadovoljava Markovljevo svojstvo te iz nezavisnosti trenutne opservacije o prošlosti:

$$\begin{aligned} P(X_{n+1} = i, &Y_{n+1} = j | X_n = i_n, Y_n = j_n, \ldots, X_0 = i_0, Y_0 = j_0) \\ &= P(Y_{n+1} = j | X_{n+1} = i, X_n = i_n, Y_n = j_n, \ldots, X_0 = i_0, Y_0 = j_0) \\ &\quad \cdot P(X_{n+1} = i | X_n = i_n, Y_n = j_n, \ldots, X_0 = i_0, Y_0 = j_0) \\ &= P(Y_{n+1} = j | X_{n+1} = i) \cdot P(X_{n+1} = i | X_n = i_n). \end{aligned}$$

Iz te jednakosti vidimo da vjerojatnost da X_{n+1} i Y_{n+1} poprime vrijednosti i i j možemo računati kao umnožak vjerojatnosti prijelaza iz trenutnog stanja skrivenog procesa $X_n = i_n$ u stanje i i vjerojatnosti da će ono u trenutku $n+1$ emitirati opservaciju $Y_{n+1} = j$. Za razliku od Markovljevog lanca, neposredna budućnost ne ovisi samo o sadašnjosti već i o uzročno posljedičnoj vezi između skrivenog stanja i opservacije.

Također, možemo zaključiti da za trenutak $n = 0$ vrijedi sljedeće:

$$P(X_0 = i, Y_0 = j) = P(Y_0 = j | X_0 = i) \cdot P(X_0 = i) = q_{ij} \cdot \lambda_i.$$

Analogno, kao kod Markovljevih lanaca, lako se može izvesti formula za računanje konačnodimenzionalne distribucije skrivenog Markovljevog lanca pomoću parametara (Π, Φ, λ) koji ga određuju:

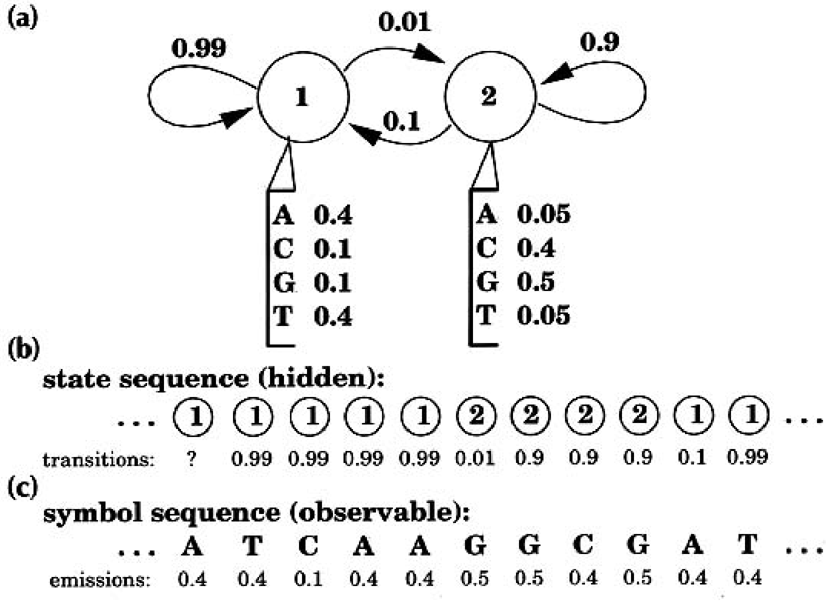
$$\begin{aligned} &P(X_n = i_n, Y_n = j_n, \ldots, X_0 = i_0, Y_0 = j_0) \\ &\quad = P(X_n = i_n, Y_n = j_n | X_{n-1} = i_{n-1}, Y_{n-1} = j_{n-1}, \ldots, X_0 = i_0, Y_0 = j_0) \\ &\qquad \cdot P(X_{n-1} = i_{n-1}, Y_{n-1} = j_{n-1}, \ldots, X_0 = i_0, Y_0 = j_0) \\ &\quad \overset{(2.1)}{=} p_{i_{n-1} i_n} q_{i_n j_n} P(X_{n-1} = i_{n-1}, Y_{n-1} = j_{n-1} | X_{n-2} = i_{n-2}, Y_{n-2} = j_{n-2}, \ldots, X_0 = i_0, Y_0 = j_0) \\ &\qquad \cdot P(X_{n-2} = i_{n-2}, Y_{n-2} = j_{n-2}, \ldots, X_0 = i_0, Y_0 = j_0) \\ &\quad = \cdots = p_{i_{n-1} i_n} q_{i_n j_n} p_{i_{n-2} i_{n-1}} q_{i_{n-1} j_{n-1}} \ldots p_{i_0 i_1} q_{i_1 j_1} P(X_0 = i_0, Y_0 = j_0) \\ &\quad = q_{i_0 j_0} \lambda_{i_0} \prod_{k=1}^{n} p_{i_{k-1} i_k} q_{i_k j_k}. \end{aligned}$$

2.2 Primjena skrivenih Markovljevih modela

U ovom poglavlju ćemo pokušati demonstrirati široku primjenu skrivenih Markovljevih modela pomoću konkretnih primjera.

Primjer 2.1. (Bioinformatika) Genetske informacije su kodirane u DNK, dugački polimer koji se nalazi u gotovo svim živim sustavima se sastoji od niza osnovnih parova A, C, G, T (tj. genetski kod je vrlo dugačka riječ sačinjena od četiri slova). Kako genetski kod igra važnu ulogu u unutarnjem djelovanju žive stanice, dešifriranje ove informacije trebao bi dovesti do značajnog znanstvenog i medicinskoga napretka.

Međutim, tumačenje genetskih podataka je vrlo netrivijalan zadatak. Na primjer, susrećemo se sa sljedećim problemom. Genetski kod se sastoji od regija kodiranja i nekodiranja. Regije koje kodiraju izravno kodiraju strukturu proteina koji nastaju u stanici zamršenim procesom koji započinje prepisivanjem odgovarajućeg dijela lanca DNK. Nekodirajuće regije, međutim, ne kodiraju izravno molekularnu strukturu, ali mogu služiti za regulaciju kada i koliko će se bjelančevina proizvesti. Da bi se mogao interpretirati genetski kod, prvo moramo razdvojiti područja kodiranja i nekodiranja. Nažalost, ne postoji jasan znak kada započinje područje kodiranja ili završava tako da se obično ta identifikacija mora obaviti statističkim metodama.



Slika 2: Shematski prikaz skrivenog Markovljevog lanca koji modelira regije kodiranja DNK

Korištenje skrivenih Markovljevih modela bilo je izuzetno uspješno u pristupu rješavanja ovog problema. Najjednostavniji pristup je sljedeći. Vremenski parametar n predstavlja položaj duž lanca DNK. Skriveni proces $(X_n, n \in \mathbb{N})$ je Markovljev lanac sa stanjima $S = \{1, 2\}$, gdje bi 1 označavao da se trenutni par nalazi u području kodiranja, a 2 da se nalazi u nekodirajućem području. Opservacijski lanac $(Y_n, n \in \mathbb{N})$ bi imao skup stanja

$O = \{A, C, G, T\}$, tako da Y_n predstavlja tip n-tog para. Prijelazne i emitirane vjerojatnosti matrica Π i Φ procjenjuju se iz podataka DNK niza. Nakon što to učinimo, možemo pokrenuti postupak obrnute procjene za određivanje područja DNK niza koja kodiraju ili nekodiraju. Taj je pristup prilično naivan, ali već daje iznenađujuće dobre rezultate.

Primjer 2.2. (Financijsko tržište) Najpoznatiji način modeliranja tržišne cijene dionica S_k u trenutku k je Black-Scholes model

$$S_k = e^{\mu - \sigma^2/2 + \sigma Z_k} S_{k-1},$$

gdje je $Z_k \sim \mathcal{N}(0,1)$ IID niz, $\sigma \in \mathbb{R}$ volatilnost i $\mu \in \mathbb{R}$ stopa povrata. Visoka volatilnost znači da cijene dionica pokazuju velike slučajne fluktuacije, dok visoka stopa povrata znači da se vrijednost dionica u prosjeku brzo povećava.

Unutar kraćih vremenskih razdoblja ovaj jednostavan model daje poprilično dobre rezultate jer se u takvim slučajevima ispunjavaju pretpostavke normalnosti i nekoreliranosti log povrata na kojima se zasniva ovakav model. Međutim, na duže periode te pretpostavke nisu zadovoljene i pokazuje se da se cijene dionica najčešće ne ponašaju kao Markovljev proces. Intuitivno se to moglo i očekivati jer μ i σ ovise o raznim vanjskim (ekonomskim, političkim, ekološkim) čimbenicima koji nisu konstantni na duljim vremenskim periodima. Kako bi uključili ovu pretpostavku zavisnosti i nestabilnosti uvodimo Markovljev proces $(X_k, k \geq 0)$ (koji je nezavisan s Z_k) i postavimo

$$S_k = e^{\mu(X_k) - \sigma(X_k)^2/2 + \sigma(X_k) Z_k} S_{k-1}.$$

U slučaju da želimo promatrati kretanje log-povrata $Y_k = \log(\frac{S_k}{S_{k-1}})$ onda bi proces $((X_k, Y_k), k \geq 0)$ bio skriveni Markovljev proces. Ukoliko se na odgovarajući način odabere X_k, može se dobiti model za kretanje cijene dionica koji je realniji od Black-Scholes modela.

Primjer 2.3. (Praćenje objekta) Još jedan problem koji je interesantan za spomenuti je problem praćenja pokretnog objekta pomoću nesigurnih senzora. Promatramo objekt koji se nasumično kreće u ravnini, njegove koordinate položaja mogu se mijenjati prema sljedećim formulama:

$$X_k^1 = X_{k-1}^1 + Z_k^1 + \alpha^1(U_k) \qquad X_k^2 = X_{k-1}^2 + Z_k^2 + \alpha^2(U_k),$$

gdje je $\alpha(U_k)$ brzina promatranog objekta (eventualno neki kontrolirani vanjski proces), dok je $(Z_k)_{k>1}$ IID niz koji odgovaraju slučajnim promjenama brzine. Uzmemo li neki Markovljev lanac za U_k može se modelirati kretanje objekta koji nas pokušava zbuniti nasumičnim prebacivanjem brzine u različitim unaprijed postavljenim smjerovima (npr. praćenje položaja neprijateljskoga aviona).

Signale koje dobivamo pomoću senzora bi bile naše opservacijske vrijednosti Y_k koje se najčešće računaju kao

$$Y_k = h((X_k^1, X_k^2)) + W_k,$$

gdje je $(W_k)_{k>0}$ IID niz i $h(\cdot)$ neka funkcija opservacije. Funkcija $h(\cdot)$ može biti i nelinearna. Na primjer, ako pratimo lokaciju objekta s fiksnog položaja na zemlji, možemo si zamisliti situaciju u kojoj možemo samo promatrati smjer vidne linije između senzora i objekta, a ne njihovu udaljenost. U ovom slučaju se uzima $h(X_k^1, X_k^2) = \arctan(X_k^2/X_k^1)$. Cilj je što je moguće bolje pratiti položaj objekta pomoću poznavanja njegovih prijašnjih položaja i podataka dobivenih putem senzora.

Primjer 2.4. (Seizmologija) Potres je iznenadna i kratkotrajna vibracija tla uzrokovana magmatskom aktivnošću (vulkanski potres), tektonskim poremećajima (tektonski potres) u litosferi i drugim sličnim prirodnim ili pak ljudskim djelovanjem. Potresi se često događaju u klasterima. Konvencionalno, najveći potres u skupu naziva se glavni potres, oni koji ga prethode nazivaju se *foreshocks*, a oni nakon njega nazivaju se *aftershocks* i stoga ih je moguće identificirati tek nakon što se cijeli klaster smirio. Razumijevanje potresnih klastera nam pruža priliku predviđanja sljedećeg glavnog potresa te eventualno pružanje informacija važnih za analizu opasnosti. Prilikom svake analize podataka o potresima, često je prvi korak upravo klasifikacija potresa na pripadnike klastera ili kao pojedinačni potres. Tek su se nedavno počeli skriveni Markovljevi modeli primjenjivati za rješavanje takvog problema.

U ovom slučaju bi skriveni proces $(X_n, n \in \mathbb{N}_0)$ predstavljao klasu promatranog potresa koje se mogu proizvoljno izabrati. Na primjer, mogu se uzeti dva moguća skrivena stanja koja bi označavala je li je potres glavni ili samo dio klastera, dok se sa tri stanja mogu rasporediti na glavni potres, *foreshock* i *aftershock*. Opservacijski proces $(Y_n, n \in \mathbb{N}_0)$ bi predstavljao čimbenike zabilježenog potresa koji direktno ovise o klasi tog potresa, npr. njihova jačina po Richteru, geografski položaj, dubina epicentra i sl. Detaljnije rješenje ovog problema možete pronaći u [7].

3 Osnovni problemi skrivenih Markovljevih lanaca

U prijašnjem poglavlju smo definirali pojmove koji određuju skriveni Markovljevi model te izveli formulu za računanje njegovih konačnodimenzionalnih distribucija, međutim zbog primjene nam se nameću druga pitanja koja još trebamo proučiti. Postoje tri osnovna problema koja su karakteristična za skrivene Markovljeve lance.

1. Pretpostavimo da nam je zadan skriveni Markovljev model (Π, Φ, λ) i niz opservacijskih stanja $o = [j_0, \ldots, j_n]$. Potrebno je izračunati posteriorne marginalne distribucije skrivenih stanja X_k, $0 \le k \le n$, uz dane podatke, tj. tražimo vjerojatnost

$$P(X_k = i | Y_0 = j_0, \ldots, Y_n = j_n), \quad \forall i \in S.$$

   Problem traženja takvih distribucija se naziva **problem izglađivanja** (*smoothing*).

2. Uz dani slijed opservacija $o = [j_0, \ldots, j_n]$ i skriveni Markovljev model (Π, Φ, λ) potrebno je odrediti slijed skrivenih stanja $s = [i_0, \ldots, i_n]$ koji bi s najvećom vjero-

jatnošću emitirali dani niz opservacija. Ovakav problem je poznat kao **problem dekodiranja** (*decoding*). Tražena vjerojatnost koju je potrebno maksimizrati po skrivenim stanjima s onda glasi

$$P(X_0 = i_0, \dots, X_n = i_n | Y_0 = j_0, \dots, Y_n = j_n).$$

3. Pretpostavimo da nam matrice emitiranih i prijelaznih vjerojatnost, te početna distribucija nisu poznate i neka je dan skup opservacija $o = [j_0, \dots, j_n]$. Potrebno je odrediti skriveni Markovljev model (Π, Φ, λ) koji maksimizira vjerojatnost emitiranja danog skupa opservacija. Ovakav problem se zove **problem treniranja** (*training*).

Za svaki od navedenih problema ćemo izvesti rješenje i navesti algoritam pomoću kojega se računaju tražene vjerojatnosti. Kako bi si lakše mogli predočiti i interpretirati probleme navesti ćemo primjer kojim ćemo se služiti kroz čitavo poglavlje za demonstraciju.

Primjer 3.1. Neka imamo sljedeću situaciju: kćerka i otac žive razdvojeno u daleko udaljenim gradovima i svaki dan bi se čuli putem telefona. Kćerka zna da se očevo raspoloženje mijenja s obzirom na vrijeme i to na sljedeći način:

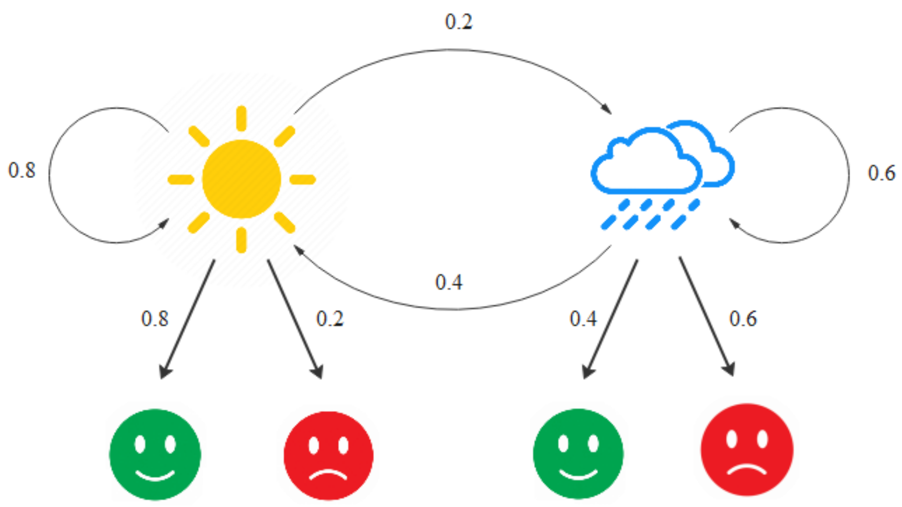
- ako je otac dobre volje vjerojatnost da je vrijeme sunčano je 0.8, a da je kišovito 0.4
- ako je otac loše volje vjerojatnost da je vrijeme sunčano iznosi 0.2, a da je kišovito 0.6.

Iz iskustva, kćerka zna na koji način se vrijeme mijenja iz dana u dan te da sutrašnje vrijeme ovisi samo o današnjem:

- ako je danas sunčano, vjerojatnost da će sutra biti sunčano je 0.8, a da je kišovito 0.2
- ako je danas kišovito, vjerojatnost da će sutra biti sunčano je 0.4, a da je kišovito 0.6.

Tijekom njihovih razgovora, otac nikad ne spominje kakvo je vrijeme, ali kćerka prepoznaje kakve je volje i svakodnevno je bilježila podatke o njegovom raspoloženju tijekom proteklih 6 dana. Kćerka bi htjela na temelju svih informacija saznati stanje vremena. Ovakav problem se može modelirati skrivenim Markovljevim lancem.

U ovom slučaju nam je skriveni lanac $(X_t, t \in \mathbb{N}_0)$ lanac koji modelira vremenske prilike na dan t koje imaju dva moguća stanja $S = \{s, k\}$ $(N = 2)$, gdje je s oznaka za sunčano vrijeme, a k oznaka za kišovito vrijeme. Opservacijski lanac $(Y_t, t \in \mathbb{N}_0)$ nam predstavlja očevo raspoloženje na dan t koji također ima dva moguća stanja $O = \{d, l\}$ $(M = 2)$, gdje je d oznaka za dobro raspoloženje, a l oznaka za loše raspoloženje. Također imamo i niz opservacijskih stanja $o = [d, d, l, l, l, d]$.



Slika 3: Shematski prikaz skrivenog Markovljevog lanca iz *Primjera 3.1*

Iz opisa primjera vidimo da su nam zadani svi parametri koji određuju skriveni Markovljev model.

$$\Pi = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix} \quad \Phi = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix} \quad \lambda = \left[\frac{2}{3}, \frac{1}{3}\right].$$

Prateći osnovne probleme skrivenog Markovljevog modela pitamo se sljedeće:

1. Koja je vjerojatnost da je neki slučajno odabrani dan sunčan ili kišovit?
2. Ako znamo kakvog je raspoloženja otac bio u proteklih 6 dana, kakvo je bilo vrijeme?
3. U slučaju da se otac preseli u grad za koji kćerka ne zna pravilo mijenjanja vremena niti očevog raspoloženja, kako možemo samo na temelju novih opservacija o_2 izračunati te vjerojatnosti?

U naredna tri poglavlja ćemo izvesti i objasniti rekurzije pomoću kojih ćemo odgovoriti na navedena tri pitanja.

3.1 Problem izglađivanja

Upravo zbog Markovljevog svojstva se računanje nekih vjerojatnosti može svesti na rekurziju. Unutar ovog poglavlja ćemo izvesti formulu i navesti *Forward-Backward algoritam* pomoću kojeg se računa vjerojatnost

$$P(X_k = i | Y_0 = j_0, \dots, Y_n = j_n), \quad 0 \leq k \leq n,\, i \in S, \tag{3.1}$$

ako nam je poznat skriveni Markovljev model (Π, Φ, λ) i niz opservacija $o = [j_0, \dots, j_n]$. Odnosno, želimo izračunati vjerojatnost da se nalazimo u skrivenom stanju i u trenutku k uz poznate sve opservacije. Ova vjerojatnost će biti korisna i u kasnijim poglavljima kada se budemo bavili procjenom vjerojatnosti matrica Π i Φ.

Vjerojatnost (3.1) nije trivijalna za izračunati, pogotovo uz dugi niz opservacijskih stanja, pa ćemo u tu svrhu uvesti dvije nove vjerojatnosti pomoću kojih ćemo ju moći izraziti:

$$\begin{aligned} \alpha_i(k) &= P(X_k = i, Y_0 = j_0, \dots, Y_k = j_k) \\ \beta_i(k) &= P(Y_{k+1} = j_{k+1}, \dots, Y_n = j_n | X_k = i). \end{aligned}$$

Uočimo da vrijedi:

$$\begin{aligned} P(X_k = i | Y_0 = j_0, \dots, Y_n = j_n) &= \frac{P(X_k = i, Y_0 = j_0, \dots, Y_n = j_n)}{P(Y_0 = j_0, \dots, Y_n = j_n)} \\ &\propto P(X_k = i, Y_0 = j_0, \dots, Y_n = j_n) \\ &= P(Y_{k+1} = j_{k+1}, \dots, Y_n = j_n | X_k = i, Y_0 = j_0, \dots, Y_k = j_k) \cdot P(X_k = i, Y_0 = j_0, \dots, Y_k = j_k) \\ &= P(Y_{k+1} = j_{k+1}, \dots, Y_n = j_n | X_k = i) \cdot P(X_k = i, Y_0 = j_0, \dots, Y_k = j_k) \\ &= \alpha_i(k) \cdot \beta_i(k). \end{aligned}$$

Primijetimo da u četvrtom redu izostavljamo slučajne varijable $Y_0 = j_0, \dots, Y_k = j_k$ iz uvjeta jer prošle opservacije ne utječu na buduće, već ih emitiraju skrivene vrijednosti. Na taj način *Forward-Backward algoritam* možemo razdvojiti na dva dijela: *Forward* dio pomoću kojega računamo $\alpha_i(k)$ i *Backward* pomoću kojega ćemo računati $\beta_i(k)$.

Forward algoritam

Kako samo ime algoritma nagovješta, krenimo od početka, tj. od trenutka $k = 0$ kako bi došli do rekurzije koja nas zanima. Neka je $i \in S$ proizvoljan.

$$k = 0: \quad \alpha_i(0) = P(X_0 = i, Y_0 = j_0) = P(Y_0 = j_0 | X_0 = i) P(X_0 = i) = q_{ij_0} \lambda_i.$$

$$\begin{aligned} k = 1: \quad \alpha_i(1) = P(X_1 = i, Y_0 = j_0, Y_1 = j_1) &= \sum_{i_0 \in S} P(X_1 = i, X_0 = i_0, Y_0 = j_0, Y_1 = j_1) \\ &\overset{def}{=} \sum_{i_0 \in S} P(X_0 = i_0, Y_0 = j_0) P(X_1 = i | X_0 = i_0) P(Y_1 = j_1 | X_1 = i) \\ &= \sum_{i_0 \in S} \alpha_i(0) p_{i_0 i} q_{ij_1} \end{aligned}$$

$$
\begin{aligned}
k=2: \quad \alpha_i(2) &= P(X_2=i, Y_0=j_0, Y_1=j_1, Y_2=j_2)\\
&= \sum_{i_0\in S}\sum_{i_1\in S} P(X_2=i, X_1=i_1, X_0=i_0, Y_0=j_0, Y_1=j_1, Y_2=j_2)\\
&= \sum_{i_0\in S}\sum_{i_1\in S} p_{i_1 i} q_{ij_2} p_{i_0 i_1} q_{i_1 j_1} P(X_0=i_0, Y_0=j_0)\\
&= \sum_{i_1\in S} p_{i_1 i} q_{ij_2} \left(\sum_{i_0\in S} p_{i_0 i_1} q_{i_1 j_1} P(X_0=i_0, Y_0=j_0)\right)\\
&= \sum_{i_1\in S} p_{i_1 i} q_{ij_2} \alpha_{i_1}(1)\\
&\vdots
\end{aligned}
$$

Općenito bi onda rekurzija glasila:

$$
\alpha_i(k) = \sum_{i_{k-1}\in S} p_{i_{k-1} i} q_{ij_k} \alpha_{i_{k-1}}(k-1). \tag{3.2}
$$

Kako bismo dobili distribuciju određenog X_k, $0 \le k \le n$, potrebno je izračunati $\alpha_i(k)$ za svaki pojedini $i \in S$, ali srećom postoji jednostavniji i brži zapis. *Forward algoritam* dopušta i matrični zapis rekurzije. Definiramo matricu $Q_j, j \in O$ na sljedeći način:

$$
Q_j = \mathrm{diag}(q_{ij})_{i\in S} = \begin{bmatrix} q_{i_1 j} & 0 & \dots & 0\\ 0 & q_{i_2 j} & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & \dots & 0 & q_{i_N j} \end{bmatrix}.
$$

Tada rekurziju (3.2) za sve $i \in S$ možemo zapisati kao

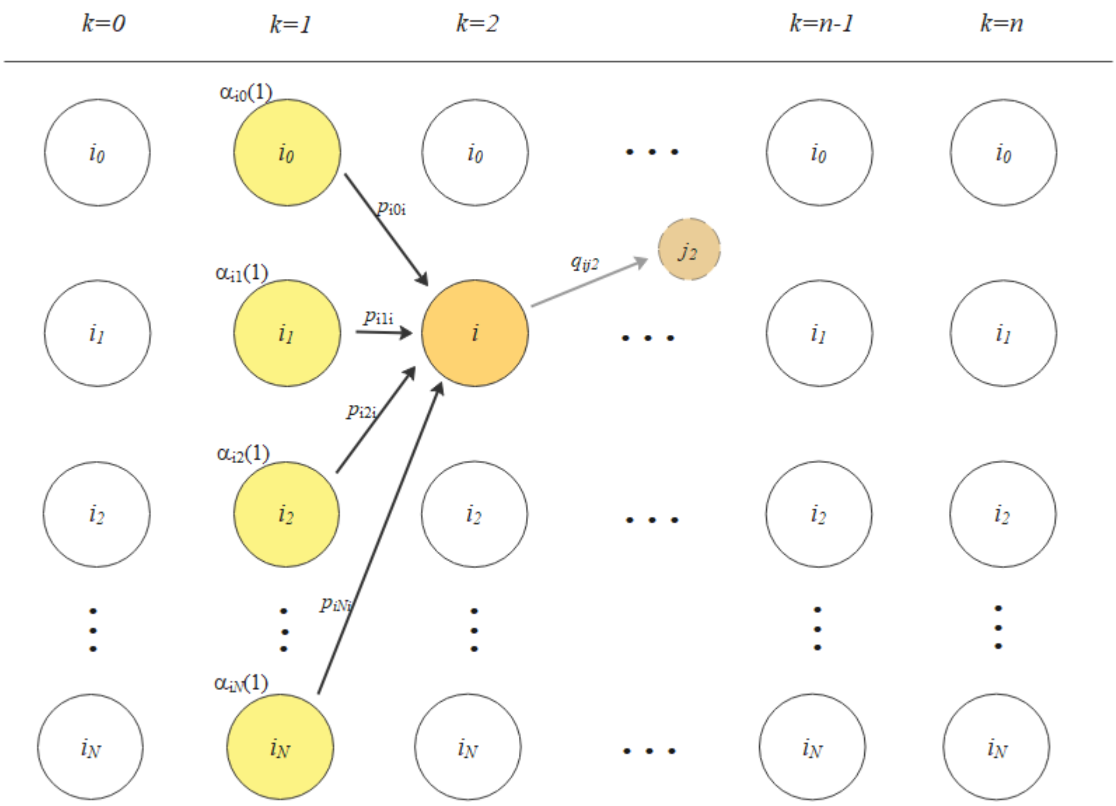
$$
\alpha(k) = \begin{bmatrix} \alpha_{i_0}(k)\\ \alpha_{i_1}(k)\\ \vdots\\ \alpha_{i_N}(k) \end{bmatrix} = \begin{cases} \lambda \cdot Q_{j_0}, & k=0\\ Q_{j_k}\cdot \Pi \cdot \alpha(k-1), & k>0. \end{cases}
$$

Forward algoritam koji računa distribucije za sve X_k, $1 \le k \le n$, smo naveli dolje:

Algoritam 1. - Forward algoritam

1. **initialize** $k=0$, (Π, Φ, λ), $o=[j_0, ..., j_n]$
2. **for** $k \longleftarrow k+1$ **do**
3. $\quad \alpha_i(k) = \sum_{i_{k-1}\in S} \alpha_{i_{k-1}}(k-1) p_{i_{k-1} i} q_{ij_k}, \forall i \in S$ ili $\alpha(k) = Q_{j_k}\cdot \Pi \cdot \alpha(k-1)$
4. **end for** $k=n$
5. **return** $\alpha(n)$
6. **end**

Koristeći se *Slikom 4*, pokušajmo ilustrirati na jednostavnom primjeru što pokušavamo postići *Forward algoritmom*. Pretpostavimo da se model nalazi u stanju i u trenutku $k = 2$ i da je u tom trenutku emitirao opservaciju j_2. Uočimo da se u koraku $k = 1$ model nalazio u nekom od stanja i_u s pripadnom vjerojatnošću $\alpha_{i_u}(1)$, $0 \leq u \leq N$. Vjerojatnost $\alpha_i(2)$ se računa kao suma svih umnožaka $\alpha_{i_u}(1) \cdot p_{i_u i}$ pomnoženih sa emitiranom vjerojatnošću q_{ij_2}.



Slika 4: Skica računanja vjerojatnosti pomoću *Forward algoritma*

Backward algoritam

U slučaju računanja vjerojatnosti $\beta_i(k)$ krećemo od kraja, tj. od zadnje vrijednosti $X_n = i$. Po dogovoru se uzima da je

$$\beta_i(n) = P(\emptyset \,|X_n = i) = 1,$$

što se može interpretirati kao da je vjerojatnost ne proučavanja novih opservacija nakon trenutka n uz poznato zadnje skriveno stanje i jednaka jedan. Pošto uzimamo fiksan niz opservacijskih stanja (u ovom slučaju od trenutka $k = 0$ do $k = n$) i kako je skup opservacija nakon trenutka n prazan skup, ovakva jednakost ima smisla. Općenito za proizvoljan $i \in S$ imamo sljedeće:

$$
\begin{aligned}
\beta_i(k) &= P(Y_{k+1} = j_{k+1}, \dots, Y_n = j_n | X_k = i) \\
&= \sum_{i_{k+1} \in S} P(Y_{k+1} = j_{k+1}, \dots, Y_n = j_n, X_{k+1} = i_{k+1} | X_k = i) \\
&= \sum_{i_{k+1} \in S} P(Y_{k+2} = j_{k+2}, \dots, Y_n = j_n | X_{k+1} = i_{k+1}, X_k = i, Y_{k+1} = j_{k+1}) \\
&\qquad \cdot P(Y_{k+1} = j_{k+1} | X_{k+1} = i_{k+1}, X_k = i) P(X_{k+1} = i_{k+1} | X_k = i) \\
&= \sum_{i_{k+1} \in S} P(Y_{k+2} = j_{k+2}, \dots, Y_n = j_n | X_{k+1} = i_{k+1}) \\
&\qquad \cdot P(Y_{k+1} = j_{k+1} | X_{k+1} = i_{k+1}) P(X_{k+1} = i_{k+1} | X_k = i) \\
&= \sum_{i_{k+1} \in S} \beta_i(k+1) q_{i_{k+1} j_{k+1}} p_{i i_{k+1}}.
\end{aligned}
$$

Rekurzija bi onda glasila:

$$\beta_i(k) = \sum_{i_{k+1} \in S} \beta_{i_{k+1}}(k+1) q_{i_{k+1} j_{k+1}} p_{i i_{k+1}}.$$

Analogno kao kod *Forward algoritma*, ovakva rekurzija se može zapisati i pomoću matrica

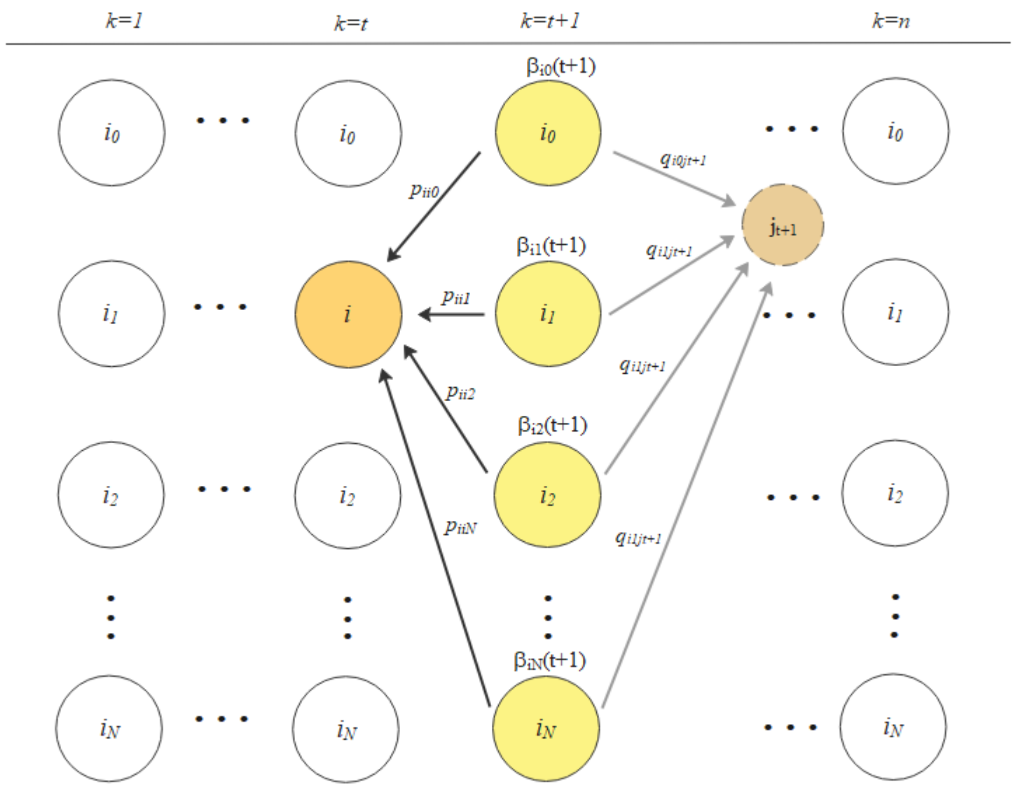
$$\beta(k) = \begin{bmatrix} \beta_{i_0}(k) \\ \beta_{i_1}(k) \\ \vdots \\ \beta_{i_N}(k) \end{bmatrix} = \Pi \cdot Q_{j_{k+1}} \cdot \beta(k+1), \quad 0 \leq k \leq n-1.$$

Nakon što smo došli do tražene rekurzije, pseudokod za *Backward algoritam* nije teško zapisati.

Algoritam 2. - Backward algoritam

1. **initialize** $k = n$, (Π, Φ, λ), $O = [j_0, ..., j_n]$
2. **for** $k \longleftarrow k - 1$ **do**
3. $\quad \beta_i(k) = \sum_{i_{k+1} \in S} \beta_{i_{k+1}}(k+1) p_{i i_{k+1}} q_{i_{k+1} j_{k+1}}, \forall i \in S$ ili $\beta(k) = \Pi \cdot Q_{j_{k+1}} \cdot \beta(k+1)$
4. **end for** $k = 0$
5. **return** $\beta(0)$
6. **end**

Radi boljeg razumijevanja pogledajmo na jednostavnom primjeru računanje β-vrijednosti. Pretpostavimo da se model nalazi u stanju i u trenutku $k = t$ i da se u trenutku $k = t+1$ može nalaziti u jednom od stanja i_u, $0 \leq u \leq N$ s pripadnom vjerojatnošću $\beta_{i_u}(t+1)$. Naravno, bilo koji od tih stanja i_u je mogao emitirati zadano opservacijsko stanje j_{t+1} pa to moramo uzeti u obzir. Vjerojatnost $\beta_i(t)$ se tada računa kao suma svih umnožaka $\beta_{i_u}(t+1) \cdot p_{i i_u} \cdot q_{i_u j_{t+1}}$.



Slika 5: Skica računanja vjerojatnosti pomoću *Backward algoritma*

Koristeći oba algoritam, tj. množenjem njihovih vraćenih vrijednosti i eventualno normalizirati ih, dobijemo traženu vjerojatnost. Pod normalizirati izraz smatramo podijeliti ga sa $P(Y_n = j_n, \ldots, Y_0 = j_0)$ kako bi dobili uvjetne vjerojatnosti, a nju možemo računati na sljedeći način:

$$\begin{aligned}
P(Y_0 = j_0, \ldots, Y_n = j_n) &= \sum_{i \in S} P(X_k = i, Y_0 = j_0, \ldots, Y_n = j_n) \\
&= \sum_{i \in S} P(Y_{k+1} = j_{k+1}, \ldots, Y_n = j_n | X_k = i) P(X_k = i, Y_0 = j_0, \ldots, Y_n = j_n) \\
&= \sum_{i \in S} \alpha_i(k) \beta_i(k).
\end{aligned}$$

Primjer 3.2. Sada se možemo vratiti na *Primjer 3.1* i odgovoriti na prvo pitanje koje se temelji na osnovnom problemu skrivenih Markovljevih modela: *Koja je vjerojatnost da je neki slučajno odabrani dan sunčan ili kišovit?* Odnosno tražimo vjerojatnosti

$$P(X_t = i|\, o), \quad i \in \{s, k\}, \quad t \in \{0, 1, \ldots, 5\},$$

gdje je o niz zadanih opservacija $\{Y_0 = d, Y_1 = d, Y_2 = l, Y_3 = l, Y_4 = l, Y_5 = d\}$. Prvo pratimo *Forward algoritam* kako bi izračunali $\alpha_i(t)$, $i \in \{s, k\}$ vrijednosti i radi jednostavnosti ćemo se koristiti matričnim zapisom rekurzije.

$$t = 0: \quad \alpha(0) = \lambda \cdot Q_d = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix} \cdot \begin{bmatrix} 0.8 & 0 \\ 0 & 0.4 \end{bmatrix} = \begin{bmatrix} 2/3 \cdot 0.8 \\ 1/3 \cdot 0.4 \end{bmatrix} = \begin{bmatrix} 0.53333 \\ 0.13333 \end{bmatrix}$$

$$t = 1: \quad \alpha(1) = Q_d \cdot \Pi \cdot \alpha(0) = \begin{bmatrix} 0.8 & 0 \\ 0 & 0.4 \end{bmatrix} \cdot \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix} \cdot \begin{bmatrix} 0.53333 \\ 0.13333 \end{bmatrix} = \begin{bmatrix} 0.38400 \\ 0.07467 \end{bmatrix}$$

$$t = 2: \quad \alpha(2) = Q_l \cdot \Pi \cdot \alpha(1) = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.6 \end{bmatrix} \cdot \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix} \cdot \begin{bmatrix} 0.38400 \\ 0.07467 \end{bmatrix} = \begin{bmatrix} 0.06741 \\ 0.07296 \end{bmatrix}$$

$$\vdots$$

Iteracije nastavimo dok ne dođemo do kraja opservacijskog niza, tj. do $n = 5$. Na taj način ćemo dobiti sve $\alpha(t)$ vrijednosti.

t	$\alpha_s(t)$	$\alpha_k(t)$
0	0.53333	0.13333
1	0.38400	0.07467
2	0.06741	0.07296
3	0.01662	0.03436
4	0.00541	0.01436
5	0.00806	0.00388

Tablica 1: Tablica vrijednosti $\alpha_i(t)$

Nakon toga, pratimo *Backward algoritam* kako bismo dobili $\beta(t)$ vrijednosti za svaki t.

$$t = 5: \quad \beta(5) = \begin{bmatrix} P(\emptyset \,|X_n = s) \\ P(\emptyset \,|X_n = k) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$t = 4: \quad \beta(4) = \Pi \cdot Q_d \cdot \beta(5) = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix} \cdot \begin{bmatrix} 0.8 & 0 \\ 0 & 0.4 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.72 \\ 0.56 \end{bmatrix}$$

$$t = 3: \quad \beta(3) = \Pi \cdot Q_l \cdot \beta(4) = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix} \cdot \begin{bmatrix} 0.2 & 0 \\ 0 & 0.6 \end{bmatrix} \cdot \begin{bmatrix} 0.72 \\ 0.56 \end{bmatrix} = \begin{bmatrix} 0.18240 \\ 0.25920 \end{bmatrix}$$

$$\vdots$$

Nakon što smo izračunali sve vrijednosti za $\beta_i(t)$ možemo rezultate zapisati u sljedeću tablicu.

t	$\beta_s(t)$	$\beta_k(t)$
0	0.01795	0.01771
1	0.02259	0.04367
2	0.06029	0.10790
3	0.18240	0.25920
4	0.72000	0.56000
5	1.00	1.00

Tablica 2: Tablica vrijednosti $\beta_i(t)$

Još nam preostaje pomnožiti vrijednosti $\alpha_i(t)$ i $\beta_i(t)$ za svaki $i \in S$ te trenutak t i potom podijeliti sa

$$P(o) = P(Y_0 = d, \ldots, Y_5 = d) = 0.01193689$$

kako bi dobili distribuciju slučajne varijable X_t za sve $t \in \{0, 1, \ldots, 5\}$ uz dani niz opservacija. Krajnje rezultate i odgovor na postavljeno pitanje smo zapisali u donjoj tablici.

t	$P(X_t = s\|\, o)$	$P(X_t = k\|\, o)$
0	0.80217	0.19783
1	0.72685	0.27315
2	0.34048	0.65952
3	0.25400	0.74599
4	0.32620	0.67379
5	0.67498	0.32502

Tablica 3: Tablica vrijednosti $P(X_t = i|\, o)$

3.2 Problem dekodiranja

Pretpostavimo da imamo niz opservacija $o = [\, j_0, \ldots, j_n]$ i da su nam poznate matrice prijelaznih i emitiranih vjerojatnosti Π i Φ zajedno sa početnom distribucijom skrivenog lanca λ. Na temelju poznatog skrivenog Markovljevog modela (Π, Φ, λ) želimo pronaći skrivena stanja $[\, i_0, \ldots, i_n]$ koja su najvjerojatnije emitirala opservacije o. Drugim riječima, cilj nam je izračunati sljedeće

$$\underset{i_0,\ldots,i_n}{\operatorname{argmax}} P(X_0 = i_0, \ldots, X_n = i_n | Y_0 = j_0, \ldots, Y_n = j_n).$$

Algoritam koji se koristi za dekodiranje je *Viterbijev algoritam*, nazvan po Andrewu Viterbiju, koji ga je 1967. godine predložio kao algoritam za dekodiranje konvolucijskih kodova preko bučnih digitalnih komunikacijskih veza. Cilj ovog poglavlja je upravo izvesti njegov algoritam. Prisjetimo se iz prijašnjeg poglavlja da vrijedi

$$P(X_0 = i_0, \ldots, X_n = i_n | Y_0 = j_0, \ldots, Y_n = j_n) \propto P(X_0 = i_0, Y_0 = j_0, \ldots, X_n = i_n, Y_n = j_n),$$

pa ćemo radi jednostavnosti računanja koristiti vjerojatnost sa desne strane izraza. Također, u raspisu ćemo umjesto argmax$(\cdot)$ koristiti funkciju max$(\cdot)$ zbog njenih određenih svojstava, ali i jedna i druga funkcija maksimizira traženu vjerojatnost pa ćemo uz pamćenje niza $i_0, \ldots, i_n$ u samom algoritmu doći do istih rezultata. Uvedimo sada funkciju od koje polazimo:

$$\mu_{i_k}(X_k) = \max_{i_0,\ldots,i_{k-1}} P(X_0 = i_0, Y_0 = j_0, \ldots, X_k = i_k, Y_k = j_k). \tag{3.3}$$

Primijetimo da gornji izraz (3.3) računa vjerojatnost posjeta skrivenog lanca $(X_t, t \in \mathbb{N}_0)$ stanju i_k u trenutku k, ako je putanja skrivenog lanca do tog trenutka bila najvjerojatnija moguća s obzirom na dana opservacijska stanja. Drugim riječima, pretpostavljamo da smo do trenutka k našli u tom smislu optimalan put kojim se skriveni lanac kretao, znajući vrijednosti opservacija do tog trenutka.

Na temelju toga, prirodno se nameće ideja o nalaženju rekurzije za rješavanje problema dekodiranja. Krenimo od trenutka $k = 0$, pomoću poznatih parametara modela (Π, Φ, λ) lako možemo izračunati početnu vrijednost funkcije (3.3):

$$\mu_{i_0}(X_0) = P(X_0 = i_0, Y_0 = j_0) = q_{i_0 j_0}\lambda_{i_0}, \quad \forall i_0 \in S.$$

U većini slučajeva se skriveno stanje i_0 za koje je $\mu_{i_0}(X_0)$ maksimum funkcije (3.3) u nuli zove početno stanje i od njega algoritam kreće. Prije nego što raspišemo općeniti izraz za rekurziju algoritma, uvest ćemo svojstvo kojim ćemo se koristiti pri njenom izvođenju.

Napomena 3.1. Za funkcije $f(a) \geq 0$, $\forall a$, i $g(a,b) \geq 0$, $\forall a, b$, vrijedi

$$\begin{aligned} \max_b \{f(a)g(a,b)\} &= f(a)\max_b \{g(a,b)\} \\ \max_{a,b} \{f(a)g(a,b)\} &= \max_a \left\{ f(a) \max_b g(a,b) \right\}. \end{aligned} \tag{3.4}$$

Krećući od općenitog izraza u trenutku k, raspisati ćemo vjerojatnost dok ne dođemo do tražene rekurzije:

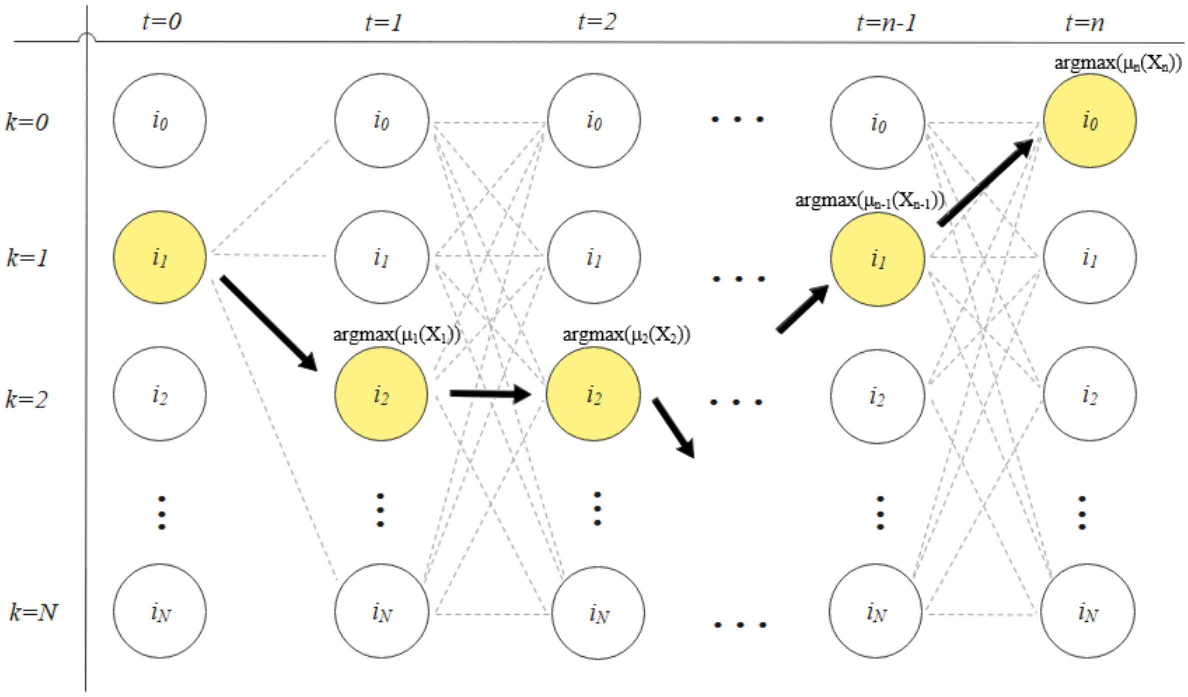
$$\begin{aligned} \mu_{i_k}(X_k) &= \max_{i_0,\dots,i_{k-1}} P(X_0 = i_0, Y_0 = j_0, \dots, X_k = i_k, Y_k = j_k) \\ &\overset{def.}{=} \max_{i_0,\dots,i_{k-1}} p_{i_{k-1}i_k} q_{i_k j_k} P(X_0 = i_0, Y_0 = j_0, \dots, X_k = i_{k-1}, Y_k = j_{k-1}) \\ &\overset{(3.4)}{=} \max_{i_{k-1}} \left\{ p_{i_{k-1}i_k} q_{i_k j_k} \max_{i_0,\dots,i_{k-2}} P(X_0 = i_0, Y_0 = j_0, \dots, X_k = i_{k-1}, Y_k = j_{k-1}) \right\} \\ &= \max_{i_{k-1}} \left\{ p_{i_{k-1}i_k} q_{i_k j_k} \mu_{i_{k-1}}(X_{k-1}) \right\}. \end{aligned}$$

Na ovaj način, pomoću rekurzije pamtimo koji je do tog trenutka bio najvjerojatniji put i tom putu samo dodamo skriveno stanje koje će lanac najvjerojatnije sljedeće posjetiti. Sada možemo zapisati *Viterbi algoritam* pomoću sljedećeg pseudokoda:

Algoritam 2. - Viterbi algoritam

1. **initialize** $t = 0$, $Put = \{\}$, (Π, Φ, λ), $o = [j_0, ..., j_n]$
2. **for** $t \longleftarrow t + 1$ **do**
3. $\quad k \leftarrow 0$
4. $\quad$ **for** $k \longleftarrow k + 1$ **do**
5. $\quad\quad \mu_{i_k}(X_t) = \max_{i_{t-1} \in S} \left\{ p_{i_{t-1}i_k} q_{i_k j_t} \mu_{i_{t-1}}(X_{t-1}) \right\}$
6. $\quad$ **end for** $k = N$
7. $\quad$ Dopuni Put sa $\underset{i_t \in S}{\operatorname{argmax}} \{\mu_{i_t}(X_t)\}$
8. **end for** $t = n$
9. **return** Put
10. **end**

Unutarnja petlja koja prolazi kroz skup skrivenih stanja S služi za nalaženje maksimalne vjerojatnosti posjeta svakom od stanja i_k, $1 \leq k \leq N$, u određenom trenutku t na temelju prijelazne vjerojatnosti u to stanje, dok vanjska petlja prolazi kroz sve promatrane vremenske trenutke (od $t = 0$ do $t = n$) i pamti skriveno stanje koje je najvjerojatnije bilo posjećeno u trenutku t. Na *Slici 6.* bismo si to mogli predočiti kao prolaženje algoritma prvo kroz stupac pa tek onda prelazi na sljedeći vremenski trenutak. Slijed tih stanja predstavlja najvjerojatniji put kroz model, a još se naziva i *Viterbijev put*.



Slika 6: Ilustracija Viterbijevog algoritama

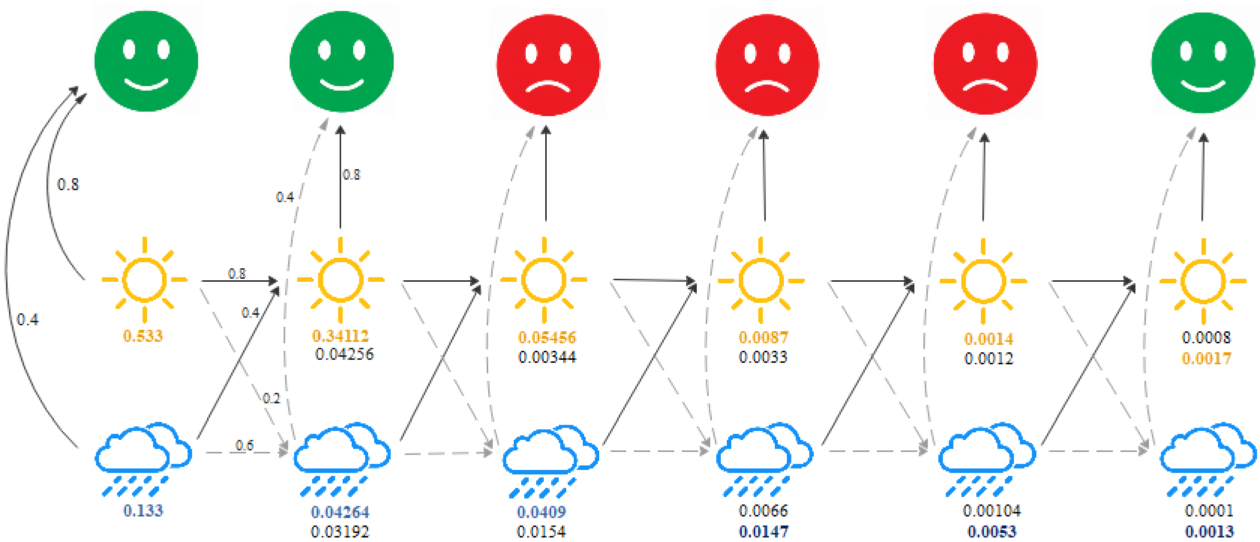
Primjer 3.3. Obratimo se sada drugom pitanju iz *Primjera 3.1*: *Ako znamo kakvog je raspoloženja otac bio u proteklih 6 dana, kakvo je bilo vrijeme?* Odnosno želimo dekodirati poznati niz opservacija $o = [\,d, d, l, l, l, d\,]$ i tako saznati koji niz skrivenih stanja ih je najvjerojatnije prouzročio ako znamo skriveni Markovljev model (Π, Φ, λ). Zato će nam poslužiti *Viterbijev algoritam* od kojeg ćemo raspisati prvih nekoliko iteracija:

$$
\begin{aligned}
t = 0: \quad & i_k = s: \ \mu_s(0) = q_{sd} \cdot \lambda_s = 0.8 \cdot 0.667 = 0.533 \\
& i_k = k: \ \mu_k(0) = q_{kd} \cdot \lambda_k = 0.4 \cdot 0.333 = 0.133 \\
& s = \operatorname{argmax}\{\mu_s(0), \mu_k(0)\}
\end{aligned}
$$

$$
\begin{aligned}
t = 1: \quad & i_k = s: \ \mu_s(1) = \max\{p_{ss}q_{sd}\mu_s(0),\ p_{ks}q_{sd}\mu_k(0)\} = 0.8 \cdot 0.8 \cdot 0.533 = 0.34112 \\
& i_k = k: \ \mu_k(1) = \max\{p_{sk}q_{kd}\mu_s(0),\ p_{kk}q_{kd}\mu_k(0)\} = 0.2 \cdot 0.4 \cdot 0.533 = 0.04264 \\
& s = \operatorname{argmax}\{\mu_s(1), \mu_k(1)\}
\end{aligned}
$$

$$
\begin{aligned}
t=2: \quad & i_k = s: \ \mu_s(2) = \max\{p_{ss}q_{sl}\mu_s(1),\, p_{ks}q_{sl}\mu_k(1)\} = 0.8 \cdot 0.2 \cdot 0.34112 = 0.05456 \\
& i_k = k: \ \mu_k(2) = \max\{p_{sk}q_{kl}\mu_s(1),\, p_{kk}q_{kl}\mu_k(1)\} = 0.2 \cdot 0.6 \cdot 0.34112 = 0.0409 \\
& s = \operatorname{argmax}\{\mu_s(2), \mu_k(2)\} \\
& \vdots
\end{aligned}
$$

Nastavljamo proceduru dok ne prođemo kroz cijeli niz zadanih opservacija, u našem slučaju je to do trenutka $t = 5$. Puni postupak provođenja algoritma kroz primjer možemo vidjeti na *Slici 7.*



Slika 7: Ilustracija *Viterbi algoritma* na *Primjeru 3.3*

Na taj način smo dobili slijed skrivenih stanja koji je najvjerojatnije emitirao zadani niz opservacija $o = [d, d, l, l, l, d]$, a on glasi $[s, s, s, k, k, s]$. Odnosno, na temelju očevog raspoloženja korištenjem *Viterbijevog algoritma* slijedi da su prva tri dana bila sunčana, sljedeća dva kišovita i na posljetku da je vrijeme bilo sunčano.

3.3 Problem treniranja

Osnovna pretpostavka u prijašnjim poglavljima je bila kako nam je u potpunosti poznat model skrivenog Markovljevog lanca, tj. poznate su nam matrice Π, Φ i početna distribucija λ. Skraćeno ćemo sve parametre koji određuju skriveni Markovljev model označiti kao $\theta = (\Pi, \Phi, \lambda)$. Međutim, takva pretpostavka nije realistična i u većini slučajeva nam prijelazne i emitirane vjerojatnosti nisu dane, već bismo ih na temelju niza podataka opservacija morali procijeniti.

Algoritam koji je zato napravljen se zove *Baum-Welch algoritam.* Dobio je ime po njegovim izumiteljima Leonardu E. Baumu i Lloydu R. Welchu, koji su algoritam kao i sam skriveni Markovljev model po prvi puta opisali u nizu članaka krajem 1960-ih godina. *Baum-Welch algoritam* traži parametre skrivenog Markovljevog modela koji bi maksimizirali vjerojatnost emitiranja poznatog niza opservacija $o = [\, j_0, \dots, j_n \,]$, odnosno

$$\underset{\theta}{\operatorname{argmax}}\, P(Y_0 = j_0, \dots, Y_n = j_n | \, \theta \,).$$

On se jednim dijelom koristi *Forward-Backward algoritmom* kojeg smo obradili u *Poglavlju 3.1* jer će nam $\alpha(t)$ i $\beta(t)$ vrijednosti služiti za izračun potrebnih izraza za algoritam. *Baum-Welch algoritam* radi na način da pri svakoj njegovoj iteraciji ažurira prijelazne i emitirane vjerojatnosti, kao i početnu distribuciju lanca, dok ne dođemo do željene točnosti procjene. Sada ćemo definirati i pojasniti vjerojatnosti kojima se služi algoritam, a vjerojatnost koju koristimo ćemo označiti s $P_\theta(\cdot)$ kako bi naglasili da koristimo upravo prijelazne i emitirane vjerojatnosti iz parametra θ pri računanju. Prvo promotrimo vjerojatnost

$$\begin{aligned} \gamma_i(t) = P_\theta(X_t = i | Y_0 = j_0, \dots, Y_n = j_n) &= \frac{P_\theta(X_t = i, Y_0 = j_0, \dots, Y_n = j_n)}{P_\theta(Y_0 = j_0, \dots, Y_n = j_n)} \\ = \frac{P_\theta(X_t = i, Y_0 = j_0, \dots, Y_n = j_n)}{P_\theta(Y_0 = j_0, \dots, Y_n = j_n, X_t \in S)} &= \frac{\alpha_i(t)\beta_i(t)}{\sum_{j \in S} \alpha_j(t)\beta_j(t)}. \end{aligned} \tag{3.5}$$

Izraz $\gamma_i(t)$ predstavlja vjerojatnost da se nalazimo u skrivenom stanju i u trenutku t uz dani niz opservacija $o = [\, j_0, \dots, j_n \,]$ i parametra θ. Nadalje, promotrimo vjerojatnost

$$\begin{aligned} \varepsilon_{ij}(t) &= P_\theta(X_t = i, X_{t+1} = j | Y_0 = j_0, \dots, Y_n = j_n) \\ &= \frac{P_\theta(X_t = i, X_{t+1} = j, Y_0 = j_0, \dots, Y_n = j_n)}{P_\theta(Y_0 = j_0, \dots, Y_n = j_n)} \\ &= \frac{\alpha_i(t) p_{ij} \beta_j(t+1) q_{j j_{t+1}}}{\sum_{u \in S} \sum_{v \in S} \alpha_u(t) p_{uv} \beta_v(t+1) q_{v j_{t+1}}}. \end{aligned} \tag{3.6}$$

Izraz $\varepsilon_{ij}(t)$ predstavlja vjerojatnost da se nalazimo u skrivenom stanju i u trenutku t i u stanju j u trenutku $t+1$ uz dani niz opservacija $[\, j_0, \dots, j_n \,]$ i parametra θ. Primijetimo, da su nazivnici u oba izraza jednaki i opisuju vjerojatnost da je skriveni Markovljev model definiran parametrom θ emitirao niz opservacija $[\, j_0, \dots, j_n \,]$. Također, ukoliko želimo prikazati te vjerojatnosti za sve kombinacije stanja $i \in S$ i $j \in O$ u određenom trenutku t, možemo ih zapisati u obliku matrica:

$$\gamma(t) = [\gamma_i]_{i \in S} = \begin{bmatrix} \gamma_{i_0}(t) \\ \gamma_{i_1}(t) \\ \vdots \\ \gamma_{i_N}(t) \end{bmatrix}, \qquad \varepsilon(t) = [\varepsilon_{ij}]_{i \in S, j \in O} = \begin{bmatrix} \varepsilon_{i_0 j_0}(t) & \varepsilon_{i_0 j_1}(t) & \dots & \varepsilon_{i_0 j_M}(t) \\ \varepsilon_{i_1 j_0}(t) & \varepsilon_{i_1 j_1}(t) & \dots & \varepsilon_{i_1 j_M}(t) \\ \vdots & \vdots & \vdots & \vdots \\ \varepsilon_{i_N j_0}(t) & \varepsilon_{i_N j_1}(t) & \dots & \varepsilon_{i_N j_M}(t) \end{bmatrix}.$$

Napomena 3.2. Ukoliko želimo saznati koliko puta je lanac $(X_n, n \in \mathbb{N}_0)$ bio u stanju $i \in S$ unutar promatranog vremena $t \in \{0, 1, \dots, n\}$ možemo definirati slučajnu varijablu

$$N_i = \sum_{t=0}^{n} \mathcal{I}_{\{X_t = i\}},$$

gdje je $\mathcal{I}_{\{X_t=i\}}$ pripadna indikator funkcija. Očekivani broj posjeta lanca $(X_n, n \in \mathbb{N}_0)$ stanju i ako su nam poznate opservacije $Y = \{Y_0 = j_0, \ldots, Y_n = j_n\}$ i parametri skrivenog Markovljevog modela θ bi se računao na sljedeći način

$$\begin{aligned} E_\theta[N_i|\, Y] &= E_\theta\left[\sum_{t=0}^{n} \mathcal{I}_{\{X_t=i\}}|\, Y\right] = \sum_{t=0}^{n} E_\theta[\mathcal{I}_{\{X_t=i\}}|\, Y] \\ &= \sum_{t=0}^{n} P_\theta(X_t = i|\, Y) = \sum_{t=0}^{n} \gamma_i(t). \end{aligned}$$

Znajući to, sumu vjerojatnosti $\gamma_i(t)$ možemo sada interpretirati kao očekivani broj posjeta lanca stanju i uz dani niz opservacija Y i parametar θ, što će nam biti od koristi za lakše razumijevanje budućih izraza. Analogno bi zaključili i za interpretaciju sume vjerojatnosti $\varepsilon_{ij}(t)$.

Za pokretanje algoritma, u prvoj iteraciji dajemo slučajne inicijalne vrijednosti za parametar θ, tj. dajemo slučajne vrijednosti za prijelazne i emitirane vjerojatnosti te početnu distribuciju. Ukoliko imamo prethodnih informacija o parametrima to bi trebali ukomponirati u inicijalne vrijednosti pošto bi to moglo ubrzati algoritam i usmjeriti ga prema željenom lokalnom maksimumu. Za daljnje iteracije obnavljamo i korigiramo njihove vrijednosti na sljedeći način:

$$\widehat{\lambda}_i = \gamma_i(0), \quad \forall i \in S, \tag{3.7}$$

što je očekivana vrijednost provedenoga vremena u stanju i u trenutku $t = 0$;

$$\widehat{p}_{ij} = \frac{\sum_{t=0}^{n-1} \varepsilon_{ij}(t)}{\sum_{t=0}^{n-1} \gamma_i(t)}, \quad \forall i, j \in S, \tag{3.8}$$

što je očekivani broj prijelaza skrivenog lanca $(X_n, n \in \mathbb{N}_0)$ iz stanja i u stanje j tijekom promatranog vremena kroz očekivani ukupni broj prijelaza lanca iz stanja i u bilo koje drugo stanje;

$$\widehat{q}_{ij} = \frac{\sum_{t=0}^{n} \mathcal{I}_{\{j_t=j\}}\gamma_i(t)}{\sum_{t=0}^{n} \gamma_i(t)}, \quad \forall i \in S, \forall j \in O, \tag{3.9}$$

gdje je $\mathcal{I}_{\{j_t=j\}}$ indikator funkcija

$$\mathcal{I}_{\{j_t=j\}} = \begin{cases} 1, & j_t = j \\ 0, & \text{inače}, \end{cases}$$

a izraz $\widehat{q}_{ij}$ predstavlja očekivani broj posjeta lanca skrivenom stanju i, a da je ono emitiralo opservaciju j, kroz očekivani ukupni broj posjeta stanju i.

Ovi koraci ažuriranja se ponavljaju iterativno sve dok se ne postigne željena razina konvergencije. Na način kako je implementiran *Baum-Welch algoritam*, garantirana je konvergencija prema najbližem lokalnom maksimumu, a dokaz takve tvrdnje i detaljnije objašnjenje možete pronaći u radu [6]. U većini slučajeva se kao kriteriji zaustavljanja uzima suma razlika između uzastopnih procjena matrica prijelaznih i emitiranih vjerojatnosti u l_2 normi, tj.

trenutak kada je razlika uzastopnih procjena manja od unaprijed određenog $\delta > 0$:

$$\|\widehat{\Pi}_t - \widehat{\Pi}_{t-1}\|_2 + \|\widehat{\Phi}_t - \widehat{\Phi}_{t-1}\|_2 < \delta, \quad t \in \mathbb{N}_0.$$

Više o matričnim normama možete pronaći u [2]. Pseudokod za *Baum-Welch algoritam* smo naveli dolje.

Algoritam 3. - Baum-Welch algoritam

1. **initialize** $k = 0$, $(\Pi_0, \Phi_0, \lambda_0)$, $o = [j_0, ..., j_n]$, δ
2. **repeat**
3. $\quad k \leftarrow k + 1$
4. $\quad$ Izračun α i β vrijednosti pomoću *Forward-Backward algoritma*
5. $\quad$ Izračun $\gamma(t)$ i $\varepsilon(t)$, $0 \leq t \leq n$, pomoću izraza (3.5) i (3.6)
6. $\quad$ ažuriranje $\widehat{\lambda}_i, \forall i \in S$ pomoću izraza (3.7)
7. $\quad$ ažuriranje $\widehat{p}_{ij}, \forall i, j \in S$ pomoću izraza (3.8)
8. $\quad$ ažuriranje $\widehat{q}_{ij}, \forall i, \in S, j \in O$ pomoću izraza (3.9)
9. **until** $\|\widehat{\Pi}_k - \widehat{\Pi}_{k-1}\|_2 + \|\widehat{\Phi}_k - \widehat{\Phi}_{k-1}\|_2 < \delta$
10. **return** $\widehat{\theta}$
11. **end**

Primjer 3.4. Konačno možemo odgovoriti i na treće pitanje iz *Primjera 3.1* koje glasi: *U slučaju da se otac preseli u grad za koji kćerka ne zna pravilo mijenjanja vremena niti očevog raspoloženja, kako možemo samo na temelju novih opservacija o_2 izračunati te vjerojatnosti?* Pretpostavimo da je kćerka bilježila promjene raspoloženja svog oca tokom 35 dana i tako dobila sljedeći niz novih opservacija:

$$o_2 = [\,d, d, d, d, l, d, d, l, l, d, d, d, d, d, l, d, d, l, d, d, d, l, l, d, d, d, d, l, l, d, d, d, d, d, d\,].$$

Pošto se otac preselio u daleko udaljen i kćerki nepoznat grad, nemamo nikakva predznanja koja bi mogli ukomponirati u inicijalne procjene početne distribucije te matrice prijelaznih i emitiranih vjerojatnosti. Zbog toga smo izabrali približno uniformne početne vrijednosti:

$$\Pi_0 = \begin{bmatrix} 0.55 & 0.45 \\ 0.45 & 0.55 \end{bmatrix}, \quad \Phi_0 = \begin{bmatrix} 0.58 & 0.42 \\ 0.43 & 0.57 \end{bmatrix}, \quad \lambda_0 = [1/2, 1/2].$$

Također, u svrhu ovog primjera ćemo za zaustavni kriteriji algoritma uzeti vrijednost $\delta = 0.0032$, koja će nam dozvoliti da ipak prođemo kroz nekoliko iteracija algoritma jer se u ovom slučaju brzo dođe do konvergencije. Držeći se *Baum-Welch algoritma* krenimo sa prvom iteracijom, tj. $k = 0$. Prvi zadatak nam je izračunati sve $\alpha(t)$ i $\beta(t)$, $0 \leq t \leq n$, vrijednosti kako bi mogli prijeći na sljedeći korak. Njih ćemo izračunati pomoću *Forward-Backward algoritma* analogno kao i u *Primjeru 3.2.* Nakon njegove primjene ćemo dobiti sljedeće vrijednosti:

t	0	1	...	33	34
$\alpha_s(t)$	0.290000	0.148625	...	4.0453e-11	2.0478e-11
$\alpha_k(t)$	0.215000	0.106963	...	2.9020e-11	1.4691e-11

Tablica 4: Tablica vrijednosti $\alpha_i(t)$

t	0	1	...	33	34
$\beta_s(t)$	7.0615e-11	1.3946e-10	...	0.512500	1.0000
$\beta_k(t)$	6.8332e-11	1.3502e-10	...	0.497500	1.0000

Tablica 5: Tablica vrijednosti $\beta_i(t)$

Time smo završili prvi korak te sada možemo dobivene vrijednosti iskoristiti za računanje vjerojatnosti $\gamma(t)$, $0 \leq t \leq 34$, prema izvedenoj formuli (3.5):

$$
\begin{aligned}
t = 0: \quad \gamma_s(0) &= \frac{\alpha_s(0)\beta_s(0)}{\alpha_s(0)\beta_s(0) + \alpha_k(0)\beta_k(0)} = 0.5652 \\
\gamma_k(0) &= \frac{\alpha_k(0)\beta_k(0)}{\alpha_s(0)\beta_s(0) + \alpha_k(0)\beta_k(0)} = 0.4348 \\
t = 1: \quad \gamma_s(1) &= \frac{\alpha_s(1)\beta_s(1)}{\alpha_s(1)\beta_s(1) + \alpha_k(1)\beta_k(1)} = 0.5724 \\
\gamma_k(1) &= \frac{\alpha_k(1)\beta_k(1)}{\alpha_s(1)\beta_s(1) + \alpha_k(1)\beta_k(1)} = 0.4276 \\
&\vdots
\end{aligned}
$$

Ovaj postupak nastavimo sve dok ne dođemo do kraja niza opservacija, tj. do trenutka $t = 34$. U donjoj tablici možemo vidjeti nekoliko tako dobivenih vjerojatnosti, a radi preglednosti smo ih izlistali pomoću matričnog zapisa:

$$\gamma(t) = \begin{array}{c} s \\ k \end{array} \begin{bmatrix} \gamma_s(t) \\ \gamma_k(t) \end{bmatrix}.$$

t	0	1	...	33	34
$\gamma(t)$	$\begin{bmatrix} 0.5652 \\ 0.4348 \end{bmatrix}$	$\begin{bmatrix} 0.5724 \\ 0.4276 \end{bmatrix}$	...	$\begin{bmatrix} 0.5725 \\ 0.4275 \end{bmatrix}$	$\begin{bmatrix} 0.4079 \\ 0.5921 \end{bmatrix}$

Tablica 6: Tablica vrijednosti $\gamma(t)$

Na sličan način ćemo izračunati vrijednosti i za vjerojatnosti $\varepsilon(t)$, $0 \leq t \leq 33$, pomoću formule (3.6):

$$
\begin{aligned}
t=0: \quad \varepsilon_{ss}(0) &= \frac{\alpha_s(0)p_{ss}\beta_s(1)q_{sd}}{\sum_{u\in S}\sum_{v\in S}\alpha_u(0)p_{uv}\beta_v(1)q_{vd}} = 0.2583\\
\varepsilon_{sk}(0) &= \frac{\alpha_s(0)p_{sk}\beta_k(1)q_{kd}}{\sum_{u\in S}\sum_{v\in S}\alpha_u(0)p_{uv}\beta_v(1)q_{vd}} = 0.3069\\
\varepsilon_{ks}(0) &= \frac{\alpha_k(0)p_{ks}\beta_s(1)q_{sd}}{\sum_{u\in S}\sum_{v\in S}\alpha_u(0)p_{uv}\beta_v(1)q_{vd}} = 0.1567\\
\varepsilon_{kk}(0) &= \frac{\alpha_k(0)p_{kk}\beta_k(1)q_{kd}}{\sum_{u\in S}\sum_{v\in S}\alpha_u(0)p_{uv}\beta_v(1)q_{vd}} = 0.2781\\
t=1: \quad \varepsilon_{ss}(1) &= \frac{\alpha_s(1)p_{ss}\beta_s(2)q_{sd}}{\sum_{u\in S}\sum_{v\in S}\alpha_u(1)p_{uv}\beta_v(2)q_{vd}} = 0.2608\\
&\vdots
\end{aligned}
$$

Postupak nastavimo sve dok ne dođemo do trenutka $t = 33$, a rezultate u matričnom zapisu možemo vidjeti u donjoj tablici.

$$
\varepsilon(t) = \begin{array}{c c} & \begin{array}{cc} s & k \end{array} \\ \begin{array}{c} s \\ k \end{array} & \begin{bmatrix} \varepsilon_{ss}(t) & \varepsilon_{sk}(t) \\ \varepsilon_{ks}(t) & \varepsilon_{kk}(t) \end{bmatrix} \end{array}.
$$

t	0	1	...	32	33
$\varepsilon(t)$	$\begin{bmatrix} 0.2583 & 0.3069 \\ 0.1567 & 0.2781 \end{bmatrix}$	$\begin{bmatrix} 0.2608 & 0.3116 \\ 0.1536 & 0.2740 \end{bmatrix}$	...	$\begin{bmatrix} 0.2616 & 0.3116 \\ 0.1535 & 0.2733 \end{bmatrix}$	$\begin{bmatrix} 0.2571 & 0.3155 \\ 0.1509 & 0.2766 \end{bmatrix}$

Tablica 7: Tablica vrijednosti $\varepsilon(t)$

U konačnici smo dobili sve vrijednosti koje su nam potrebne za novu procjenu prijelaznih i emitiranih vjerojatnosti, te početne distribucije.

Prema formuli (3.7) lako dobivamo novu procjenu za početnu distribuciju λ:

$$
\widehat{\lambda}_1 = \gamma(0) = \begin{bmatrix} 0.5652 \\ 0.4348 \end{bmatrix}.
$$

Kako bismo procijenili Π moramo prvo izračunati očekivani broj uzastopnih posjeta skrivenog lanca stanju $i \in S$ u trenutku t te stanju $j \in S$ u trenutku $t+1$ i očekivani broj ukupnih posjeta skrivenog lanca određenom stanju $i \in S$. Navedena očekivanja dobijemo zbrajanjem vjerojatnosti $\varepsilon(t)$ i $\gamma(t)$ po trenucima t.

$$
\sum_{t=0}^{33} \varepsilon(t) = \begin{bmatrix} 8.650269 & 9.409446 \\ 6.093399 & 9.846886 \end{bmatrix}, \qquad \sum_{t=0}^{33} \gamma(t) = \begin{bmatrix} 18.05972 \\ 15.94028 \end{bmatrix}.
$$

Uočimo da je suma prvog retka matrice $\sum_{t=0}^{33} \varepsilon(t)$ upravo očekivani broj posjeta skrivenog lanca stanju s, dok je suma drugog retka očekivani broj posjeta stanju k. Dijeljenjem pripadnih komponenti matrica kao u formuli (3.8) dobivamo novu procjenu za matricu prijelaznih vjerojatnosti Π:

$$\widehat{\Pi}_1 = \begin{bmatrix} 0.4789815 & 0.5210185 \\ 0.3822641 & 0.6177359 \end{bmatrix}.$$

Također, kako bi dobili procjenu za Φ moramo pronaći očekivani broj posjeta skrivenog lanca stanju $i \in S$, a da je pri tome ono emitiralo opservaciju $j \in O$, i očekivani broj ukupnih posjeta skrivenog lanca stanju $i \in S$. Tražena očekivanja računamo na temelju formule (3.9):

$$\begin{bmatrix} \sum_{t=0}^{34} \mathbb{I}_{\{j_t=d\}} \gamma_s(t) & \sum_{t=0}^{34} \mathbb{I}_{\{j_t=l\}} \gamma_s(t) \\ \sum_{t=0}^{34} \mathbb{I}_{\{j_t=d\}} \gamma_k(t) & \sum_{t=0}^{34} \mathbb{I}_{\{j_t=l\}} \gamma_k(t) \end{bmatrix} = \begin{bmatrix} 14.67294 & 3.794704 \\ 11.32706 & 5.205296 \end{bmatrix}, \qquad \sum_{t=0}^{34} \gamma(t) = \begin{bmatrix} 18.46764 \\ 16.53236 \end{bmatrix}.$$

Uočimo i ovdje, kako suma prvog retka matrice odgovara očekivanom broju posjeta skrivenog lanca stanju s, a suma drugog retka očekivanom broju posjeta stanju k. Nova procjena za matricu emitiranih vjerojatnosti Φ tada glasi:

$$\widehat{\Phi}_1 = \begin{bmatrix} 0.7945215 & 0.2054785 \\ 0.6851450 & 0.3148550 \end{bmatrix}.$$

Time smo završili prvu iteraciju *Baum-Welch algoritma*. Preostaje nam provjeriti kriteriji zaustavljanja, odnosno je li zbroj traženih normi manji od našeg izabranog $\delta = 0.0032$. Nakon prve iteracije imamo sljedeće:

$$\|\widehat{\Pi}_1 - \widehat{\Pi}_0\|_2 + \|\widehat{\Phi}_1 - \widehat{\Phi}_0\|_2 = 0.54207 \not< \delta.$$

Pošto izraz ne zadovoljava kriteriji zaustavljanja, nastavljamo sa algoritmom i ulazimo u drugu iteraciju čije vrijednosti sada računamo na temelju novih procjena $(\widehat{\Pi}_1, \widehat{\Phi}_1, \widehat{\lambda}_1)$. Ispostavlja se kako bi u trećem koraku već zadovoljili kriteriji i izašli iz petlje.

$$\|\widehat{\Pi}_3 - \widehat{\Pi}_2\|_2 + \|\widehat{\Phi}_3 - \widehat{\Phi}_2\|_2 = 0.0031453 < \delta.$$

Procjene za parametre skrivenog Markovljevog modela nakon treće procjene, a time ujedno i konačne, su sljedeće:

$$\widehat{\lambda}_3 = \begin{bmatrix} 0.71631 \\ 0.28369 \end{bmatrix},$$

$$\widehat{\Pi}_3 = \begin{bmatrix} 0.5847486 & 0.4152514 \\ 0.4848827 & 0.5151173 \end{bmatrix},$$

$$\widehat{\Phi}_3 = \begin{bmatrix} 0.7964379 & 0.2035621 \\ 0.6802875 & 0.3197125 \end{bmatrix}.$$

Ako nove procjene usporedimo sa parametrima skrivenog Markovljevog modela prije nego što se otac preselio možemo uočiti određene razlike. Vidimo da su vjerojatnosti promjena vremenskih prilika iz dana u dan približno jednake pa ih je stoga teže predvidjeti nego što je to bio slučaj u prošlom gradu. Također, možemo zaključiti kako je otac češće dobre volje, bez obzira na vremenske uvjete.

Sažetak

Skriveni Markovljevi lanci su poznati alat za matematičko modeliranje problema koji mogu biti karakterizirani pomoću jednog nama skrivenog proces $(X_n, n \in \mathbb{N}_0)$ koji emitira drugi nama vidljiv opservacijski proces $(Y_n, n \in \mathbb{N}_0)$. Ovakvi modeli su pronašli primjenu u raznovrsnim područjima znanosti poput bioinformatike, komunikacijske tehnologije te financijske matematike, a neke primjere smo naveli u samom radu.
Skriveni Markovljev model je sasvim određen poznavanjem njegovih matrica prijelaznih i emitiranih vjerojatnosti Π i Φ te početne distribucije skrivenog lanca λ. Sama funkcionalnosti skrivenih Markovljevih modela se zasniva na tri osnovna problema, a pod uvjetom poznavanja parametara modela možemo riješiti problem izglađivanja pomoću *Forward-Backward algoritma* i problem dekodiranja pomoću *Viterbi algoritma*, koja su oba navedena u radu i ilustrirana na primjerima. U slučaju potrebe procjene samih parametara modela, što je poznato pod nazivom problem treniranja, koristimo se *Baum-Welch algoritmom*, koji je također pojašnjen u radu.

Ključne riječi: Skriveni Markovljev model, matrica prijelaznih vjerojatnosti, matrica emitiranih vjerojatnosti, početna distribucija, problem izglađivanja, Forward-Backward algoritam, problem dekodiranja, Viterbi algoritam, problem treniranja, Baum-Welch algoritam.

Abstract

Hidden Markov chains are a powerful tool for mathematical modeling of problems that can be characterized by one hidden process $(X_n, n \in \mathbb{N}_0)$ which emits another visible observational process $(Y_n, n \in \mathbb{N}_0)$. Such models have found its application in various scientific fields such as bioinformatics, communication technology and financial mathematics, and we have given some examples in the paper itself.
The hidden Markov model is completely determined by knowing its matrices of transition and emission probabilities Π and Φ and the initial distribution of the hidden chain λ. The functionality of hidden Markov models is based on three basic problems, and provided we know the parameters of the model we can solve the smoothing problem using the *Forward-Backward algorithm* and the decoding problem by using the *Viterbi algorithm*, both of which are listed in the paper and illustrated on examples. In case we need to estimate the parameters of the model themselves, which is known as the training problem, we use the *Baum-Welch algorithm*, which is also explained in the paper.

Key words: Hidden Markov model, transsition probability matrix, emission probability matrix, initial distribution, smoothing problem, Forward-Backward algorithm, decoding problem, Viterbi algorithm, training problem, Baum-Welch algorithm.

Literatura

[1] D. Jurafsky, J. H. Martin, Speech and Language Processing, Stanford University, Stanford, 2019.

[2] N. Truhar, Numerička linearna algebra, Odjel za matematiku, Osijek, 2010.

[3] R. van Handel, Hidden Markov Models, Princeton University, New Jersey, 2008.

[4] Z. Vondraček, Markovljevi lanci, PMF-Matematički odsjek, Zagreb, 2008.

[5] R. S. Mamon, R. J. Elliott, Hidden Markov Models in Finance, Springer, New York, 2007.

[6] F. Yang, S. Balakrishnan, M. J. Wainwright, Statistical and Computational Guarantees for the Baum-Welch Algorithm, Journal of Machine Learning Research 18 (2017) 1-53.

[7] Z. Wu, A hidden Markov model for earthquake declustering, Journal of geophysical research 115 (2010).

Životopis

Moje ime je Sanja Stjepanović i rođena sam 26.11.1996. u Osijeku. Pohađala sam osnovnu školu "Frana Krste Frankopana" u Osijeku od 2003. do 2011. godine i u sklopu nje bila član odbojkaškog kluba i likovnjaka, te izvan škole pohađala plesni klub "MIDAL". Školovanje sam nastavila 2011. godine upisom u III. gimnaziju Osijek koju sam završila 2015. godine. Iste te godine se upisujem na Preddiplomski studiji matematike na Odjelu za matematiku Sveučilišta J.J. Strossmayera u Osijeku i započinjem držati instrukcije iz matematike kao i programiranja. Godine 2018. stekla sam zvanje prvostupnika matematike na osnovu završnog rada pod nazivom *Reprezentacija linearnog funkcionala i hermitsko adjungiranje* uz mentorstvo doc. dr. sc. Suzane Miodragović. Nastavljam studiranje upisom na Diplomski studiji financijske matematike i statistike na Odjelu za matematiku u Osijeku te sam u sklopu njega odradila stručnu praksu u osiguravajućem društvu *GRAWE Hrvatska d.d.* u odjelu za aktuare.