REJTETT MARKOV MODELLEK I.

Gerencsér László   Michaletzky György   Molnár-Sáska Gábor
Vágó Zsuzsanna

(Oktatási segédlet)

Budapest, Eötvös Loránd Tudományegyetem
Valószínűségelméleti és Statisztika Tanszék

2001. október

1. Markov-láncok

Legyen ϕ_n diszkrét idejű folyamat, az állapottér $\mathcal{X}$ megszámlálható.

1.1. Definíció. (ϕ_n) *Markov-lánc, ha*

$$P(\phi_n = x_n|\phi_{n-1} = x_{n-1}, \dots, \phi_0 = x_0) = P(\phi_n = x_n|\phi_{n-1} = x_{n-1}).$$

A fenti lánc átmeneti mátrixa $P = (P(x, y))$, melyet így definiálunk:

$$P(x, y) = P(\phi_n = y|\phi_{n-1} = x).$$

Ennek alaptulajdonságai:

$$P(x, y) \geq 0, \qquad \sum_y P(x, y) = 1.$$

Tekintsük az általános esetet, amikor az állapottér tetszőleges $\mathcal{X}$ mérhető halmaz, és adott ezen egy $\mathcal{B}(\mathcal{X})$ végesen generált σ-algebra.

1.2. Definíció. *Egy $P = P(x, A)$ leképezést, ahol $x \epsilon \mathcal{X}$, $A \epsilon \mathcal{B}(\mathcal{X})$ átmeneti valószínűség-magnak nevezünk, ha*

- *$\forall A \epsilon \mathcal{B}(\mathcal{X})$ esetén $P(\cdot, A)$ x-ben mérhető, és nemnagatív.*
- *$\forall x \epsilon \mathcal{X}$ esetén $P(x, \cdot)$ valószínűségi mérték.*

Markov-lánc esetén ezt a függvényt értelemszerűen így definiáljuk:

$$P(x, A) = P(\phi \epsilon A|\phi_{n-1} = x).$$

Általános állapottér esetén nem világos, hogy a Markovitási tulajdonságot hogyan definiáljuk. Az előző definíció mintájára történő kísérleti definíció:

$$P(\phi_n \epsilon A_n|\phi_{n-1} \epsilon A_{n-1}, \dots \phi_0 \epsilon A_0) = P(\phi_n \epsilon A_n|\phi_{n-1} \epsilon A_{n-1})$$

nem teljesül, mert $\phi_{n-1}\epsilon A_{n-1}$ túl kevés információt tartalmaz.

Mielőtt a definíciót megadnánk, néhány példát nézünk.

1. Példa. Tegyük fel, hogy az X_n folyamat kielégíti az alábbi rekurziót:

$$X_{n+1} = AX_n + BW_{n+1}$$

ahol a W_n i.i.d. sorozat független X_0-tól.

2. Példa. (1. speciális esete.) Legyen

$$X_{n+1} = X_n + W_{n+1}.$$

Ez a klasszikus bolyongási folyamat.

3. Példa. Legyen

$$X_{n+1} = [X_n + W_{n+1}]^+ .$$

Ez egy bolyongás a pozitív félsíkon.

4. Példa. $G1/G/1$ sorbanállási modell. Legyenek $T_0, T_0 + T_1, \ldots$ az érkezési idők, ahol T_i i.i.d. sorozat, továbbá legyenek a kiszolgálási idők S_i szintén i.i.d.. Ekkor a W_n hátralévő kiszolgálási idő az n-dik kliens érkezése előtt-re azt kapjuk, hogy $W_{n+1} = [W_n + S_{n+1} - T_{n+1}]^+$, ami a 3. példa általánosítása.

Általánosítás: Legyen $\Phi_{n+1} = F(\Phi_n, W_{n+1})$, ahol W_n i.i.d..

2. Feltételes függetlenség

Emlékeztetünk rá, hogy hogyan definiáltuk események feltételes függetlenségét. Legyenek A, B, C véletlen események valamilyen véletlen eseménytérben. Azt mondjuk, hogy *A és C feltételesen függetlenek B-re nézve*, ha

$$P(AC|B) = P(A|B)P(C|B).$$

Ezt így jelöljük:

$$A \perp C | B$$

Triviális átírása a fenti relációnak:

$$\frac{P(AC|B)}{P(A|B)} = P(C|AB).$$

Ezt állítás formájában is megfogalmazzuk:

2.1. Állítás. *A és C pontosan akkor feltételesen függetlenek B-re nézve, ha*

$$P(C|AB) = P(C|B) \tag{1}$$

teljesül.

Megjegyezük, hogy a fenti állí tásban A és C szerepe nem szimmetrikus. Ebben a formájában a feltételes függetlenség éppen a markovitást jelenti. Ha ugyanis (ϕ_n) Markov-lánc a diszkrét $\mathcal{X}$ állapottéren, akkor

$$A = \{\phi_{n-1} = x_{n-1}\}, \qquad B = \{\phi_n = x_n\}, \qquad C = \{\phi_{n+1} = x_{n+1}\}$$

jelöléssel az (1) összefügges épp a lánc markovitását jelenti.

Általában a fenti definícióban A a *múlt*, B a *jelen* és C a *jövő*nek felel meg.

2.1. Definíció. *Legyenek $\mathcal{F}$, $\mathcal{G}$ és $\mathcal{H}$ σ-algebrák az $(\Omega, \mathcal{A}, P)$ valószínűségi mező felett. Azt mondjuk, hogy $\mathcal{F}$ és $\mathcal{H}$ feltételesen függetlenek $\mathcal{G}$-re nézve, ha $\forall A \epsilon \mathcal{F}$ és $\forall C \epsilon \mathcal{H}$ esetén*

$$P(AC|\mathcal{G}) = P(A|\mathcal{G})P(C|\mathcal{G}).$$

Jelölés: Ezt a tulajdonságot így jelöljük: $\mathcal{F} \perp \mathcal{H} | \mathcal{G}$.

2.2. Állítás. *$\mathcal{F} \perp \mathcal{H} | \mathcal{G}$ pontosan akkor teljesül, ha $\forall C \epsilon \mathcal{H}$ esetén*

$$P(C|\mathcal{F} \cup \mathcal{G}) = P(C|\mathcal{G}). \tag{2}$$

Bizonyítás. Elegendő belátni, hogy $P(C|\mathcal{G})$ $\mathcal{F} \cup \mathcal{G}$- mérhető. Ehhez be kell látni, hogy minden $D \epsilon \mathcal{F} \cup \mathcal{G}$ esetén

$$\int_D \chi_C dP = \int_D P(C|\mathcal{G}) dP. \tag{3}$$

Most elegendő azt belátni, hogy (3) teljesül minden $D = A \cup B$ halmazra, ahol $A \epsilon \mathcal{F}$, $B \epsilon \mathcal{G}$. Erre a halmazra (3) azzal ekvivalens, hogy

$$\mathrm{E}(\chi_A \chi_B \chi_C) = \mathrm{E}(\chi_A \chi_B P(C|\mathcal{G})).$$

Ennek $\mathcal{G}$-re vett feltételes várható értéke

$$\mathrm{E}(\chi_B \mathrm{E}(\chi_A \chi_C|\mathcal{G})) = \mathrm{E}(\chi_B P(C|\mathcal{G}) \mathrm{E}(\chi_A|\mathcal{G})).$$

Ez a B halmazon vett integrálok azonosságaát jelenti, minden B- re, azaz

$$\mathrm{E}(\chi_A \chi_C|\mathcal{G}) = P(C|\mathcal{G}) \mathrm{E}(\chi_A|\mathcal{G}),$$

ez pedig nyilván igaz.

Megjegyzés. $\mathcal{F} \perp \mathcal{H}|\mathcal{G}$ nem azt jelenti, hogy $\forall A \epsilon \mathcal{F}$, $B \epsilon \mathcal{G}$, $C \epsilon \mathcal{H}$ -re $A \perp C|B$!

Ezek után megfogalmazhatjuk az általános Markov-lánc fogalmát.

2.2. Definíció. *Legyen* (X_n) *sztochasztikus folyamat valamely általános* $\mathcal{X}$ *állapottéren. Ez Markov lánc, ha*

$$\begin{aligned} \mathcal{F} &= \sigma(X_k : \ k \leq n) \\ \mathcal{G} &= \sigma(X_n) \\ \mathcal{H} &= \sigma(X_k : \ k \geq n+1) \end{aligned}$$

választással $\mathcal{F} \perp \mathcal{H}|\mathcal{G}$ *teljesül minden* n*-re. A Markov lánc homogén, ha*

$$P(X_{n+1} \epsilon A | X_n = x) = P(x, A)$$

független n*-től minden* x*,* A *mellett.*

Megjegyzés. A fenti definícióban a múlt és jövő szerepe szimmetrikus. Tehát ha egy folyamat Markov lánc, akkor "visszafele" is Markov lánc.

Legyen (X_n) Markov lánc, amelynek állapottere $\mathcal{X}$ megszámlálható. Vezessük be az alábbi jelöléseket:

$$\begin{aligned} \mathcal{F}_n^{X-} &= \sigma(X_k :\ 0 \leq k \leq n) && (4)\\ \mathcal{F}^{X_n} &= \sigma(X_n) && (5)\\ \mathcal{F}_n^{X+} &= \sigma(X_k :\ k \geq n). && (6) \end{aligned}$$

2.3. Állítás. *Az (X_n) Markov láncra teljesülnek az alábbi feltételes függetlenségek:*

$$\mathcal{F}^{X_{n+1}} \perp \mathcal{F}_n^{X-} | \mathcal{F}^{X_n}$$

$$\mathcal{F}_{n+1}^{X+} \perp \mathcal{F}_n^{X-} | \mathcal{F}^{X_n}$$

Bizonyítás. Atomonként beláthatóak, épp a markovitási tulajdonságot kapjuk.

3. Rejtett Markov folyamat

3.1. Definíció. *Az (X_n, Y_n) rejtett Markov folyamatot a következőképpen definiáljuk. Legyen (X_n) egy homogén Markov lánc valamely $\mathcal{X}$ diszkrét állapottéren. (X_n) az állapotfolyamat. Megfigyelünk egy (Y_n) folyamatot, ahol $Y_n \epsilon \mathcal{Y}$, $\mathcal{Y}$ a kiolvasási tér. Feltesszük, hogy a kiolvasási valószínűségekre teljesül, hogy*

$$P(Y_n = y_n, \ldots Y_0 = y_o | X_n = x_n, \ldots X_0 = x_0) = \prod_{i=0}^{n} P(Y_i = y_i | X_i = x_i).$$

Bevezetjük az alábbi jelöléseket:

$$P(Y_k = y_k | X_k = x_k) = P(y_k | x_k),$$

$$P(y|x) = b^x(y).$$

3.1. Állítás. *Ha (X_n, Y_n) rejtett Markov folyamat, akkor $Z_n = (X_n, Y_n)$ Markov folyamat*

Ennél az állításnál többet fogunk belátni:

3.2. Állítás.
$$\mathcal{F}_{n+1}^{Z+} \perp \mathcal{F}_n^{Z-} | \mathcal{F}^{X_n}.$$

1. Bizonyítás. Kihasználjuk, hogy diszkrét állapottérben vagyunk. Legyen

$$C \epsilon \sigma(Z_k: \ n+1 \leq k \leq n+m), \qquad m \geq 1.$$

m-re vonatkozó teljes indukciót fogunk végezni. Be kell látni, hogy

$$P(C|\mathcal{F}_n^{Z-}) = P(C|\mathcal{F}^{X_n}).$$

Vegyük $\mathcal{F}_n^{Z-}$ egy elemét:

$$A = \{Y_n = y_n, X_n = x_n, \ldots Y_0 = y_0, X_0 = x_0\}.$$

Belátjuk, hogy $P(C|A) = P(C|X_n = x_n)$. Ezt elegendő olyan C-re belátni, mely $C = DB$ alakú, ahol $D \epsilon \sigma(Z_k: \ n+2 \leq k \leq n+m)$, $B \epsilon \sigma(Z_{n+1})$.

Tudjuk, hogy
$$P(DB|A) = P(D|BA) \cdot P(B|A).$$

Az indukció szerint $P(D|BA) = P(D|B)$, másrészt $P(B|A) = P(B|X_n = x_n)$. Innen az állítás következik.

2. Bizonyítás. Általános állapottérben bizonyítunk. Legyen $(Z_n) = (X_n, Y_n)$ egy $\mathcal{Z}$ lokálisan kompakt, szeparábilis, teljes topologikus téren értelmezett folyamat. A bizonyítás a következő lemmán alapul:

3.1. Lemma. *A fenti jelöleéseket használva $\mathcal{F} \perp \mathcal{H} | \mathcal{G}$ pontosan akkor teljesül, ha minden η integrálható, $\mathcal{H}$-mérhető valószínűségi változóra*

$$\mathrm{E}(\eta | \mathcal{F} \cup \mathcal{G}) = \mathrm{E}(\eta | \mathcal{G}).$$

További példákat mutatunk rejtett Markov folyamatokra.

1. Keverék eloszlások
Legyen X_n egy véges állapottéren értelmezett i.i.d. valószínűségi változó sorozat. Az állapottér $\mathcal{X} = \{1, 2, \ldots, d\}$, és X_n eloszlása $P(X_n = i) = p_i$. A megfigyelt Y_n eloszlása X_n értékétől függ, $\mathcal{L}(Y_n \mid X_n = i) = (f_i)$.

2. Perspektív identifikáció.
Legyen X_n egy olyan folyamat, melyre

$$X_{n+1} = AX_n + W_n,$$

ahol A tetszőleges mátrix, W_n i.i.d. zajfolyamat. X_n nyilván Markov. A megfigyelt érték legyen X_n szöge, azaz

$$Y_n = \frac{X_n}{|X_n|}.$$

3. GARCH modellek (Pénzügyi idősorok)
Legyen P_t valamely pénzügyi ár-folyamat. Ennek fontos jellemzője a *volatilitás*, ami azt mutatja, hogy "mennyire erősen változik" a folyamat. (Pontos definíciót itt nem adunk.) A fenti folyamattal kapcsolatosan szokás vizsgálni a hozamot, melyet így definiálunk:

$$X_t = \log \frac{P_t}{P_{t-1}}.$$

Ennek a folyamatnak a modellezésére használják a GARCH(p,q) modelleket:

$$\begin{aligned} X_t &= \sigma_t Z_t \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \end{aligned}$$

ahol (Z_t) i.i.d. sorozat, $\mathrm{E}Z_t = 0$. Az egyenletben szereplő együtthatókra $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$. Ebben a modellben σ_t a volatilitás. Ha $\beta_j \equiv 0$, akkor a fenti folyamat ARCH típusú.

Könnyen látható például az ARCH esetben, hogy a

$$\phi_t = (X_{t-1}^2, \ldots, X_{t-p}^2)^T$$

vektor folyamat eleget tesz egy

$$\phi_t = A_t \phi_{t-1} + B_t$$

rekurziónak, ahol (A_t, B_t) i.i.d. sorozat. Tehát (X_t, ϕ_t) rejtett Markov-folyamat.

4. Véges állapotterű rejtett Markov folyamatok

Adott az (X_n, Y_n) rejtett Markov folyamat. Feltesszük, hogy X_n és Y_n értékei egy véges halmazból valóak, $X_n \epsilon \mathcal{X} = \{1, \ldots, N\}$, $Y_n \epsilon \mathcal{Y} = \{1, \ldots M\}$. Mivel *absztrakt* folyamatokkal foglalkozunk, igazából csak az átmenet valószínűségek számítanak. Így az $\mathcal{X}$ elemei kicserélhetőek, legyen

$$\mathcal{X} = \{e_1, \ldots, e_N\},$$

ahol e_i az N-dimenziós tér i-dik egységvektora. Vezessük be az átmenet valószínűségekre az alábbi jelölést:

$$P(X_{n+1} = e_j \mid X_n = e_i) = A_{ji}.$$

(Vegyük észre, hogy az indexek nem a megszokott sorrendben vannak.) Ez az A mátrix oszlop-sztochasztikus, azaz

$$\mathbf{1}^T A = \mathbf{1}^T, \qquad A_{ji} \geq 0 \qquad \sum_j A_{ji} > 0.$$

$(\mathbf{1}^T = (1, \ldots, 1)$).

Bevezetjük az alábbi σ-algebrákat:

$$\mathcal{F}_n^X = \sigma(X_i,\ 0 \leq i \leq n), \qquad \mathcal{G}_n = \mathcal{F}_n^{X-} \vee \mathcal{F}_n^{X+}.$$

4.1. Állítás.

$$\mathrm{E}(X_{n+1} \mid \mathcal{G}_n) = AX_n.$$

Bizonyítás. A rejtett Markov tulajdonság alapján tudjuk, hogy az alábbi σ-algebrák feltételesen ortogonálisak:

$$\mathcal{F}^{X_{n+1}} \vee \mathcal{F}^{Y_{n+1}} \perp \mathcal{F}_n^{X-} \vee \mathcal{F}_n^{Y-} \mid \mathcal{F}_n^X.$$

Emiatt
$$\mathrm{E}(X_{n+1} \mid \mathcal{G}_n) = \mathrm{E}(X_{n+1} \mid X_n).$$
Mivel X_n értékkészlete atomos, az állítást atomonként látjuk be:
$$\mathrm{E}(X_{n+1} \mid X_n = e_i) = \sum_{j=1}^{N} e_j P(X_{n+1} = e_j \mid X_n = e_i) = \sum_{j=1}^{N} e_j A_{ji}.$$

Definiáljuk az alábbi folyamatot:
$$V_n = X_{n+1} - AX_n.$$
Ekkor a fenti állítás értelmében $\mathrm{E}(V_n \mid \mathcal{G}_n) = 0$.

Válasszuk meg az absztrakt $\mathcal{Y}$ tér elemeit is speciálisan, nevezetesen legyen
$$\mathcal{Y} = \{f_1, \dots, f_M\},$$
ahol f_j az M-dimenziós Euklideszi tér j-dik egységvektora. A kiolvasási valószínűségeket így jelöljük:
$$C_{ji} = P(Y_n = f_j \mid X_n = e_i).$$

4.2. Állítás.
$$\mathrm{E}(Y_n \mid \mathcal{G}_n) = CX_n.$$

Bizonyítás. Az előző bizonyításhoz hasonlóan, az $\{X_n = e_i\}$ halmazon igazoljuk az állítást:
$$\mathrm{E}(Y_n \mid X_n = e_i) = \sum_{j=1}^{M} f_j P(Y_n = f_j \mid X_n = e_i) = \sum_{j=1}^{M} f_j C_{ji}.$$

Bevezetve a
$$W_n = Y_n - CX_n$$
folyamatot, az (X_n, Y_n) rejtett Markov folyamat alábbi realizációját kapjuk:

4.3. Állítás. *Legyen* (X_n, Y_n) *véges állapotterű rejtett Markov folyamat. Ekkor megadható ennek az alábbi reprezentációja:*

$$X_{n+1} = AX_n + V_n, \qquad \mathrm{E}(V_n \mid \mathcal{G}_n) = 0 \tag{7}$$
$$Y_n = CX_n + W_n, \qquad \mathrm{E}(W_n \mid \mathcal{G}_n) = 0 \tag{8}$$

Felmerül a kérdés, hogy igaz-e a fenti állítás megfordítása, azaz ha (X_n, Y_n) eleget tesz a (7), (8) összefüggésnek, akkor vajon rejtett Markov folyamat-e. Ez így nem igaz, kell egy kis kiegészítés.

4.4. Állítás. *Tegyük fel, hogy* (X_n) *eleget tesz az alábbi rekurziónak:*

$$X_{n+1} = AX_n + V_n,$$

ahol $\mathrm{E}(V_n \mid \mathcal{F}_n^{X-}) = 0$. *Ekkor* (X_n) *Markov folyamat, melynek átmenet valószínűsége* A.

Bizonyítás. Azt kell belátni, hogy

$$P(X_{n+1} = e_j \mid \mathcal{F}_n^{X-}) = P(X_{n+1} = e_j \mid X_n).$$

Egyszerű átalakítással:

$$P(X_{n+1} = e_j \mid \mathcal{F}_n^{X-}) = \mathrm{E}(\langle X_{n+1}, e_j \rangle \mid \mathcal{F}_n^{X-}) = \langle \mathrm{E}(X_{n+1} \mid \mathcal{F}_n^{X-}), e_j \rangle = \langle AX_n, e_j \rangle,$$

ami valóban X_n-mérhető. A fenti egyenlet utolsó egyenlőségében felhasználtuk az alábbi azonosságot:

$$\mathrm{E}(X_{n+1} | \mathcal{F}_n^{X-}) = \mathrm{E}(AX_n + V_n | \mathcal{F}_n^{X-}) = AX_n.$$

Ezek után az átmenet valószínűség:

$$P(X_{n+1} = e_j | X_n = e_i) = \langle Ae_i, e_j \rangle = A_{ji}.$$

4.1. Következmény. *Ha* (X_n, Y_n) *eleget tesz az (7), (8) összefüggésnek, és* W_n *i.i.d. független* (V_n)*-től, akkor* (X_n, Y_n) *rejtett Markov folyamat.*

Az (7), (8) leírás nagyon emlékeztet a lineáris sztochasztikus rendszerek reprezentációjára. Az ott használt fogalmakat, technikákat adaptáljuk rejtett Markov folyamatok esetére.

Elsőként az $\mathrm{E}(X_n X_n^T)$ várható értéket határozzuk meg. A (7) egyenletet transzponáltjával megszorozva

$$X_{n+1}X_{n+1}^T = AX_nX_n^TA^T + V_nV_n^T + AX_nV_n^T + V_nX_n^TA^T.$$

Feltételes várható értéket véve $\mathcal{G}_n$-re használjuk ki, hogy

$$\mathrm{E}(AX_nV_n^T|\mathcal{G}_n) = AX_n\mathrm{E}(V_n^T|\mathcal{G}_n) = 0.$$

Emiatt a keresztszorzatok kiesnek, és azt kapjuk, hogy

$$\mathrm{E}(X_{n+1}X_{n+1}^T|\mathcal{G}_n) = AX_nX_n^TA^T + \mathrm{E}(V_nV_n^T|\mathcal{G}_n). \tag{9}$$

Ez egy feltételes Ljapunov-egyenlet a második momentumra, ami még nincs normálva. (Várható értéket ki kell vonni.)

Tegyük fel, hogy az A mátrix primitív, azaz van olyan hatványa, melynek minden eleme pozitív. Ekkor a Markov-folyamatok elméletéből ismert, hogy egyértelműen létezik invariáns eloszlás $\mathcal{X}$-en. Könnyen látható, hogy erre az eloszlásra $A\pi = \pi$. Legyen ugyanis $P(X_0 = e_i) = \pi_i$. Ekkor

$$\pi_j = P(X_1 = e_j) = \sum_{i=1}^{N} P(X_1 = e_j|X_0 = e_i)P(X_0 = e_i) = \sum_{i=1}^{N} A_{ji}\pi_i.$$

4.5. Állítás.

$$\mathrm{E}X_n = \pi, \qquad \mathrm{E}X_nX_n^T = \mathrm{diag}\ \pi.$$

Bizonyítás. Egyszerű számolással adódik, hogy

$$\mathrm{E}X_n = \sum_{i=1}^{N} e_iP(X_n = e_i) = \pi.$$

Minden egységvektorra teljesül, hogy

$$e_ie_i^T = \mathrm{diag}\ (e_i),$$

amiből az állítás második része következik.

Ezek után kiszámoljuk az autokovarianciafüggvényt.

4.6. Állítás.

$$R(0) = \mathrm{E}(X_n - \pi)(X_n - \pi)^T = (I - \pi\mathbf{1}^T)(\mathrm{diag}\ \pi).$$

Bizonyítás. Vegyünk várható értéket a (9) egyenletben:

$$(\mathrm{diag}\ \pi) = A\ (\mathrm{diag}\ \pi)\ A^T + \Sigma_V.$$

Ezt felhasználva

$$\mathrm{E}(X_n - \pi)(X_n - \pi)^T = (\mathrm{diag}\ \pi) - \mathrm{E}\pi X_n^T - \mathrm{E}X_n\pi^T + \pi\pi^T =$$

$$= (\mathrm{diag}\ \pi) - \pi\pi^T = (I - \pi\mathbb{1}^T)(\mathrm{diag}\ \pi).$$

Itt felhasználtuk, hogy $\pi^T = \mathbf{1}^T(\mathrm{diag}\ \pi)$.

Hasonlóan számolható az autokovarianciafüggvény általános esetben:

4.7. Állítás. *Tetszőleges $\tau \geq 1$ természetes számra*

$$R(\tau) = \mathrm{E}(X_{n+\tau} - \pi)(X_n - \pi)^T = (A^\tau - \pi\mathbf{1}^T)(\mathrm{diag}\ \pi).$$

Bizonyítás. Fokozatos behelyettesítéssel kapjuk, hogy

$$X_{n+\tau} = A^\tau X_n + \sum_{i=n+1}^{n+\tau} A^{n+\tau-i}V_{i-1}.$$

Szorozzuk meg az egyenletet jobbról X_n^T-nel, majd vegyünk várható értéket. Az utolsó tag kiesik, mert

$$\mathrm{E}(V_{i-1}X_n^T) = \mathrm{E}\left(\mathrm{E}(V_{i-1}X_n^T|\mathcal{F}_n^{X-})\right) = \mathrm{E}\left(\mathrm{E}(V_{i-1}|\mathcal{F}_n^{X-})X_n^T\right) = 0.$$

Tehát azt kapjuk, hogy

$$\mathrm{E}X_{n+\tau}X_n^T = A^\tau \mathrm{E}X_nX_n^T.$$

Kis módosítással az adódik, hogy

$$\mathrm{E}(X_{n+\tau} - \pi)(X_n - \pi)^T = A^\tau\mathrm{E}X_nX_n^T - \pi\pi^T = A^\tau\ (\mathrm{diag}\ \pi) - \pi\mathbf{1}^T\ (\mathrm{diag}\ \pi).$$

5. Szűrés

Ebben a fejezetben visszatérünk arra a modellre, amikor az (X_n, Y_n) rejtett Markov folyamat értékeit az $\mathcal{X} = \{1, \dots, N\}$ és $\mathcal{Y} = \{1, \dots, M\}$ térből vesszük.

A szűrés feladata a következő: Adott $Y_{n-1}, \dots Y_0$ megfigyelés alapján adjunk becslést X_n értékére. Pontosabban, a *prediktív szűrő* az alábbi valószínűségi változó megadását jelenti:

$$P(X_n = j | Y_{n-1}, \dots, Y_0).$$

Mivel atomos σ-algebrákkal dolgozunk, elegendő a következő valószínűségeket megadni:

$$P(X_n = j | Y_{n-1} = y_{n-1}, \dots, Y_0 = y_0).$$

Ezt rekurzív módon fogjuk elvégezni. Definiáljuk az alábbi halmazokat:

$$\begin{aligned} A &= \{Y_{n-1} = y_{n-1}, \dots, Y_0 = y_0\} \\ B_i &= \{X_n = i\} \\ C &= \{Y_n = y_n\} \\ D &= \{X_{n+1} = x_{n+1}\}. \end{aligned}$$

Feladatunk így fogalmazható meg: *ismert* $P(B_i|A)$ minden i-re, *kiszámítandó* $P(D|CA)$.

Első lépésként a teljes valószínűség tétel feltételes eloszlásokra átfogalmazott alakját használjuk. Eszerint

$$P(D|CA) = \sum_{i=1}^{N} P(D|B_iCA)P(B_i|CA). \tag{10}$$

5.1. Állítás.

$$P(D|B_iCA) = P(D|B_i).$$

Bizonyítás. Rejtett Markov folyamatokra tudjuk, hogy

$$P(X_{n+1} = x_{n+1} | \mathcal{F}_n^{X-} \vee \mathcal{F}_n^{Y-}) = P(X_{n+1} = x_{n+1} | \mathcal{F}^{X_n}).$$

Ezt az összefüggést atomokra kiírva a kívánt állítást kapjuk.

A szokásos jelölésekkel

$$P(D|B_i) = P(X_{n+1} = x_{n+1}|X_n = i) = Q_{ix_{n+1}},$$

tehát azt kapjuk, hogy

$$P(D|B_iCA) = Q_{ix_{n+1}}. \tag{11}$$

Következő lépésként a Bayes-formula feltételes alakját használjuk. Eszerint

$$P(B_i|CA) = \frac{P(C|B_iA)P(B_i|A)}{P(C|A)}.$$

Ezt fogjuk visszahelyettesíteni a (10) formulába. Mivel a nevező, $P(C|A) = \gamma_n$ i-től független , ezért ez az összegzésből kiemelhető. A számlálóban levő értékek

$$P(B_i|A) = P(X_n = i|Y_{n-1} = y_{n-1}, \dots, Y_0 = y_0)$$

a feltétel szerint ismertek. Végül az utolsó tényező

$$P(C|B_iA) = P(Y_n = y_n|X_n = i, Y_{n-1} = y_{n-1}, \dots, Y_0 = y_0) =$$

$$= P(Y_n = y_n|X_n = i) = b^i(y_n)$$

a Markov tulajdonság miatt.

Összegezve a fenti számításokat, azt kapjuk, hogy

$$P(X_{n+1} = j|Y_n = y_n, \dots, Y_0 = y_0) = \sum_{i=1}^{N} Q_{ij} b^i(y_n) P(B_i|A)/\gamma_n.$$

Vezessük be a

$$p^j_{n+1} = P(X_{n+1} = j|Y_n, \dots, Y_0)$$

jelölést, illetve ennek vektor alakját:

$$p_{n+1} = (p^1_{n+1}, \dots, p^N_{n+1})^T.$$

Megjegyezzük, hogy Y egy rögzített realizációjával dolgozunk. Ekkor a fenti egyenlet így írható:

$$p^j_{n+1} = \sum_{i=1}^{N} Q_{ij} b^i(Y_n) p^i_n / \gamma_n.$$

További jelöléseket vezetünk be:

$$B(y) = \mathrm{diag}\ (b^i(y)),$$

illetve tetszőleges $x \neq 0 \ \epsilon\ \mathbb{R}^N$ vektor esetén $\pi(x)$ jelöli ennek vetületét az egységgömbön:

$$\pi(x)_i = \frac{x_i}{\sum_{j=1}^{N} x_j}.$$

Tehát beláttuk a következő tételt:

5.1. Tétel. *(Baum-egyenlet) A fenti jelölésekkel:*

$$p_{n+1} = \pi(Q^* B(Y_n) p_n). \tag{12}$$

A γ_n normalizáló tényezőt vizsgáljuk.

5.2. Állítás.

$$\gamma_n = b^T(Y_n) p_n,$$

ahol $b^T(y) = (b^1(y), \ldots, b^N(y))$.

Az állításra kétféle bizonyítást adunk.

1.Bizonyítás. $Q^* B(Y_n) p_n$ normalizáló tényezője

$$\mathbf{1}^T Q^* B(Y_n) p_n.$$

Mivel Q sor-sztochasztikus mátrix, ezért Q^* oszlop-sztochasztikus, ezért $\mathbf{1}^T Q^* = \mathbf{1}^T$. Így az állítás abból következik, hogy $\mathbf{1}^T B(Y_n) = b^T(Y_n)$.

2.Bizonyítás. Legyen $Y_n = j$. Ekkor

$$\gamma_n = P(Y_n = j | Y_{n-1}, \dots, Y_0) =$$

$$= \sum_{i=1}^{N} P(Y_n = j | X_n = i, Y_{n-1}, \dots, Y_0) P(X_n = i | Y_{n-1}, \dots, Y_0) =$$

$$= \sum_{i=1}^{N} P(y_n = j | X_n = i) p_n^i = \sum_{i=1}^{N} b^i(j) p_n^i = b^T(j) p_n.$$

Ez pedig igaz minden $j = 1, \dots, M$ esetén.

Kérdés, hogy a Baum-egyenlet indítását honnan kezdjük, hiszen X_0 eloszlását nem ismerjük.

Mielőtt erre válaszolnánk felírjuk a Baum-egyenlet *nem prediktív* alakját. Ehhez vezessük be az alábbi jelöléseket:

$$q_n^i = P(X_n = i \mid Y_n, \dots, Y_0), \qquad q_n = (q_n^1, \dots, q_n^N)^T.$$

5.3. Állítás.

$$q_n = \pi(B(Y_n) Q^* q_{n-1}).$$

Bizonyítás. A korábbi jelöléseket használjuk. Legyenek $X_n = i$, $Y_n = y_n, \dots, Y_0 = y_0$ rögzítettek. Ekkor

$$q_n^i = P(B_i | CA) = \frac{P(C | B_i A) P(B_i | A)}{P(C | A)}.$$

A markovitás miatt $P(C|B_i A) = P(C|B_i)$, és $P(C|A) = \gamma_n$. Azt kapjuk, hogy

$$q_n^i = b^i(y_n) p_n^i / \gamma_n,$$

azaz vektoriálisan

$$q_n = B(y_n) p_n / \gamma_n.$$

Innen kis számolással az állítás következik.

Megjegyzés. A fenti rekurzióban a műveletek során felváltva történik szorzás és normálás. Mivel a normálás során az egységmátrix egy többszörösével szorzunk, elegendő csak a végén normálni. általában is igaz pozitív elemű A, B mátrixok és pozitív p vektor esetén, hogy

$$\pi(A\pi(Bp)) = \pi(ABp).$$

Ennek megfelelően a fenti formula így írható:

$$p_n = \pi\left(Q^* B(Y_n) \dots Q^* B(Y_0) p_0\right).$$

Példa. Tekintsük azt a példát, amikor a kiolvasás egyenletes valószínűséggel történik, azaz

$$P(Y_n = y_n | X_n = x_n) = \frac{1}{M}, \qquad \forall (x_n, y_n).$$

5.4. Állítás. *Ekkor*

$$p_n^i = P(X_n = i).$$

Bizonyítás. A (12) egyenletet használjuk. Mivel $B(Y_n) = \frac{1}{M} I$, ezért a normalizáláskor ez a tag kiesik. Így

$$p_{n+1} = \pi(Q^* p_n) = Q^* p_n,$$

hiszen Q^* sztochasztikus mátrix, és nem kell normalizálni.

5.5. Állítás. *Ebben az esetben* $(X_1, \dots, X_n)$ *és* $(Y_1, \dots, Y_n)$ *teljesen függetlenek.*

Bizonyítás.

$$P(Y_n = y_n, \dots, Y_0 = y_0, X_n = x_n, \dots, X_0 = x_o) =$$

$$= \prod_{i=0}^{n} P(Y_i = y_i | X_i = x_i) P(X_n = x_n, \dots, X_0 = x_0) =$$

$$\left(\frac{1}{M}\right)^{n+1} P(X_n = x_n, \dots, X_0 = x_0)$$

Ezt az egyenlőséget összegezzük $\{x_o, \ldots, x_n\}$-re, azt kapjuk, hogy

$$P(Y_n = y_n, \ldots, Y_0 = y_0) = \left(\frac{1}{M}\right)^{n+1},$$

amit visszaírva az eredeti egyenletbe a függetlenséget kapjuk.

Következmény. Ekkor X és Y kölcsönös infromációja 0,

$$I(X \wedge Y) = 0.$$

6. Visszafelé szűrés

Fix adattömb esetén, ha $X_0, \ldots, X_N$ Markov folyamat, akkor $X_N, \ldots, X_0$ is az. Ekkor az alábbi valószínűséget szeretnénk meghatározni:

$$\overline{p}_n^i = P(X_n = i | Y_{n+1}, \ldots, Y_T).$$

A Markov-lánc átmeneti valószínűségei

$$\overline{Q}_{ji} = P(X_n = i | X_{n+1} = j) = \frac{P(X_{n+1} = j | X_n = i) P(X_n = i)}{P(X_{n+1} = j)}.$$

Tegyük fel, hogy X-nek létezik stacionárius eloszlása, π. Ekkor a fenti reláció röviden így írható:

$$\overline{Q}_{ji} = Q_{ij} \frac{\pi_i}{\pi_j}.$$

Megjegyezzük, hogy ez könnyebben megjegyezhető a $\pi_j \overline{Q}_{ji} = \pi_i Q_{ij}$ formában.

Ez alapján a visszefelé szűrés egyenlete felírható.

7. Inicializálás

Azt a problémát vizsgáljuk meg, hogy a Baum egyenlet indításához használt

$$p_0^i = P(X_0 = i)$$

valószínűségek megválasztása mennyiben befolyásolja az egyenlet aszimptotikus viselkedését. Tekintsünk egy példát, amelyen jól látható, mi okozhat gondot.

Példa. Kvantálás. Legyen $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$ a Markov lánc állapottere. A kiolvasás a következő:

$$Y_n = \begin{cases} 0 & \text{ha } X_n \in \mathcal{X}_1 \\ 1 & \text{ha } X_n \in \mathcal{X}_2 \end{cases} .$$

Tegyük fel, hogy a kiindulás determinisztikus, $X_0 = x_0 \in \mathcal{X}_1$. Ha ezek után a p_0 valószínűséget az $\mathcal{X}_2$-re koncentráltan választjuk, akkor $b(Y)p_0 \equiv 0$ lesz, tehát nem tud az algoritmus elindulni.

Feltétel. Tegyük fel, hogy a kezdeti eloszlás kompatibilis Y_0-lal; ez alatt a következőt értjük. Legyen

$$q_0 := B(Y)p_0, \qquad q \geq 0,$$

ekkor feltesszük, hogy

$$Q^* q_0 > 0.$$

8. ML becslés

Legyen X_n véges állapotterű Markov lánc, állapottere $\mathcal{X} = \{1, \ldots, N\}$. Az átmenet valószínűségek $a_{ji} = P(X_{n+1} = j | X_n = i)$. Az $A = (A_{ji})$ mátrix elemeit szeretnénk megbecsülni egy $X_0, \ldots X_n$ megfigyelés alapján.

A megfigyelés adott realizációjának valószínűsége:

$$P(X_n = x_n, \ldots, X_0 = x_0) = \prod_{t=1}^{n} a_{x_t x_{t-1}} \cdot P(X_0 = x_0).$$

Számláljuk össze, hogy bizonyos átmenetek hányszor fordulnak elő:

$$\begin{aligned} J_{sr} &= \#\{t : \ X_t = s, X_{t-1} = r, 1 \leq t \leq n\} \\ O_r &= \#\{t : \ X_t = r, 0 \leq t \leq n-1\}. \end{aligned}$$

Gyűjtsük össze ezeket az értékeket egy $J = (J_{sr})$ mátrixba ill. $O = (O_r)$ vektorba. Ekkor nyilván $\mathbf{1}^T J = O$. Természetes választás, hogy $P(X_{t+1} = s, X_t = r)$ becslése legyen J_{sr}/n, $P(X_t = r)$ becslése legyen O_r/n. Ez alapján legyen

$$\widehat{a}_{sr} = \begin{cases} \dfrac{J_{sr}}{O_r} & \text{ha } J_{sr} \neq 0 \\ \\ 0 & \text{ha } J_{sr} = 0. \end{cases} \qquad (13)$$

8.1. Állítás. *A fenti (13) becslés a megfelelő paraméterek maximum likelihood becslésével esik egybe.*

Bizonyítás. Legyen E a következő esemény:

$$E = \{(s, r) : \ J_{sr} = 0\}.$$

A log-likelihood függvényt így írhatjuk fel:

$$\log P(x_n, \ldots, x_0) = C + \sum_{s,r} J_{sr} \log a_{sr},$$

ahol C a kezdetiértékből ered, a második tagban pedig felhasználtuk, hogy

$$\prod_{t=1}^{n} a_{x_t x_{t-1}} = \prod_{s,r} a_{sr}^{J_{sr}}.$$

Tegyük fel először, hogy $J_{sr} > 0$ minden (s, r)-re. Ekkor a feladat egy feltételes optimalizálás:

$$\max_{\{\sum_s a_{sr}=1, \ r=1,\ldots,M\}} (\log P).$$

A Lagrange féle szükséges feltétel szerint az alábbi függvény stacionárius pontjait kell meghatározni:

$$\log P - \sum_{r=1}^{M} \lambda_r \left(\sum_{s=1}^{N} a_{sr} - 1 \right).$$

A gradiens zérushelyét meghatározó egyenletrendszer:

$$J_{sr} \frac{1}{a_{sr}} - \lambda_r = 0, \qquad r = 1, \ldots M,$$

ahonnan

$$\widehat{a}_{sr} = \frac{J_{sr}}{\lambda_r}$$

adódik. Szorozzuk meg ezt az egyenletet balról $\mathbf{1}^T$-tal:

$$\mathbf{1}^T \widehat{a}_{sr}^T = \mathbf{1}^T J_{sr}/\lambda_r,$$

ahonnan

$$\mathbf{1}^T \widehat{a}_{sr}^T = 1, \qquad \mathbf{1}^T J_{sr} = O_r$$

miatt $\lambda_r = O_r$ adódik.

Ezzel az állítást beláttuk.

8.1. Következmény. *Az átmeneti mátrix ML becsése*

$$\widehat{A} = J(\text{diag } O)^{-1},$$

ahol

$$\text{diag } O = \begin{pmatrix} O_1 & & \\ & \ddots & \\ & & O_n \end{pmatrix}.$$

Térjünk át most újra arra a reprezentációra, amikor az állapottér $\mathcal{X} = \{e_1, \dots, e_N\}$. Ekkor a Markov láncnak megadható

$$X_{k+1} = AX_k + V_k$$

reprezentációja, ahol V_k martingál-differencia sorozat. Az A átmeneti mátrix becslését fogjuk meghatározni. Az IV (instrumentális változók módszere) módszert fogjuk használni. Szorozzuk meg a fenti egyenletet jobbról X_k^T-tal:

$$X_{k+1}X_k^T = AX_kX_k^T + V_kX_k^T.$$

Összegezzük $k = 0, \dots n-1$ mellett. A baloldal:

$$\left(\sum_{k=0}^{n-1} X_{k+1}X_k^T \right)_{sr} = \sum_{k=0}^{n-1} \langle X_{k+1}, e_s \rangle \langle X_k, e_r \rangle = J_{sr},$$

míg a jobboldal

$$\sum_{k=0}^{n-1} X_kX_k^T = \sum_{k=0}^{n-1} \text{diag } (X_k) = \text{diag } (\sum_{k=0}^{n-1} X_k),$$

és ennek r-dik eleme

$$\left(\sum_{k=0}^{n-1} X_k\right)_r = O_r.$$

Így

$$J = A^*\mathrm{diag}\ O + \sum_{k=0}^{n-1} V_k X_k^T.$$

A jobboldal második tagja a nagy számok törvénye szerint 0-hoz tart. így a kapott becslés:

$$\widehat{A} = J(\mathrm{diag}\ O)^{-1}.$$

Ennek statisztikai tulajdonságait nézzük meg. Ha az igazi paramétert A^* jelöli, akkor az összefüggések:

$$\begin{aligned} J &= A^*\ \mathrm{diag}\ (O) + \sum_{k=0}^{n-1} V_k X_k^T \\ J &= \widehat{A}\ \mathrm{diag}\ (O). \end{aligned}$$

Ebből

$$\widehat{A} - A^* = \sum_{k=0}^{n-1} V_k X_k^T\ (\mathrm{diag}\ (O))^{-1}.$$

Ebből kiindulva fogjuk a becslés aszimptotikus tulajdonságait meghatározni.

Rögzített (s, r) index-pár mellett legyen

$$Z_k = (V_k X_k^T)_{sr}.$$

Ekkor

$$\mathrm{E}(Z_k^2|\mathcal{F}_k) = \mathrm{E}((v_k x_k)^2|\mathcal{F}_k) = x_k^2\mathrm{E}(v_k^2|\mathcal{F}_k) \le K < \infty$$

teljesül valamely k-tól független K konstans mellett.

Általánosan kimondunk egy tételt martingálok konvergenciájára vonatkozóan:

8.2. Állítás. *Legyen $(Z_k, \mathcal{F}_k)$ martingál-differencia sorozat, melyre*

$$\mathrm{E}(Z_k^2|\mathcal{F}_k) \le K < \infty,$$

k-tól függetlenül. Legyen továbbá $\rho(n)$ egy monoton növő számsorozat, melyre $\sum \rho(n)^{-1} < \infty$. Ekkor

$$\frac{1}{\rho(n)^{1/2}} \sum_{k=1}^{n} Z_k \to 0.$$

Bizonyítás. Tekinsük az

$$S_n = \sum_{k=1}^{n} \rho(k) Z_k$$

sorozatot. Nyilván S_n $\mathcal{F}_n$-martingál. Második momentuma:

$$\mathrm{E}(S_n^2) = \mathrm{E}(\sum_{k,l} \rho(k)^{-1/2} \rho(l)^{-1/2} Z_k Z_l).$$

A várható értéken belül a (k, l)-dik tagban vegyünk először feltételes várható értéket $\mathcal{F}_{l \wedge k}$ σ-algebrára. Könnyen látható, hogy ha $k \neq l$, akkor

$$\mathrm{E}(Z_k Z_l | \mathcal{F}_{l \wedge k}) = 0.$$

Így a fenti összegből ez marad:

$$\mathrm{E}(S_n^2) = \mathrm{E}\left((\sum_{k=1}^{n} \rho(k)^{-1} \mathrm{E}(Z_k^2 | \mathcal{F}_k) \right) = \sum_{k=1}^{n} \rho(k)^{-1} \mathrm{E}(Z_k^2) \leq K \sum_{k=1}^{\infty} \rho(k)^{-1} < \infty.$$

Ebből az is következik, hogy

$$\sup \mathrm{E} S_n^2 < \infty.$$

Felhasználjuk az alábbi tételt:

8.1. Tétel. *(Hull-Heyde, 20. old.) Ha (S_n) martingál, melyre*

$$\sup_n |S_n|^p < \infty$$

valamely $p > 1$ mellett, akkor $\lim_{n \to \infty} S_n$ létezik 1 valószínűséggel.

A fenti tételt alkalmazhatjuk $p = 2$-re. Ezek után a Kronecker-lemmából következik az állítás.

Alkalmazzuk az állítást $\rho(n) = n^2$ választással. Ekkor az

$$\widehat{A} - A^* = \sum_{k=0}^{n-1} V_k X_k^T \; (\mathrm{diag}\ (O))^{-1}$$

egyenletben a jobboldal első tényezője n-nel normálva 0-hoz tart. A második tényező viselkedését csak bizonyos további feltételek mellett tehetjük meg.

8.3. Állítás. *Tegyük fel, hogy (X_n) irreducibilis és aperiodikus Markov lánc. Ekkor létezik egyértelmű invariáns π eloszlás. Továbbá*

$$\frac{1}{n}O_r \to \pi(r)$$

1 valószínűséggel.

Ebből tehát következik, hogy a fenti feltételek mellett

$$\widehat{A} - A^* \to 0$$

1 valószínűséggel.

9. Rejtett Markov-folyamat becslése

Az előző fejezetben meghatároztuk Markov lánc paramétereinek ML becslését. Ezt fogjuk kiterjeszteni arra az esetre, amikor rejtett Markov folyamatot tekintünk.

Legyen (X_n, Y_n) rejtett Markov-folyamat. A diszkrét állapotterek $\mathcal{X} = \{1, \ldots, N\}$, $\mathcal{Y} = \{1, \ldots, M\}$. Tegyük fel (irreális feltevés), hogy (X_n) ismert. Ekkor átmeneti valószínűségei becsülhetőek. A kiolvasási valószínűségek becslését fogjuk megadni. Ezek:

$$C_{ir} = P(Y_n = r | X_n = i).$$

Adottak $(x_1, \ldots, x_n)$, $(y_1, \ldots, y_{n-1})$. A hozzá tartozó likelihood függvény:

$$P = \prod_{t=1}^{n} a_{x_t x_{t-1}} \prod_{t=0}^{n-1} c_{y_t x_t} p(x_0).$$

összegyűjtjük az azonos tagokat:

$$P = \prod_{(i,j)} a_{ji}^{J_{ji}} \prod_{(r,i)} c_{ri}^{T_{ri}} p(x_o),$$

ahol

$$T_{ri} = \#\{t: \ 0 \leq t \leq n-1, x_t = i, y_t = r\}.$$

Ez alapján a log-likelihood függvény:

$$logP = \sum_{j,i=1}^{N} J_{ji} \log a_{ji} + \sum_{r=1}^{M} \sum_{i=1}^{N} T_{ri} \log c_{ri} + \log p(x_0).$$

Így a ML becslés az alábbi feltételes optimalizálási feladat megoldása:

$$\max_{a,c} \log P \qquad \text{ha} \qquad \begin{array}{ll} a_{ji} \geq 0 & \sum_j a_{ji} = 1 \quad \forall i \\ c_{ri} \geq 0 & \sum_r c_{ri} = 1 \quad \forall i \end{array}.$$

Tehát a becslés meghatározása két olyan különálló optimalizálási feladatra esik szét, mint amit az előző fejezetben oldottunk meg. Ebből következik az alábbi állítás:

9.1. Állítás. *A fenti feltételek mellett C ML becslése*

$$\widehat{C} = T(\text{diag } O)^{-1}.$$

Megjegyzés. A bizonyítás során itt is a $\{T_{ir} > 0\}$ és $\{T_{ir} = 0\}$ eseményeket külön kell kezelni.

A ML becslés más levezetését kaphatjuk meg az állapotteres reprezentáció alapján. A rejtett Markov lánc kiolvasási egyenletét így reprezentáltuk:

$$Y_k = CX_k + W_k, \tag{14}$$

ahol

$$\mathrm{E}(W_k | \mathcal{F}_X^{k-} \wedge \mathcal{F}_Y^{k-}) = 0.$$

C meghatározását lineáris regressziós feladatként tekintjük, és az IV módszert alkalmazzuk. Szorozzuk meg a (14) egyenletet X_k^T-tal, és összegezzünk $k = 0, \ldots n-1$-re. Ebből

$$\sum_{k=0}^{n-1} \widehat{C}(X_k X_k^T) = \sum_{k=0}^{n-1} Y_k X_k$$

adódik, felhasználva hogy az utolsó tag várható értéke 0. Könnyen látható, hogy az így kaptt becslés egybeesik a fenti állításban szereplővel. Nagy előnye ennek az algebrai reprezentációnak, hogy a hibatagot pontosan ismerjük:

$$(\widehat{C} - C)\text{diag } O = \sum_{k=0}^{n-1} W_k X_k^T.$$

Ennek nagyságrendje ugyanúgy kezelhető, mint az előző fejezetben.

10. Az EM módszer

Rátérünk arra a valós feladatra, hogy a rejtett Markov folyamat paramétereit a kiolvasás ismeretében kell megbecsülni. Az EM (Expectation Maximization) módszert tipikusan ilyen feladatok megoldására fejlesztették ki; amikor parciális információ alapján kell paraméterbecslést elvégezni. Szokás erre a módszerre úgy is hivatkozni, mint *Baum-Welch algoritmus.*

Általános Baum-Welch algoritmus. Adottak az X, Y valószínűségi változók , értékkészletük $\mathcal{X}$ ill. $\mathcal{Y}$. Együttes eloszlásuk $p(x, y, \theta^*)$, ahol θ^* ismeretlen paraméter. Megfigyeljük az $\{Y = y\}$ eseményt. A teljes log-likelihood függvény $\log P(x, y, \theta)$ lenne, de x ismeretlen. Ehelyett tekintjük az alábbi függvényt:

$$Q(\theta, \overline{\theta}) = \mathrm{E}\left(\log P(X, Y, \theta)|Y = y, \overline{\theta}\right).$$

A fenti kifejezésben θ és $\overline{\theta}$ általában különböző.

Most nézzük meg a fenti $Q(\theta, \overline{\theta})$ és az információelméletből jólismert entrópia függvény közötti kapcsolatot. Induljunk ki a

$$P(x, y, \theta) = P(x|y, \theta)P(y, \theta)$$

azonosságból. Vegyük mindkét oldal logaritmusát, majd szorozzuk mindkét oldalt $P(x|y, \overline{\theta})$-val, végül mindkét oldalt összegezzük x-ben. Így átrendezve kapjuk, hogy

$$\sum_x P(x|y, \overline{\theta}) \log P(y, \theta) = \sum_x P(x|y, \overline{\theta}) \log P(x, y, \theta) - \sum_x P(x|y, \overline{\theta}) \log P(x|y, \theta)$$

Felhasználva, hogy az entrópiafüggvény

$$H(\theta, \overline{\theta}) = \sum_x -P(x|y, \overline{\theta}) \log P(x, y, \theta),$$

kapjuk, hogy

$$\log P(y, \theta) = Q(\theta, \overline{\theta}) + H(\theta, \overline{\theta})$$

Másképpen felírva

$$Q(\theta, \overline{\theta}) = -H(\theta, \overline{\theta}) + \log P(y, \theta)$$

Megjegyezzük, hogy $\theta = \overline{\theta}$ esetén Q és H véges.

10.1. Állítás. *Legyen* $\overline{\theta}$ *olyan, hogy* $\log P(y,\overline{\theta}) > -\infty$, *valamint* θ *olyan, hogy* $Q(\theta,\overline{\theta}) > Q(\overline{\theta},\overline{\theta})$. *Ekkor* $\log P(y,\theta) > \log P(y,\overline{\theta})$.

Bizonyítás.

$$\log P(y,\overline{\theta}) = Q(\overline{\theta},\overline{\theta}) + H(\overline{\theta},\overline{\theta})$$

$\theta = \overline{\theta}$ esetén,

$$Q(\theta,\overline{\theta}) > Q(\overline{\theta},\overline{\theta})$$

a feltételből, illetve a Kullbach-Leifer alapján

$$H(\theta,\overline{\theta}) \geq H(\overline{\theta},\overline{\theta}),$$

amiből rögtön adódik az állítás.

Mit is jelent konkrétan az EM módszer a jelen esetben. Kiindulunk egy $\overline{\theta}_0$ paraméterből, majd ebből kapjuk a

$$\overline{\theta}_1 = \max_{\theta} Q(\theta,\overline{\theta})$$

első iterációt,...

Kérdés, hogy ezzel az iterációsorozattal eljutunk-e a maximum likelihood becsléshez, azaz $\overline{\theta}_k \longrightarrow \widehat{\theta}$ teljesül-e, ahol $\widehat{\theta} = \max_{\theta} \log P(y,\theta)$

Azonban mint ahogy általában, most is van feltétel a paraméterre vonatkozóan, így az iterációs lépéskor a maximalizálás igazából egy feltételes maximalizálást jelent.

Baum-Welch algoritmus: Tekintsük az (X_n, Y_n) Rejtett Markov folyamatot. Álljon a paramétertér az átmenetvalószínűségekből és a kiolvasási valószínűségekből, azaz $\theta = (A, C)$. (Legyen $\mathcal{X}$ illetve $\mathcal{Y}$ véges).

Mintmár korábban láttuk, a teljes loglikelihood függvény

$$\log P(x,y,\theta) = \sum_{s,r} J^{sr} \log a_{sr} + \sum_{jr} T^{jr} \log c_{jr}$$

és a paraméterre vonatkozó feltételek

$$a_{sr} \geq 0 \quad \sum_{s} a_{sr} = 1 \quad \forall r,$$

illetve

$$c_{jr} \geq 0 \quad \sum_j c_{jr} = 1 \quad \forall r.$$

Mi most az EM algoritmus?

$x = (x_0, \dots x_n)$ és $y = (y_0, \dots y_n)$ jelölések mellett

$$\overline{J}^{sr} = E(y^{sr}|y, \overline{\theta}) \quad \overline{T}^{jr} = E(T^{jr}|y, \overline{\theta}),$$

$$Q(\theta, \overline{\theta}) = \sum_{sr} \overline{J}^{sr} \log a_{sr} + \sum_{jr} \overline{T}^{jr} \log c_{jr}$$

Az M-lépés eredménye (azaz amikor maximalizálunk θ-ban)

$$a_{sr} = \frac{\overline{J}^{sr}}{\overline{O}^{r}}, \text{ ahol } \overline{O}^{r} = \sum_s \overline{J}^{sr},$$

illetve

$$c_{jr} = \frac{\overline{T}^{jr}}{\overline{O}^{r}}$$

Még az a kérdés persze hátra van, hogy hogyan számoljuk $\overline{J}^{sr}$ értékét, azaz hogyan számoljuk a $E(J^{sr}|y_{n-1}, \dots y_0, \overline{\theta})$ feltételes várhatóértéket?

Vegyük észre, hogy $J^{sr} = \sum\limits_{t=1}^{n-1} < X_{t+1}, e_s >< X_t, e_r >$ a vektoros felírásban. Így elegendő az

$$E(< X_{t+1}, e_s >< X_t, e_r > |y_{n-1} \dots y_0, \overline{\theta})$$

feltételes várhatóértéket kiszámolni, ami nem más, mint a

$$P(X_{t+1} = s, X_t = r|y_{n-1}, \dots, y_0, \overline{\theta})$$

feltételes várhatóérték.

Legyenek A, B, C, D a következő események:

$$A = \{Y_t = y_t, \dots Y_0 = y_0\}$$

$$B = \{X_t = r\}$$
$$C = \{X_{t+1} = s\}$$
$$D = \{Y_{n-1} = y_{n-1}, \ldots, Y_{t+1} = y_{t+1}\}$$

(a $t \leq n-2$ feltehető)

Feladatunk $P(CB|DA)$ meghatározása.

Azt tudjuk, hogy

$$P(CB|DA) = P(DA|CB)\frac{P(CB)}{P(DA)}$$

a feltételes valószínűségre vonatkozó egyszerű egyenlőtlenség alapján. Ezen túlmenően még szükségünk van a következő állításra.

10.2. Állítás. *A Rejtett Markov folyamat definíciója alapján*

$$P(DA|CB) = P(D|C)P(A|B)$$

Bizonyítás. A feltételes függetlenség miatt

$$P(DA|CB) = \frac{P(DAC|B)}{P(C|B)} = P(DC|B)\frac{P(A|B)}{P(C|B)},$$

ugyanakkor a Markovitás miatt

$$P(DC|B) = P(D|CB)P(C|B) = P(D|C)P(C|B),$$

ahonnan következik az állítás.

Innen egyszerű számolással adódik, hogy a keresett $P(CB|DA)$ mennyiséget a következő módon kell kiszámolni:

$$P(CB|DA) = P(C|D)P(D)P(B|A)P(A)\frac{\frac{P(CB)}{P(C)P(B)}}{P(DA)}$$

11. Kvantált Gauss AR folyamatok

Tekintsünk egy elsőrendű AR folyamatot

$$y_n + a^* y_{n-1} = e_n, \qquad n = 1, \dots, N,$$

ahol (e_n) független azonos $\mathcal{N}(0, \sigma^2)$ eloszlású valószínűségi változó sorozat. A stacionaritáshoz feltesszük, hogy $|a^*| < 1$. a^* becslése teljes megfigyelés alapján elvégezhető például a ML becsléssel a következő képpen. (Pontosabban, a feltételes ML becslést mutatjuk be, $y_0 = 0$ feltétel mellett.)

Tetszőleges $|a| < 1$ mellett definiáljuk az

$$\varepsilon_n(a) = y_n + a y_{n-1}, \qquad n = 1, 2, \dots$$

folyamatot. Mivel az $y^N = (y_1, \dots, y_N)$ vektort az $\varepsilon^N = (\varepsilon_1, \dots, \varepsilon_N)$ vektorba vivő transzformáció determinánsa 1, ezért y^N sűrűségfüggvénye megegyezik ε^N sűrűségfüggvényével. Igy

$$p(\varepsilon^N, a) = \frac{1}{(2\pi)^{N/2} \sigma^N} \exp\left\{ -\sum_{n=1}^{N} \varepsilon_n^2(a) / 2\sigma^2 \right\} = p(y^N, a).$$

Az egyszerűség kedvéért feltesszük, hogy σ^2 ismert. Ekkor a ML becsléshez $\log p(y^N, a)$ gradiensének zérushelyét kell meghatározni:

$$\frac{\partial}{\partial a} \big(\log p(y^N, a) \big) = 0 = \sum_{n=1}^{N} \varepsilon_n(a) y_{n-1} / \sigma^2.$$

ε_n definícióját beírva megkapjuk a ML becsést:

$$\widehat{a} = \frac{\sum_{n=1}^{N} y_n y_{n-1}}{\sum_{n=1}^{N} y_{n-1}^2}.$$

Megjegyezzük, hogy ebben az egyszerű esetben a fenti becslés egybeesik a LS becsléssel.

Tekintsük most azt az esetet, amikor kvantált megfigyelésünk van. Ekkor az AR folyamat

$$\xi_n + a^* \xi_{n-1} = e_n,$$

és a megfigyelt értékek:

$$y_n = q(\xi_n),$$

ahol

$$q(\xi) = nh, \qquad \text{ha} \qquad nh - \frac{1}{2}h \le \xi < nh + \frac{1}{2}h.$$

Ekkor (y_n) rejtett Markov folyamat, hiszen a (ξ_n) Markov folyamat kiolvasásával kapható meg. Gondot okoz azonban, hogy a HMM módszerek nem alkalmazhatóak, hiszen a ML likelihood függvény nehezen számolható.

Legyen $|a| < 1$ tetszőleges. Jelölje továbbra is $(\varepsilon_n(a))$ az (y_n) alapján számított folyamatot: $\varepsilon_n(a) = y_n + ay_{n-1}$. Alkalmazzuk a korábbi fejezet eredményeit a ML függvény kiszámítására. Legyen adott egy $y^N = (y_1, \ldots, y_N)$ megfigyelés. Ekkor

$$p(y^N, a) = \int_{C(y^N)} p(\xi^N, a) d\xi^N =$$

$$= \int_{C(y^N)} \frac{1}{(2\pi)^{N/2}\sigma^N} \exp\left\{ -\sum_{n=1}^N \varepsilon_n^2(a)/2\sigma^2 \right\} d\xi^N,$$

ahol

$$C(y^N) = \{\xi^N \ : \ q(\xi^N) = y^N\}.$$

A logaritmus gradiense így írható

$$\frac{\partial}{\partial a} \log p(y^N, a) = \frac{1}{p(y^N, a)} \frac{\partial}{\partial a} p(y^N, a).$$

a szerint deriválva azt kapjuk, hogy

$$\frac{\partial}{\partial a} p(y^N, a) = \int_{C(y^N)} \frac{1}{(2\pi)^{N/2}\sigma^N} \exp\left\{ -\sum_{n=1}^N \varepsilon_n^2(a)/2\sigma^2 \right\} \left(-\sum_{n=1}^N \varepsilon_n(a) y_{n-1}/\sigma^2 \right) d\xi^N$$

Ezért

$$\frac{\partial}{\partial a} \log p(y^N, a) = \int_{C(y^N)} \frac{p(\xi^N, a)}{p(y^N, a)} \left(-\sum_{n=1}^N (y_n + ay_{n-1}) y_{n-1}/\sigma^2 \right) d\xi^N.$$

12. Szűrőstabilitás

Legyenek $\mathcal{X}$ és $\mathcal{Y}$ véges halmazok.

$$p_n^i = P(X_n = i | Y_{n-1}, ..., Y_0)$$

jelöléssel legyen

$$p_n = (p_n^1, ..., p_n^N),$$

ahol $|\mathcal{X}| = N$. Továbbá legyen

$$Q^{ij} = P(X_{n+1} = j | X_n = i),$$

$$b^x(y) = P(Y_n = y | X_n = x),$$

valamint

$$B(y) = \mathrm{diag} b^x(y).$$

Ezen jelölések mellett a Baum-egyenlet

$$p_{n+1} = \pi(Q^T B(Y_n) p_n),$$

ahol $p_0^i = P(X_0 = i)$.

Legyen $M_n = Q^T B(Y_n)$, és így a rekurziót kifejtve

$$p_{n+1} = \pi(M_n p_n) = \pi(M_n M_{n-1} p_{n-1}) = \pi(M_n ... M_0 p_0).$$

Kérdés, hogy különböző p_0,$p_0^,$ kiinduló eloszlásokból indulva mi lesz $\pi(M_n...M_0 p_0)$ és $\pi(M_n...M_0 p_0^,)$ eltérése.

12.1. Definíció. *$M \geq 0$ ($M^{ij} \geq 0 \quad \forall i, j$-re) oszlopmegengedett, ha $\sum_{i=1}^{N} M^{ij} > 0 \ \forall j$-re ($i, j = 1, ..., N$)*

12.2. Definíció. *M oszlop-sztochasztikus, ha $\sum_{i=1}^{N} M^{ij} = 1 \quad \forall j$-re.*

Legyen $\mathcal{P}$ az $\mathbf{R}^N$-ben lévő valószínűségi vektorok halmaza, azaz

$$p\epsilon\mathcal{P} \text{ ha } \quad p = (p_1, ..., p_N)^T, \text{ahol} \quad p_i \geq 0 \text{ és } \quad \sum_{i=1}^{N} p_i = 1.$$

Tetszőleges $x\epsilon\mathbb{R}$ valós számra

$$x^+ = \begin{cases} x & \text{ha } x \geq 0 \\ 0 & \text{ha } x < 0 \end{cases}, \qquad x^- = \begin{cases} 0 & \text{ha } x \geq 0 \\ -x & \text{ha } x < 0 \end{cases}$$

12.3. Definíció. *Legyen* $s = (s_1, ..., s_N)^T$ *esetén az* $\|s\| = \sum_{i=1}^{N} s_i^+ - \sum_{i=1}^{N} s_i^- = \sum_{i=1}^{N} |s_i|$ *a totális variáció.*

Nyilvánvaló, hogy $\|s\| \leq 2$, illetve, hogy

$$\|s + t\| \leq \|s\| + \|t\|,$$

$$\|\lambda s\| = |\lambda|\|s\|.$$

A továbbiakban legyen M oszlop-sztochasztikus mátrix.

12.4. Definíció. $\tau_1(M) = \frac{1}{2} \max_{i,i^,} \|M^{\cdot i} - M^{\cdot i^,}\|_{TV}$

Tehát a definíció közvetlen következménye, hogy $\tau_1(M) \leq 1$.

12.1. Lemma. *Tetszőleges* $p, q\epsilon\mathcal{P}$*-re fennáll, hogy*

$$\|Mp - Mq\| \leq \tau_1(M)\|p - q\|$$

Bizonyítás. Legyen $s = p - q \quad s = s^+ + s^-$, valamint $\delta = \sum_{i=1}^{N} s_i^+ = -\sum_{i=1}^{N} s_i^-$. Ekkor nyilván $\|s\| = 2\delta$, illetve a bizonyítandó állítás $\|Ms^+ + Ms^-\| \leq 2\tau_1(M)\delta$.

Tehát $\delta > 0$ miatt elegendő belátni, hogy

$$\|M\frac{s^+}{\delta} + M\frac{s^-}{\delta}\| \leq 2\tau_1(M)$$

.

Felhasználva, hogy $\frac{s^+}{\delta} = (\lambda_1, \ldots, \lambda_N)$ és $\frac{s^-}{\delta} = (\mu_1, \ldots, \mu_N)$ valószínűségi vektorok (azaz $\sum_{i=1}^{N} \lambda_i = \sum_{i=1}^{N} \mu_i = 1$), kapjuk, hogy

$$\|M\frac{s^+}{\delta} + M\frac{s^-}{\delta}\| = \|\sum_{i=1}^{N} \lambda_i M^{\cdot i} + M\frac{s^-}{\delta}\| \leq \sum_{i=1}^{N} \lambda_i \|M^{\cdot i} + M\frac{s^-}{\delta}\|,$$

és a kapott összeg i. tagjában

$$\|M^{\cdot i} + M\frac{s^-}{\delta}\| = \|M^{\cdot i} - \sum_{j=1}^{N} \mu_j M^{\cdot j}\| \leq \sum_{j=1}^{N} \mu_j \|M^{\cdot i} - M^{\cdot j}\| \leq 2\tau_1(M).$$

Ezzel az állítást bebizonyítottuk.

Megjegyzés: a lemma éles, hiszen legyenek i, i' olyanok, hogy $\frac{1}{2}\|M^{\cdot i} - M^{\cdot i'}\| = \tau_1(M)$. Ekkor a $p = e_i$ és $q = e_{i'}$ választással $\|Me_i - Me_{i'}\| = 2\tau_1(M)$ illetve $\|p - q\| = 2$. Tehát azt kapjuk, hogy

12.1. Következmény.

$$\tau_1(M) = \sup_{p \neq q, p, q \epsilon \mathcal{P}} \frac{\|M_p - M_q\|}{\|p - q\|}$$

A lemma további következményei

12.2. Következmény.

$$\tau_1(M_2 M_1) \leq \tau_1(M_2)\tau_1(M_1)$$

Bizonyítás. $\|M_2 M_1 p - M_2 M_1 q\| \leq \tau_1(M_2)\|M_1 p - M_1 q\| \leq \tau_1(M_2)\tau_1(M_1)\|p - q\|$

Vezessük be a következő jelöléseket. Az i-dik oszlop elemeinek összegét jelölje:

$$M^{\sigma i} = \sum_{j=1}^{N} M^{ji},$$

továbbá

$$\pi(M)^{\cdot i} = \frac{M^{\cdot i}}{M^{\sigma i}} = \pi(M^{\cdot i}) = \overline{M},$$

illetve

$$M^{\sigma} = \mathrm{diag}(M^{\sigma i}).$$

Egyszerű állítások, hogy

12.1. Állítás.

$$M = \overline{M} M^{\sigma} \tag{15}$$

12.2. Állítás.

$$\|\pi(Mp) - \pi(Mq)\| \leq \tau_1(\overline{M}) \|\pi(M^{\sigma}p) - \pi(M^{\sigma}q)\| \quad \forall p, q \geq 0 \tag{16}$$

Valószínűségi vektorok között egy új távolságfogalmat vezetünk be.

12.5. Definíció. *Hilbert projektív távolság: $p, q \geq 0$ esetén legyenek α és β olyanok, hogy $\alpha q_i \leq p_i \leq \beta q_i$ fennáll minden i-re. Legyen α a lehető legnagyobb, β pedig a lehető legkisebb ilyen tulajdonságú szám. Ekkor legyen*

$$\Theta(p, q) = \log \frac{\beta}{\alpha}$$

a Hilbert projektív távolság.

Megjegyzés. $\alpha = 0$ és $\beta = +\infty$ is lehetséges. Ekkor $\Theta(p, q) = \infty$.

Megjegyzés. $\Theta < \infty$ pontosan akkor teljesül, ha a p és q valószínűségi mértékek abszzólút folytonosak.

A Hilbert projektív távolság szimmetrikus, és teljesül a háromszög-egyenlőtlenség, azaz

12.3. Állítás. $\Theta(p,q) = \Theta(q,p)$

Bizonyítás.
$$\alpha q_i \leq p_i \leq \beta q_i \Leftrightarrow \beta^{-1} p_i \leq q_i \leq \alpha^{-1} p_i,$$
amiből következik, hogy
$$\Theta(q,p) = \log \frac{\alpha^{-1}}{\beta^{-1}} = \log \frac{\beta}{\alpha}.$$

12.4. Állítás. $\Theta(p,r) \leq \Theta(p,q) + \Theta(q,r)$

Bizonyítás.
$$\left.\begin{aligned} \alpha q_i \leq p_i \leq \beta q_i \\ \gamma r_i \leq q_i \leq \delta r_i \end{aligned}\right\} \Longrightarrow \alpha\gamma r_i \leq p_i \leq \beta\delta r_i$$
azaz $\Theta(p,r) \leq \log \frac{\beta\delta}{\alpha\gamma} = \log \frac{\beta}{\alpha} + \log \frac{\delta}{\gamma}$.

A Hilbert projektív távolság definíciójából adódik, hogy

12.5. Állítás.
$$\Theta(p,q) = \log \sup_{i,j} \left(\frac{\frac{p_i}{q_i}}{\frac{p_j}{q_j}} \right).$$

12.2. Lemma. *Legyenek $p, q \epsilon \mathcal{P}$ olyan valószínűségi vektorok, melyekre $p_i \leq \beta q_i$ teljesül $\forall i$-re. Ekkor fennáll, hogy*
$$\|p - q\|_{TV} \leq 2(1 - \beta^{-1}).$$

Bizonyítás. A feltétel alapján
$$p_i - q_i \leq p_i - \beta^{-1} p_i = (1 - \beta^{-1}) p_i.$$
Ezt összegezve azon i-kre ahol $p_i > q_i$ kapjuk, hogy
$$\sum_{p_i > q_i} p_i - q_i \leq (1 - \beta^{-1}) \sum_{p_i > q_i} p_i \leq 1 - \beta^{-1}$$

Hasonló összegzést elvégezve $q_i \geq p_i$-re

$$\sum_{q_i \geq p_i} q_i - p_i = \sum_{p_i \geq q_i} p_i - q_i \leq (1-\beta^{-1}) \sum_{p_i \geq q_i} p_i \leq 1 - \beta^{-1}.$$

A két egyenlőtlenséget összeadva kapjuk a bizonyítandó állítást.

12.3. Lemma. *Legyen $N \geq 0$ diagonális mátrix és $p, q \geq 0$ nem azonosan 0 vektorok. Továbbá tegyük fel, hogy $Np, Nq \neq 0$. Ekkor fennáll, hogy*

$$\|\pi(Np) - \pi(Nq)\| \leq 2(1 - \frac{\alpha}{\beta}),$$

ahol

$$\alpha q_i \leq p_i \leq \beta q_i.$$

Megjegyzés. A Lemmában szereplő felső korlát független az N diagonális mátrixtól.

Bizonyítás. Legyen $N = \text{diag} N_i$. A feltétel alapján $\alpha N_i q_i \leq N_i p_i \leq \beta N_i q_i$, azaz i-re összegezve kapjuk, hogy

$$\alpha \sum_i N_i q_i \leq \sum_i N_i p_i \leq \beta \sum_i N_i q_i.$$

Bevezetve az $(\frac{N_i p_i}{\sum_j N_j p_j}) = \overline{p}_i$ és $(\frac{N_i q_i}{\sum_j N_j q_j}) = \overline{q}_i$ jelöléseket, a $\overline{p}, \overline{q} \epsilon \mathcal{P}$ valószínűségi vektorokra fennáll, hogy

$$\overline{p}_i \leq \beta \frac{N_i q_i}{\sum_j N_j p_j} \leq \beta \frac{N_i q_i}{\alpha \sum_j N_j q_j} = \frac{\beta}{\alpha} \overline{q}_i$$

A 12.2 lemmát felhasználva kapjuk a bizonyítandó állítást.

Az előző lemma megfogalmazható a Hilbert féle projektív távolság segítségével is, nevezetesen,

$$\|\pi(Np) - \pi(Nq)\| \leq 2(1 - \exp(-\Theta(p, q))). \tag{17}$$

τ_1 mintájára a Hilbert projektív távolság segítségével definiáljuk a M projektív átmérőjét.

12.6. Definíció. *Legyen* $M > 0$. *Ekkor* $\Delta(M) = \sup_{i,j} \Theta(M^{\cdot i}, M^{\cdot j})$ *(*$M_{ij} > 0$*)* M *projektív átmérője.*

A továbbiakban néhány egyszerű állítást vegyünk végig a projektív átmérővel kapcsolatban.

12.6. Állítás.

$$\Delta(M) := \sup_{i,j,k,l} \log(\frac{M^{ki}/M^{kj}}{M^{li}/M^{lj}})$$

12.3. Következmény. Δ *invariáns az átlóra való tükrözésre, azaz* $\Delta(M) = \Delta(M^T)$

12.7. Állítás. *Ha* B *egy pozitív diagonális mátrix, akkor* $\Delta(MB) = \Delta(M)$

Bizonyítás. Felhasználva, hogy $\Theta(\lambda p, \mu q) = \Theta(p, q) \quad \lambda, \mu > 0$ skalár esetén illetve, hogy $(MB)^{\cdot i} = M^{\cdot i} b_i$, ahol $B = diag(b_i)$, az állítás nyilvánvaló.

12.8. Állítás. *Legyen* M *oszlop-sztochasztikus mátrix. Ekkor*

$$\tau_1(M) \leq 1 - \exp(-\Delta(M)).$$

Bizonyítás. A (17) egyenlőtlenséget használva $N = I$ esetén tetszőleges i, j párra kapjuk, hogy

$$\|M^{\cdot i} - M^{\cdot j}\| \leq 2(1 - \exp(-\Theta(M^{\cdot i}, M^{\cdot j})))$$

Mindkét oldal felének a szuprémumát véve kapjuk a bizonyítandó állítást.

12.9. Állítás. *Legyenek* $M_k > 0$ *pozitív mátrixok (*$k = 0, \ldots, n-1$*) és legyen szorzatuk* $M = M_{n-1} \ldots M_0$. *Ekkor*

$$\tau_1(\pi(M_{n-1} \ldots M_0)) \leq \prod_{k=0}^{n-1} (1 - \exp(-\Delta(M_k))).$$

Bizonyítás. Teljes indukcióval bizonyítunk. $n = 1$ esetén az állítás nyilvánvaló, hiszen

$$\tau_1(\pi(M_0)) \leq 1 - \exp(-\Delta(\overline{M}_0)),$$

de $\Delta(\overline{M}_0) = \Delta(M_0)$.

Most legyen $M = RM_0$, ahol $R = M_{n-1} \dots M_1$. Nyilván $M^{\cdot i} = RM_0^{\cdot i}$, illetve $M^{\cdot j} = RM_0^{\cdot j}$. Tehát

$$\|\pi(RM_0^{\cdot i}) - \pi(RM_0^{\cdot j})\| \leq \tau_1(\pi(R))\|\pi(R^\sigma M_0^{\cdot i}) - \pi(R^\sigma M_0^{\cdot j})\| \leq$$
$$\leq \tau_1(\pi(R))2(1 - \exp(-\Theta(M_0^{\cdot i}, M_0^{\cdot j})))$$

Mindkét oldal felének véve a suprémumát i-ben és j-ben, majd felhasználva az indukciós állítást kapjuk amit bizonyítani akartunk.

Ezen utóbbi állítás közvetlenül alkalmazható a Rejtett Markov Modellünk (HMM) esetére. Legyen most $M_k = Q^T B(Y_k)$, ahol $B(y) = \text{diag }(b^i(y))$.

12.10. Állítás. *Tegyük fel, hogy* $Q, B(y) > 0 \quad \forall y$*-ra. Ekkor*

$$\tau_1(\pi(M_{n-1} \dots M_0)) \leq (1 - \exp(-\Delta(Q)))^n \tag{18}$$

teljesül.

Bizonyítás. Az állítás nyilvánvaló az előzőek alapján, hiszen

$$\Delta(M_k) = \Delta(Q^T B(Y_k)) = \Delta(Q^T) = \Delta(Q).$$

Megjegyezzük, hogy az állításban $\Delta(Q) < 1$, hiszen $Q > 0$ így geometriai konvergencia van M_k-ban.

A korábbi fejezetek jelöléseit használva továbbra is legyen

$$P(X_n = i | Y_{n-1} \dots Y_0) = p_n^i,$$

$$p_n = (p_n^i),$$

és a Baum-egyenlet

$$p_n = \pi(Q^T B(Y_{n-1}) p_{n-1})$$

a $p_0^i = P(X_0 = i)$ kezdeti eloszlással. Az $M_k = Q^T B(Y_k)$ jelöléssel ez utóbbi írható a

$$p_n = \pi(M_{n-1} p_{n-1}) = \cdots = \pi(M_{n-1} \ldots M_0 p_0)$$

alakban. A továbbiakban azt szeretnénk tudni, hogy különböző kezdeti eloszlásokból elindulva mennyire fog eltérni a folyamatunk, másszóval a

$$\|\pi(M_{n-1} \ldots M_0 p) - \pi(M_{n-1} \ldots M_0 q)\|$$

különbséget akarjuk becsülni, ahol $p, q \epsilon \mathcal{P}$ valószínűségi vektorok. (15) és (16) alapján azt tudjuk, hogy

$$\|\pi(M_{n-1} \ldots M_0 p) - \pi(M_{n-1} \ldots M_0 q)\| \leq \tau_1(\pi(M_{n-1} \ldots M_0)) \|\pi(M^\sigma p) - \pi(M^\sigma q)\|,$$

ahol $M^\sigma = \text{diag } (M^{\sigma i})$ diagonális mátrix és ez utóbbi (18) alapján felülről becsülhető

$$(1 - \delta)^n \|\pi(M^\sigma p) - \pi(M^\sigma q\|$$

-vel, ahol $\delta = \exp(-\Delta(Q))$. Tehát kapjuk a következő állítást:

12.11. Állítás. *Ha $Q, B(y) > 0 \quad \forall y$-ra, akkor*

$$\|p_n^p - p_n^q\| \leq 2(1 - \delta)^n,$$

ahol p_n^p, illetve p_n^q jelöli p, illetve q indítás melletti állapotot n lépés után.

12.7. Definíció. *Legyen $D = diag(d_i) > 0$ egy tetszőleges diagonális mátrix, és legyen $\omega(D) = \max\limits_{i,j} \frac{d_i}{d_j}$.*

12.4. Lemma. *$p, q \epsilon \mathcal{P}$ valószínűségi vektorokra fennáll, hogy*

$$\|\pi(Dp) - \pi(Dq)\|_{TV} \leq \omega(D) \|p - q\|_{TV}$$

Bizonyítás. Feltehetjük, hogy $1 \leq d_i \leq \omega(D) = \omega$ teljesül $\forall i$-re, hiszen D átskálázható. Ekkor $q_i \leq d_i q_i \leq \omega q_i$, azaz i-re összegezve $1 \leq \sum_i q_i d_i = \rho_q \leq \omega$. Azt is feltehetjük, hogy $\rho_p \leq \rho_q$. Ekkor tetszőleges i-re

$$\frac{d_i p_i}{\rho_p} - \frac{d_i q_i}{\rho_q} \geq \frac{d_i}{\rho_q}(p_i - q_i)$$

Innen

$$[\frac{d_i p_i}{\rho_p} - \frac{d_i q_i}{\rho_q}]^- \leq \frac{d_i}{\rho_q}[p_i - q_i]^- \leq \omega[p_i - q_i]^-$$

amit összegezve i szerint kapjuk, hogy

$$1^T(\pi(Dp) - \pi(Dq))^- \leq \omega 1^T(p-q)^-$$

Azonban $1^T(p-q)^- = 1^T(p-q)^+$ egyenlőség alapján ugyanez elmondható $+$-ra is, így kapjuk, hogy

$$\|\pi(Dp) - \pi(Dq)\|_{TV} \leq \omega\|p-q\|_{TV}$$

Általában ha $D > 0$ nem diagonális, akkor legyen

$$\omega(D) = \omega(D^\sigma)) = \max_{i,j} \frac{D^{\sigma i}}{D^{\sigma j}} = \inf\{\gamma : D^{\sigma i} \leq \gamma D^{\sigma j} \quad i,j\}.$$

Ennek mintájára vezessük be pozitív mátrixokra a következő operátort:

12.8. Definíció. *$D > 0$ esetén legyen $\beta(D) = \inf\{\beta : D^{\cdot i} \leq \beta D^{\cdot j} \quad \forall i,j\}$.*

Nézzünk néhány egyszerű állítást β-ra.

12.12. Állítás. $\omega(D) \leq \beta(D)$

Bizonyítás. Ha $D^{\cdot i} \leq \beta D^{\cdot j}$ teljesül, akkor $\mathbf{1}^T D^{\cdot i} \leq \beta \mathbf{1}^T D^{\cdot j}$ is teljesül, ahonnan az állítás adódik.

12.13. Állítás. *Legyenek $M_k > 0$ pozitív mátrixok. Ekkor fennáll, hogy*

$$\beta(M_{n-1} \ldots M_0) \leq \beta(M_0).$$

Bizonyítás. Az állítás bizonyítását egy lemmával kezdjük.

12.5. Lemma. *Legyenek $p, q > 0$ olyan vektorok, hogy $p \leq \beta q$, valamint legyen $M > 0$. Ekkor $Mp \leq \beta Mq$ teljesül.*

Bizonyítás. (lemma): $p_i \leq \beta q_i$-ből következik, hogy $p_i M^{\cdot i} \leq \beta q_i M^{\cdot i}$. Az egyenlőtlenség mindkét oldalát összegezve i-re kapjuk a lemma állítását. Most térjünk vissza az állítás bizonyítására. Teljes indukcióval bizonyitunk. $n = 1$ esetén az állítás triviális. Tegyük fel, hogy n-ig igaz, és bizonyitsunk $n+1$-re. Legyen $M' = M_n \dots M_1$. Ekkor $M' > 0$ nyilván teljesül. Alkalmazzuk a lemmát $p = M_0^{\cdot i}$ és $q = M_0^{\cdot j}$ és $M = M'$-re. Tegyük fel, hogy $M_0^{\cdot i} \leq \beta M_0^{\cdot j}$ teljesül $\forall i, j$-re. Ekkor a lemma alapján $M' M_0^{\cdot i} \leq \beta M' M_0^{\cdot j}$ is teljesül, azaz az állitást bebizonyítottuk.

Alkalmazzuk a lemmát az $M_k = Q^T B(y_k)$ mátrixokra, és így kapjuk, hogy $\omega(M_{n-1} \dots M_0) \leq \beta(M_{n-1} \dots M_0) \leq \beta(M_0)$ Tehát a 12.11 Állítás helyett kapjuk a következő becslést:

12.1. Tétel. *Ha $Q, B(y) > 0 \quad \forall y$-ra, illetve $p, q \epsilon \mathcal{P}$ valószínűségi vektorok, akkor*

$$\|\pi(M_{n-1} \dots M_0 p) - \pi(M_{n-1} \dots M_0 q)\|_{TV} \leq (1 - \exp(-\Delta(Q)))^n \beta(M_0) \|p - q\|_{TV}$$

teljesül.

Eddig feltettük, hogy mind az átmenetvalószínűség mátrix (Q), mind a feltételes valószínűség mátrixok ($B(y)$) pozitívak voltak. Most nézzük meg, mit mondhatunk abban az esetben, ha $B(y)$-ról csak annyit tudunk, hogy nemnegatív.

12.6. Lemma. *Legyen $Q > 0$, illetve $B(y) \geq 0 \quad \forall y$-ra. Ekkor*

$$\tau_1(\pi(M_{n-1} \dots M_1 Q^T)) \leq (1 - \exp(-\Delta(Q)))^n.$$

Bizonyítás. $M_{n-1} \dots M_1 Q^T > 0$, hiszen $M_1 Q^T = Q^T B(y_1) Q^T > 0$ (ez utóbbi könnyen látszik, hiszen $B(y_1)(Q^T)^{\cdot i} \neq 0$). Tehát $\pi(M_{n-1} \dots M_1 Q^T)$ egy pozitív mátrix, és ráadásul ez folytonosan függ az $M_{n-1} \dots M_1 Q^T$ szorzattól. Másfelől $\tau_1(M)$ folytonos függvénye M-nek. Így a lemma bizonyítását azonnal megkapjuk, ha perturbáljuk a $B(y)$-t, hogy $B^{\epsilon}(y) > 0$ legyen.

A lemma alapján megkapjuk a 12.1-nek megfelelő tételt.

12.2. Tétel. *Legyenek $Q > 0$, $B(y) \geq 0$, $p, q \epsilon \mathcal{P}$ valószínűségi vektorok. Ekkor*

$$\|\pi(M_{n-1} \dots M_0 p) - \pi(M_{n-1} \dots M_0 q)\|_{TV} \leq$$

$$\leq (1 - \exp(-\Delta(Q)))^n \beta(Q^T) \|\pi(B(Y_0)p) - \pi(B(Y_0)q)\|_{TV}$$

13. A gradiens folyamat

A továbbiakban használni fogjuk a $p_n(q) = p_n(X_n|y_{n-1}, \dots, y_0)$ jelölést utalva arra, hogy az n. lépés utáni eloszlás a kezdeti eloszlás függvénye ($p_0 = q \epsilon \mathcal{P}$).

13.1. Tétel. *Legyen $(y_{n-1}, \dots y_0)$ egy fix sorozat, $p_n(q)$ egy folytonosan differenciálható függvénye q-nak, $q(t)$ pedig egy folytonosan differenciálható görbe $\mathcal{P}$-n, melyre*

$$\frac{\partial q(0)}{\partial t} = w.$$

Ekkor

$$\|\frac{\partial}{\partial t} p_n(q(t))\|_{TV}|_{t=0} = \|\frac{\partial}{\partial q} p_n(q) w\|_{TV} \leq \beta(M_0)(1 - \exp(-\Delta(Q)))^n \|w\|_{TV}$$

Bizonyítás. $M_{n-1} \dots M_0 q$ lineáris leképezés, igy folytonosan differenciálható. Az $x \longrightarrow \pi(x)$ projekció szintén folytonosan differenciálható, igy a $q \longrightarrow \pi(M_{n-1} \dots M_0 q)$ folytonosan differenciálható, és legyen a deriváltja $F_n(q)$. Tehát jelölésünkkel $p_n(t) = p_n(q(t))$, illetve $\dot{p}_n = F_n \dot{q}_n$, ahol $q(t) \epsilon \mathcal{P}$ valószinűségi vektor, igy $1^T \dot{q}_n = 0$. Elemi analizisből ismert, hogy

$$\lim_{h \to 0} \frac{p_n(q) - p_n(q')}{h} = F_n w,$$

ahol $q - q' = hw$.

Tehát kapjuk, hogy

$$\|p_n(q) - p_n(q')\|_{TV} \leq \beta(M_0)(1 - \exp(-\Delta(Q)))^n h \|w\|_{TV},$$

azaz

$$\|F_n w\|_{TV} \leq \beta(M_0)(1 - \exp(-\Delta(Q)))^n h \|w\|_{TV}$$

amivel a tételt bebizonyitottuk.

Mivel a Baum-egyenlet tulajdonképpen felírható

$$p_{n+1} = \pi(Q^T B(y_n)p_n) = g(p_n, y_n)$$

alakban, így a rekurziót kibontva kapjuk, hogy

$$p_n = g(y_{n-1}, g(y_{n-2}, \dots g(y_1, g(y_0, p_0)) \dots)).$$

Tehát a derivált előáll a

$$\frac{\partial p_n}{\partial q} = \frac{\partial g(y_{n-1}, p_{n-1})}{\partial p_{n-1}} \dots \frac{\partial g(y_0, p_0)}{\partial p_0}$$

alakban ($p_0 = q$). Legyen

$$G_{n-1} = \frac{\partial g(y_{n-1}, p_{n-1})}{\partial p_{n-1}}.$$

Ezen jelölések mellett az előző tétel

13.2. Tétel.

$$\|G_{n-1} \dots G_0 w\|_{TV} \leq \beta(M_0)(1-\delta)^n \|w\|_{TV},$$

ahol $\delta = \exp(-\Delta(Q))$ *és* $\mathbf{1}^T w = 0$.

Továbbá általában $\|G_{n-1} \dots G_k w\|_{TV} \leq \beta(M_k)(1-\delta)^{n-k} \|w\|_{TV}$.

14. Markov-folyamatok reprezentációja

Legyen M az állapottér lengyel-tér (teljes szeparábilis metrikus tér), $\mathcal{X} : \mathcal{M} \longrightarrow \mathcal{M}$ Borel-mérhető leképezések egy tere. Tegyük fel, hogy $\mathcal{X}$ mérhető tér, és van rajta egy m valószínűségi mérték. Végül legyenek a T_n leképezések független azonos eloszlásúak m szerint.

Definiáljuk a következő folyamatot: $X_0 \epsilon M$, $X_{n+1} = T_{n+1} X_n$.

14.1. Állítás. *Az így megadott* X_n *folyamat Markov-folyamat.*

Most tekintsük fordítva a helyzetet. Tegyük fel, hogy van egy Markov-folyamatunk a $P(x, G)$ átmenet-valószínűségekkel ($x \epsilon M, \quad G \epsilon \mathcal{B}(M)$), ahol M lengyel tér, $P(x, .)$ valószínűségi mérték $\mathcal{B}(M)$-n, rögzített G-re $P(., G)$ Borel-mérhető). Keressük ennek egy reprezentációját oly módon, hogy $\mathcal{X}$-n legyen egy m mértékünk, amivel $P(x, G) = m\{T : Tx \epsilon G\}$.

14.2. Állítás. *Ilyen reprezentáció létezik.*

Legyen $\varphi : M \longrightarrow [0, 1]$ egy Borel-izomorfia a M lengyel térről és a $[0, 1]$ zárt intervallumra, és így $\Gamma = \varphi(M) \subseteq [0, 1]$ a φ képhalmaza, azaz φ^{-1} a Γ-n van értelmezve. Adott $x \epsilon M$-re a $P(x, .)$ mértéket átvisszük a $[0, 1]$ intervallumra a φ segítségével, és így kapjuk a $\widetilde{P}(x, .)$ mértéket, azaz $\Delta \subseteq [0, 1]$ Borel-halmazra $\widetilde{P}(x, \Delta) = P(x, \varphi^{-1}(\Delta \cap \Gamma)$

Kvantilis transzformáció: tegyük fel, hogy $P(x, [0, \gamma)) = F(\gamma)$ eloszlásfüggvény, azaz monoton növő, $F(0) = 0$, $F(1) = 1$. Definiáljuk F általánosított inverzét a következő módon: $F^{-1}(g) = \inf\{x : F(x) > g\}$. Ekkor ha U egyenletes eloszlású a $[0, 1]$ intervallumban, akkor $F^{-1}(U)$ eloszlása éppen F lesz.

Esetünkben legyen $Z(x, \omega) = \inf\{\gamma : \widetilde{P}(x, [0, \gamma] \geq \omega\}$.

14.3. Állítás. *Ha ω egyenletes eloszlású a $[0, 1]$ intervallumban, akkor $Z(x, \omega)$ eloszlása éppen $\widetilde{P}(x, .)$.*

Bizonyítás.
$$1 - \lambda\{\omega : Z(x, \omega) \leq \alpha\} = \lambda\{\omega : Z(x, \omega) > \alpha\} =$$
$$= \lambda\{\omega : \widetilde{P}(x, [0, \alpha]) < \omega\} = 1 - \widetilde{P}(x, [0, \alpha]),$$
ahol λ jelöli a $[0, 1]$ intervallumon a Lebesgue-mértéket. Innen az állítás nyilvánvaló.

14.4. Állítás. *$Z(., \omega)$ Borel-mérhető x-ben.*

Bizonyítás.

$$\{x : Z(x, \omega) > \alpha\} = \{x : \widetilde{P}(x, [0, \alpha]) < \omega\} = \{x : P(x, \varphi^{-1}([0, \alpha] \cap \Gamma)) < \omega\}$$

alapján az állítás nyilvánvaló.

Terjesszük ki a φ^{-1} függvényt a teljes $[0,1]$ intervallumra. Legyen $\psi = \varphi^{-1}$ a Γ halmazon, és $\psi(z) = x_0 \epsilon M$ rögzített ha $z \epsilon \Gamma^c$. A ψ függvény segítségével definiáljuk a véletlen leképezésünket, mely eleget fog tenni a reprezentáció feltételeinek.

Legyen $f(\omega) : M \longrightarrow M$ leképezés a következő képlettel definiálva:

$$f(\omega)x = \psi \circ Z(x, \omega).$$

Ekkor f nyilván Borel-mérhető minden ω-ra.

Ezzel megadtunk egy konstrukciót a reprezentációra. Természetesen a Markov-folyamatok fenti reprezentációja nem egyértelmű.

A továbbiakban egy újabb fontos fogalmat vezetünk be a Markov-láncok körében.

14.1. Definíció. *Egy Markov-láncra teljesül a Doeblin-feltétel, ha $\forall x \epsilon M$-re és $A \subset \mathcal{B}(M)$-re fennáll, hogy $P(x, A) \geq \delta\nu(A)$, ahol $\delta > 0$ és ν valószínűségi mérték.*

Most tekintsünk egy fontos speciális esetet, a független azonos eloszlású valószínűségiváltozó sorozatot (i.i.d. eset). Ebben az esetben az $X_{n+1} = T_{n+1}X_n$ reprezentációnál $T_n x$ független x-től, azaz T_n konstans függvény ($T_n \epsilon \Gamma_c$). Ugyanakkor ebben az esetben teljesül a Doeblin-feltétel is $\delta = 1$-gyel, hiszen $P(x, A) = \nu(A)$ - az eloszlás alapján. A Doeblin-feltétel δ-ja tulajdonképpen azt mutatja meg, hogy egy Markov-lánc "mennyire i.i.d.". A következő lemma a Doeblin-feltétel és a Markov-lánc reprezentációjának kapcsolatát mutatja.

14.1. Lemma. *Az X_n Markov láncra teljesül a Doeblin-feltétel ekvivalens azzal, hogy létezik olyan T_n független azonos eloszlású reprezentáció, melyre $P(T_n \epsilon \Gamma_c) \geq \delta$*

Bizonyítás. Először tegyük fel, hogy létezik a fenti tulajdonságú T_n reprezentáció. Ekkor $P(x, A) = P(T_1 x \epsilon A) \geq P(T_1 x \epsilon A | T_1 \epsilon \Gamma_c) P(T_1 \epsilon \Gamma_c) \geq m(A)\delta$, ahol m valószínűségi mérték.

Megfordítva tegyük fel, hogy teljesül a Doeblin-feltétel. Ekkor δ valószinűséggel vegyünk ν szerint egy x-t és $1 - \delta$ valószinűséggel vegyünk egy olyan T leképezést, amelyik a

$$\frac{P(x,A)-\delta\nu(A)}{(1-\delta)^{-1}}=\overline{P}(x,A)$$

magfüggvényű Markov-lánc reprezentációjából adódik. Ez nyilván a feltételeknek megfelelő reprezentáció lesz.

14.1. Tétel. *Tegyük fel, hogy az X_n Markov-láncra teljesül a Doeblin-feltétel. Ekkor létezik egy π invariáns eloszlás, és erre fennáll, hogy*

$$|P^n(x,A)-\pi(A)|\leq(1-\delta)^n \quad \forall A\epsilon\mathcal{B}(M)\text{-re.}$$

Bizonyítás. Legyen T_n a folyamat reprezentációja. Ekkor a 14.1 lemma alapján 1-valószínűséggel létezik a $\lim_n T_1\dots T_nx$ határérték, hiszen $P(T_k\epsilon\Gamma_c)\geq\delta>0$, és így 1-valószínűséggel $\exists k$, hogy $T_k\epsilon\Gamma_c$, és konstans leképezés után a folyamat már determinisztikus.

Legyen $\lim_n T_1\dots T_nx=y$. Ekkor y eloszlása invariáns, hiszen $\lim_n T_0T_1\dots T_nx=T_0y$ eloszlása a fentihez hasonló okokból megegyezik y eloszlásával.

Tehát

$$\begin{aligned}|P^n(x,A)-\pi(A)|&=|P(X_n\epsilon A)-\pi(Y_n\epsilon A)|=\\&=|E(\chi_A(X_n)-\chi_A(Y_n))|\leq P(X_n\neq Y_n),\end{aligned}$$

ahol $X_n=T_n\dots T_1X_0$ és $Y_n=T_n\dots T_1Y_0$.

Ugyanakkor $P(X_n\neq Y_n)\leq P(T_k\epsilon\Gamma_c^c,k\leq n)\leq(1-\delta)^n$. Ezzel a tétel állítását bizonyítottuk.

Most térjünk vissza a Rejtett Markov-folyamatok világába. Legyen (X_n,Y_n) rejtett Markov-folyamat, és tegyük fel, hogy mind az $\mathcal{X}$ állapottér, mind az $\mathcal{Y}$ kiolvasási tér lengyel tér.

14.2. Lemma. *Tegyük fel, hogy az X_n Markov-láncra teljesül a Doeblin-feltétel. Ekkor az (X_n,Y_n) párra is teljesülni fog.*

Bizonyítás. Tekintsük a 14.1 lemma szerinti T_n reprezentációját a Markov-láncnak, azaz létezik a T_n független azonos eloszlású leképezés sorozat, hogy $P(T_n \epsilon \Gamma_c) \geq \delta > 0$ és T_n független az X_0 kiindulástól, amivel $X_{n+1} = T_{n+1} X_n$.

Most tekintsük a kiolvasást. Legyen $P(x, G)$ az adott magja a kiolvasásnak, ahol $x \epsilon \mathcal{X}$ és $G \epsilon \mathcal{Y}$. Ekkor az eddigiekhez hasonló módon létezik az $\mathcal{X} \longrightarrow \mathcal{Y}$ leképezések terén egy m mérték úgy, hogy $P(x, G) = m\{U : Ux \epsilon G\}$.

Tehát az $Y_n = U_n X_n$ jelöléssel $X_{n+1} = T_{n+1} X_n$, illetve $Y_{n+1} = U_{n+1} T_{n+1} X_n$, azaz ha $T_n \epsilon \Gamma_c(X \to X)$, akkor $U_n T_n \epsilon \Gamma_c(\mathcal{X} \to \mathcal{Y})$, amivel a lemmát bebizonyítottuk.

A Doeblin-feltételt szokás a következő kicsit általánosabb formában definiálni:

14.2. Definíció. *Az X_n folyamatra teljesül az m-edrendű Doeblin-feltétel, ha $\exists m \geq 1$, hogy $P^m(x, A) \geq \delta \nu(A)$ valamilyen ν valószínűségi mértékkel.*

14.5. Állítás. *(Bhattacharya, Waymire) Legyen X_n egy Markov-lánc. Ekkor az általánosított Doeblin-feltétel ekvivalens azzal, hogy $\exists T_n$ független azonos eloszlású leképezéssorozat, hogy $P(T_m \ldots T_1 \epsilon \Gamma_c) \geq \delta$*

15. L-keverő folyamatok

A továbbiakban a motiváció a következő. Ha X_n egy Markov-lánc egy általános $\mathcal{X}$ állapottéren ($\mathcal{X}$ lengyel tér), és adott egy $G : \mathcal{X} \longrightarrow \mathbf{R}$ mérhető, korlátos függvény, akkor mikor létezik nagy számok törvénye, azaz mikor létezik a $\lim_N \frac{1}{N} \sum_{n=1}^{N} g(x_n)$ határérték?

Bevezetünk egy folyamatosztályt, az L-keverő folyamatok osztályát. Ehhez először szükségünk van az M-korlátosság fogalmára.

15.1. Definíció. *Az u_n valós értékű sztochasztikus folyamat ($n \geq 0$) M-korlátos, ha $\forall q \geq 1$-re $M_q(u) = \sup_{n \geq 0} E^{1/q} |u_n|^q < \infty$*

Legyenek $(\mathcal{F}_n)$ monoton növő. illetve $(\mathcal{F}_n^+)$ monoton fogyó σ-algebra sorozatok olyanok, hogy $\mathcal{F}_n$ és $\mathcal{F}_n^+$ függetlenek $\forall n$-re. Egy egyszerű példa ilyen σ-algebrákra az, amikor (e_n) egy független azonos eloszlású valószínűségváltozó sorozat és $\mathcal{F}_n = \sigma\{e_i : i \leq n\}$, valamint $\mathcal{F}_n^+ = \sigma\{e_i : i \geq n+1\}$.

15.2. Definíció. *Az X_n folyamat L-keverő, ha M-korlátos és*

$$\gamma_q(\tau) = \sup_{n \geq \tau} E^{1/q}|X_n - E(X_n|\mathcal{F}_{n-\tau}^+)|^q$$

mellett

$$\Gamma(q) = \sum_{\tau=0}^{\infty} \gamma_q(\tau) < \infty.$$

Ahhoz, hogy egy folyamatról lássuk, hogy L-keverő, nagyon fontos a következő lemma.

15.1. Lemma. *Legyen X olyan valószínűségi változó, melyre $E|X|^q < \infty \quad \forall q$-ra, $\mathcal{G} \subset \mathcal{F}$ σ-algebra és η $\mathcal{G}$-mérhető valószínűségi változó. Ekkor fennáll, hogy*

$$E^{1/q}|X - E(X|\mathcal{G})|^q \leq 2E^{1/q}|X - \eta|^q.$$

Az L-keverő folyamatok tulajdonságainak fontosságát mutatja, hogy

15.1. Állítás. *Legyen $X_n \epsilon \mathcal{X}$ Markov-lánc ($\mathcal{X}$ lengyel tér), melyre teljesül a Doeblin-feltétel $m = 1$ mellett, valamint $g : \mathcal{X} \longrightarrow \mathbf{R}$ korlátos, mérhető függvény. Ekkor $g(X_n)$ L-keverő folyamat.*

Bizonyítás. Legyen $n > m$ és $n - m = \tau$. Cél, hogy approximáljuk $g(X_n)$-t. Legyen $\mathcal{F}_n = \sigma\{X_0, T_k : k \leq n\}$ és $\mathcal{F}_n^+ = \sigma\{T_k : k \geq n+1\}$. Először approximáljuk X_n-t $X_{n,m}^+$-szal, ahol $X_{n.m}^+ = T_n \dots T_{m+1} X^*$ és X^* egy konstans. Nyilvánvaló, hogy $X_{n,m}^+$ az $\mathcal{F}_m^+$ mérhető. Továbbá

$$P(X_n \neq X_{n,m}^+) \leq P(T_k \text{ nem konstans } : m+1 \leq k \leq n) \leq (1-\delta)^{n-m}$$

Tehát

$$\mathrm{E}^{1/q}|g(X_n) - g(X_{n,m}^+)|^q \leq 2KP^{1/q}(X_1 \neq X_{n,m}^+) \leq 2K(1-\delta)^{\frac{n-m}{q}},$$

ahol K a g függvény korlátja, azaz $|g| \leq K$. A lemma alapján így

$$\gamma_q(\tau) \leq 4K(1-\delta)^{\frac{\tau}{q}}$$

Az egyenlőtlenség jobb oldala azonban szummálható (véges), így az állítást bebizonyítottuk.

Célunk most a Baum-egyenlet felírása kicsit általánosabb környezetben. Legyen továbbra is X_n egy Markov-lánc az $\mathcal{X}$ állapottéren, melyre teljesül a Doeblin-feltétel $m = 1$-re. Legyen $f : \mathcal{X} \times \mathcal{N} \longrightarrow \mathcal{N}$ olyan függvény, ahol $\mathcal{N}$ egy normált tér. Tekintsük a $z_{n+1} = f(x_n, z_n)$ rekurziót ($z_0 = \xi$, x_n tetszőleges). A rekurzió megoldását $z_n(\xi)$-vel fogjuk jelölni.

15.3. Definíció. *Az $f(x, z)$ leképezés egyenletesen exponenciálisan stabil, ha $\forall\{x_n\}$ sorozatra z_n korlátos ($\{x_n\}$-től független) és*

$$\|z_n(\xi) - z_n(\xi')\| \leq C(1-\delta)^n\|\xi - \xi'\|,$$

ahol $C, \delta > 0$ független az $\{x_n\}$ sorozattól.

Speciális esetként megjegyezzük, hogy ha $\mathcal{N}$ a k-dimenziós Euklideszi tér, és az x_n-eknek a $k \times k$-as mátrixok felelnek meg, akkor $Z_{n+1} = A_n Z_n$, ahol $Z_0 = \xi$ és $A_n \epsilon \mathcal{A}$ a rekurzió. Kérdés, hogy hogyan kell megválasztani $\mathcal{A}$-t, hogy egyenletesen exponenciálisan stabil legyen a folyamat. (Ez egy nehéz kérdés, de az utóbbi időben megoldották.)

15.1. Tétel. *Tekintsük a $\{X_n, Z_n\}$ folyamatot, ahol $Z_{n+1} = f(X_n, Z_n)$, $\quad Z_0 = \xi$, és X_0, Z_0 független az X_n Markov-lánc reprezentációjából adódó $\{T_n\}$-től ($n \geq 1$). Legyen $g(x, z)$ egy korlátos, mérhető Lipschitz-folytonos függvény z-ben. Ekkor $v_n = g(x_n, z_n)$ L-keverő folyamat.*

Bizonyítás. Legyen $n > m$, $\tau = n - m$, $\mathcal{F}_n, \mathcal{F}_n^+$, valamint $X_{n,m}^+$ ugyanazok, mint korábban. Legyen Z_n approximációja a következő: $Z_{k+1,m}^+ = f(X_{k,m}^+, Z_{k,m}^+)$, ahol $Z_{m,m}^+ = z^*$ konstans. Nyilvánvaló, hogy ekkor $Z_{n,m}^+$ $\mathcal{F}_m^+$ mérhető.

Legyen $m' = n - [\frac{\tau}{2}]$. Annak valószínűsége kicsi, hogy m'-ig nincsen összeakadás (azaz a reprezentációnál nem lép fel a konstans leképezés), hiszen $P(X^+_{m',m} \neq X_{m'}) \leq (1-\delta)^{[\frac{\tau}{2}]}$. Jelöljük azt az eseményt B-vel, amikor m'-ig nincsen összeakadás, azaz

$$B \subset \{\omega : \quad X^+_{n,m} \neq X_n\}$$

Most tekintsük azt az esetet, amikor összeakadás van, azaz a B^C eseményt. Ekkor $X^+_{k,m} = X_k$ minden $k \geq m'$-re.

Ekkor tekintsük a következő folyamatot:

$$Z^+_{k+1,m} = f(X_k, Z^+_{k,m}) \quad Z^+_{m',m}\text{kezdeti érték mellett},$$

másfelől az eredeti folyamat

$$Z_{k+1} = f(X_k, Z_k) \quad Z_{m'}\text{kezdeti érték mellett}$$

A B^C halmazon $\|Z^+_{n,m} - Z_n\| \leq C(1-\delta)^{[\frac{\tau}{2}]}\|Z_{m',m} - Z_{m'}\|$

Tehát alkalmazva a 15.1 lemmát, kapjuk, hogy

$$E^{\frac{1}{q}}\|g(X^+_n, Z^+_{m,n}) - g(X_n, Z_n)\|^q \leq 2KP^{\frac{1}{q}}(B) + CE(\|Z_{m',m} - Z_{m'}\|)(1-\delta)^{[\frac{\tau}{2}]} \leq$$

$$\leq 2K(1-\delta)^{\frac{[\tau/2]}{q}} + C'(1-\delta)^{[\tau/2]}$$

, azaz az L-keverőség definíciójában szereplő $\gamma_q(\tau)$-ra fennáll, hogy

$$\gamma_q(\tau) \leq 2(2K(1-\delta)^{\frac{[\tau/2]}{q}} + C'(1-\delta)^{[\tau/2]}),$$

ami szummábilis, így a $v_n = g(x_n, z_n)$ folyamat valóban L-keverő.

Végül nézzük meg, hogy ezen általános eredmények mit jelentenek a Rejtett Markov-láncok esetében. Az $\mathcal{X}$ állapottér legyen véges, és legyen $Q > 0$ az átmenetvalószínűség mátrix. Ekkor X_n-re nyilván teljesül a Doeblin-feltétel, amiből az is következik, hogy az (X_n, Y_n) párra is teljesül a Doeblin-feltétel tetszőleges $Y_n \epsilon \mathcal{Y}$ kiolvasási tér (lengyel tér) esetén. A $p^i_n = P(X_n = i | Y_{n-1}, \ldots, Y_0)$ jelölés mellett volt a

$$p_{n+1} = \pi(Q^T B(Y_n) p_n) = f(Y_n, p_n)$$

Baum egyenlet. A ??? lemmában bizonyítottuk, hogy $Q > 0$ esetén

$$\|p_n(q) - p_n(q')\| \leq C(1-\delta)^n,$$

feltéve, hogy q, q' kompatibilisek az Y_0-lal.

15.2. Állítás. *A fenti feltételek mellett ha $Q > 0$, akkor p_n L-keverő folyamat.*

Végül nézzük meg, hogy mit mondhatunk a Maximum likelihood becslésről. A maximum likelihood becslés persze azt jelenti, hogy keressük azt a θ-t, ahol maximális a

$$\log P(y_{n-1}, \ldots y_0, \theta) = \sum_{k=1}^{n-1} \log P(y_k | y_{k-1}, \ldots y_0, \theta) + \log P(y_0, \theta) \tag{19}$$

kifejezés, ahol

$$\log P(y_k | y_{k-1}, \ldots y_0, \theta) = \sum_x \log b^x(y_k) P(x | y_{n-1}, \ldots, y_0, \theta) \tag{20}$$

Természetesen merül fel a kérdés, hogy létezik-e határértéke a

$$\frac{1}{n} \log P(Y_{n-1}, \ldots, Y_0, \theta)$$

likelihood függvénynek. Ha $\theta = \theta^*$, akkor a Shannon-McMillen-Breiman tétel azt mondja, hogy ha Y stacionárius és ergodikus, akkor

$$\frac{1}{n} \log P(Y_{n-1}, \ldots, Y_0, \theta^*)$$

konvergens 1-valószínűséggel. Azonban most nem használjuk ezt az erős állítást, csak annyit, hogy az L-keverő folyamatoknál van nagy számok törvénye.

Ha a p_n folyamat L-keverő, akkor a (19) illetve a (20) jobb oldalán van nagy számok törvénye, azaz a maximum likelihood becslésnek van konzisztenciája (létezik a $\lim_n \frac{1}{n} \log P(Y_{n-1}, \ldots, Y_0, \theta)$ határérték).