REJTETT MARKOV MODELLEK STATISZTIKAI VIZSGÁLATA

TÉZISFÜZET

Molnár-Sáska Gábor

Témavezető:
Gerencsér László

2005.

Budapesti Műszaki és Gazdaságtudományi Egyetem, Matematika Intézet

és

Számítástechnikai és Automatizálási Kutatóintézet, SZTAKI

Ebben a tézisfüzetben a Rejtett Markov Modellek Statisztikai Vizsgálata című PhD dolgozat eredményeit foglaljuk össze. Először egy rövid történeti áttekintést adunk a választott témáról, majd a modell definiálása után bemutatjuk a dolgozat eredményeit.

1. Bevezető

A Rejtett Markov Model (HMM) egy diszkrét idejű, véges állapotterű homogén Markov lánc, melyet egy memória nélküli zajos csatornán figyelünk meg. A Markov lánc minden állapotához hozzá van rendelve egy kiolvasási sűrűség, ami megadja a zaj mértékét. A Markov lánc kezdeti eloszlása illetve átmenetvalószínűségi mátrixa, valamit a kiolvasási sűrűségek valamilyen paraméteres családja karakterizálja a Rejtett Markov Modellt.

A Rejtett Markov Modell sztochasztikus rendszerek modellezésének egyik alapvető eszköze lett. Széleskörű alkalmazása folyik a nanotechnológiában [31], a telekommunikációban [52], a beszédfelismerésben [30], a kapcsolt rendszerekben [16, 20], a pénzügyi matematikában [13] és a fehérje kutatásban [53]. Egy általános bevezető található a [15] összefoglaló cikkben.

A Rejtett Markov Modellek dinamikájának becslése alapvető probléma az alkalmazásokban. Az első eredmény Baumtól és Petrie-től származik, akik a kérdést véges állapottér és véges kiolvasási tér mellett vizsgálták, [5]. Eredményük a Shannon-Breiman-McMillan tételen alapszik. Erős értelemben vett konzisztenciát először Araposthatis és Marcus bizonyított véges állapotterű és bináris kiolvasási terű HMM-ekre, [1]. Ők bizonyították először a prediktív filter exponenciális felejtését. Leroux általános folytonos kiolvasási tér mellett bizonyította a maximum-likelihood becslés erős konzisztenciáját a szubadditív ergod tétel alkalmazásával.

Legland és Mevel vizsgálták a prediktív filter exponenciális felejtését általános kiolvasási tér mellett, és bizonyítottak geometriai ergodicitást arra a kibővített Markov láncra, mely az állapotból, kiolvasásból, és a prediktív filterből áll, [40] és [39]. Ezen utóbbi eredményeket terjesztették ki Douc és Matias általános kompakt állapotterű HMM-ekre, [9].

Rejtett Markov Modellek statisztikai vizsgálatának egyik alapvető kérdése, hogy a nagy számok törvénye létezik-e a log-likelihood függvényre. Az előbbiekben vázolt irodalmi hivatkozások mindegyikében találhatóak feltételek arra, hogy a nagy számok erős törvénye létezzen. Az eddigiekben a legjobb megközelítésnek a Legland és Mevel által használt eszköz, a geometriai ergodicitás tűnt. A geometriai ergodicitás segítségével centrális határeloszlás tételeket tudtak bizonyítani.

Lineáris sztochasztikus rendszerek statisztikai elméletéből ismert, hogy ezek a klasszikus eredmények nem mindig kielégítőek bizonyos természetesen adódó kérdések megválaszolására, mint például az adaptív előrejelzés hatékonysága. Erre vonatkoznak Gerencsérnek és Rissanen-nek az eredményei, lásd [28], [26]. Adaptív előrejelzők hatékonyságának vizsgálata tette szükségessé lineáris sztochasztikus rendszerek paraméter becslésének erős approximációját. Lineáris sztochasztikus rendszerek esetében az off-line becslés erős approximációjára vonatkozó eredmény van [24]-ben.

Az erős approximációra vonatkozó eredmények az L-keverőség fogalmán alapulnak, melyet Gerencsér vezetett be [23]-ben. Ez a fogalom a Caines, Rissanen [48] és Ljung [42] által használt exponenciálisan stabil folyamatok egy általánosítása.

Egy egyszerű, de fontos észrevétel, hogy az L-keverőség fogalma kiterjeszthető a Rejtett Markov Modellek osztályára. A kiterjesztéshez azt használjuk, hogy egy Markov lánc reprezentálható véletlen leképezésekkel, lásd [8]. Belátjuk például, hogy ha az állapot folyamatra teljesül a Doeblin-feltétel, akkor a Rejtett Markov folyamat tetszőleges, rögzített korlátos, mérhető függvénye L-keverő lesz, lásd 3.1 Tétel.

2. Alapfogalmak

2.1. Hidden Markov Models

Legyenek $\mathcal{X}, \mathcal{Y}$ lengyel terek. Tekintsünk egy Rejtett Markov Modellt, melynek állapottere $\mathcal{X}$ és a kiolvasási tere $\mathcal{Y}$.

2.1. Definíció. *Az (X_n, Y_n) pár Rejtett Markov folyamat, ha (X_n) egy homogén Markov lánc az $\mathcal{X}$ állapottéren és az (Y_n) kiolvasási sorozat feltételesen független és azonos eloszlású az (X_n) folyamat generálta σ-algebrára nézve.*

Az érthetőség kedvéért legyen a HMM állapottere most véges, azaz $|\mathcal{X}| = N$. Kompakt állapottérre vonatkozó fogalmak és eredmények a disszertáció 4.2 fejezetében találhatóak.

Legyen Q^* az ismeretlen Markov lánc átmenetvalószínűségi mátrixa, azaz

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

ahol $*$ jelöli azt, hogy a valódi értékét tekintjük az ismeretlen mennyiségnek. A disszertációban parametrikus problémával foglalkozunk, azaz az ismeretlen mennyiségek (átmenetvalószínűségi mátrix, kiolvasási sűrűségek) mindig valamilyen paramétertől függenek. A paraméter valódi értéke az az érték lesz, amelyet a sorozat generálásához használunk.

A kiolvasásokat a következő feltételes sűrűségfüggvényekkel definiáljuk

$$P(Y_n \in dy | X_n = x) = b^{*x}(y) \lambda(dy), \tag{1}$$

ahol λ egy rögzített σ-véges mérték.

A becsléselmélet egyik kulcs mennyisége a prediktív filter:

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

Legyen $p_{n+1} = (p^1_{n+1}, \ldots, p^N_{n+1})^T$. Ekkor [5] alapján a filter folyamat kielégíti a Baum-egyenletet:

$$p_{n+1} = \pi(Q^T B(Y_n) p_n), \tag{2}$$

ahol $p_0 = q$ a kezdeti eloszlás, $Q \in \mathbb{R}^{N \times N}$ egy valószínűségi mátrix, $B(y) = \operatorname{diag}(b^i(y))$ feltételes sűrűségfüggvényekből (valószínűségekből) álló diagonális mátrix és $q \in \mathbb{R}^N$ egy valószínűségi vektor, azaz $q^i \geq 0$, ha $i = 1, \ldots N$ és $\sum_{i=1}^N q^i = 1$.

Legyen q egy tetszőleges kezdeti valószínűségi vektor és jelölje $p_n(q)$ a Baum-egyenlet megoldását.

A statisztikai vizsgálatoknak egy kulcstényezője, hogy a Baum-egyenlet exponenciálisan stabil, azaz két különböző kezdeti vektorból indítva a rekurziót (q-ból és q'-ből), akkor $p_n(q)$ és $p_n(q')$ különbsége exponenciálisan gyorsan tart 0-hoz.

Megfelelő feltételek mellett ezt bizonyította be Legland és Mevel általános kiolvasási tér mellett ([40]). Az alábbiakban kimondjuk a tételt arra az esetre, amikor az átmenetvalószínűségi mátrix pozitív:

2.1. Állítás. *Tegyük fel, hogy $Q > 0$ és $b^x(y) > 0$ minden x, y-ra. Legyenek q, q' tetszőleges kezdeti értékek. Ekkor valamilyen $0 < \delta < 1$ mellett,*

$$\|p_n(q) - p_n(q')\|_{TV} \leq C(1-\delta)^n \|q - q'\|_{TV}, \tag{3}$$

ahol $\| \cdot \|_{TV}$ jelöli a totális variáció normát.

Legyen D egy nem üres nyílt részhalmaza az $\mathbb{R}^r$ térnek. Tekintsük a következő becslési feladatot: legyen $Q(\theta)$ és $b(\theta)$ a $\theta \in D$ függvénye, valamint legyen

$$Q^* = Q(\theta^*), \quad b^* = b(\theta^*).$$

A maximum-likelihood becslés konzisztencájához szükségünk van arra, hogy a

$$\lim_{N \to \infty} \frac{1}{N} \log p(y_0, \ldots y_N, \theta) \tag{4}$$

határérték 1-valószínűséggel létezik (egyenletesen θ-ban), [42].

A (4) határértéket sokszor vizsgálták az irodalomban, lásd [4], [41], [20], [34], [9], [10].

2.2. L-keverő folyamatok

Ebben a fejezetben áttekintést adunk az L-keverő folyamatokról. Az L-keverőség fogalma, melyet Gerencsér László vezetett be ([23]), hatékony eszköznek bizonyult a lineáris sztochasztikus rendszerek vizsgálatában. Kapcsolatot teremtve a Rejtett Markov Modellek és a lineáris sztochasztikus rendszerek között, az L-keverőség a disszertáció legfontosabb technikai eszköze. Az alábbiakban definiáljuk az L-keverő tulajdonságot:

Legyen $(\Omega, \mathcal{F}, P)$ egy adott valószínűségi mező. Tekintsünk egy $\mathbb{R}^m$-értékű (X_n), $n \geq 0$ sztochasztikus folyamatot ezen a mezőn.

2.2. Definíció. *Az (X_n), $n \geq 0$ folyamat M-korlátos, ha minden $1 \leq q < \infty$-re*

$$M_q(x) = \sup_{n \geq 0} E^{\frac{1}{q}} |X_n|^q < \infty.$$

Azt, hogy (X_n) M-korlátos, a későbbiekben úgy jelöljük, hogy $X_n = O_M(1)$. Hasonlóan, ha c_n egy pozitív tagokból álló sorozat, akkor $X_n = O_M(c_n)$ azt jelenti, hogy $X_n/c_n = O_M(1)$.

Legyen $(\mathcal{F}_n)$, $n \geq 0$ egy monoton növő σ-algebra család, illetve $(\mathcal{F}_n^+)$, $n \geq 0$ egy monoton fogyó σ-algebra család. Tegyük fel, hogy minden $n \geq 0$-ra, $\mathcal{F}_n$ és $\mathcal{F}_n^+$ függetlenek.

2.3. Definíció. *Az (X_n), $n \geq 0$ sztochasztikus folyamat L-keverő az $(\mathcal{F}_n, \mathcal{F}_n^+)$ párra nézve, ha $\mathcal{F}_n$-adaptált, M-korlátos, és tetszőleges $1 \leq q < \infty$ mellett*

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

ahol τ nem-negatív egész, teljesül, hogy

$$\Gamma_q(x) = \sum_{\tau=0}^{\infty} \gamma_q(\tau) < \infty.$$

3. Exponenciálisan stabil rendszerek

3.1. Markov láncok reprezentációja

A 3.1 fejezetben egy áttekintést adunk Markov láncok véletlen leképezésekkel történő reprezentációjáról, [8], majd közlünk néhány fontos állítást a fenti reprezentáció és a Doeblin feltétel kapcsolatáról, lásd [7]. Ezen eszközökkel bizonyítjuk a következő lemmát:

3.1. Lemma. *Tegyük fel, hogy az (X_n) Markov láncra teljesül a Doeblin feltétel. Ekkor az (X_n, Y_n) folyamatra szintén teljesülni fog a Doeblin feltétel.*

3.2. Markov láncok és az L-keverőség

Tekintsük a következő input-output rendszert: Legyen az input folyamat egy Markov lánc, mely kielégíti a Doeblin-feltételt, az output folyamatot pedig egy korlátos mérhető függvény generálja. Ekkor az output folyamatra nem teljesül a Doeblin-feltétel. A következő tétel azt mondja, hogy az output folyamat L-keverő lesz.

3.1. Tétel. *Legyen (X_n) egy Markov lánc, melynek állapottere az $\mathcal{X}$ lengyel tér. Tegyük fel, hogy a Doeblin-feltétel teljesül az X_n láncra. Továbbá legyen $g : \mathcal{X} \longrightarrow \mathbb{R}$ egy korlátos, mérhető függvény. Ekkor az*

$$U_n = g(X_n)$$

folyamat L-keverő.

3.3. Exponenciálisan stabil véletlen leképezések I.

Bevezetjük az exponenciális stabilitás általános fogalmát. Legyen $\mathcal{X}$ egy tetszőleges absztrakt mérhető tér és legyen $\mathcal{Z}$ egy Banach tér zárt részhalmaza (pl. $\mathcal{Z} \subset L_1(\mathbb{R})$ lehet a sűrűségfüggvények halmaza). Legyen $f : \mathcal{X} \times \mathcal{Z} \longrightarrow \mathcal{Z}$ egy Borel-mérhető függvény, és egy rögzített $(x_n)_{n\geq 0}$, $x_n \in \mathcal{X}$ sorozatra tekintsük a következő rekurziót:

$$z_{n+1} = f(x_n, z_n), \quad z_0 = \xi. \tag{5}$$

Jelöljük a rekurzió megoldását $z_n(\xi)$-vel. Az egyszerűbb jelölések kedvéért a generáló (x_n) sorozat jelölésétől eltekintünk.

3.1. Definíció. *Az f leképezés egyenletesen exponenciálisan stabil, ha minden $(x_n)_{n\geq 0}$, $x_n \in \mathcal{X}$ sorozatra fennáll, hogy*

$$\|z_n(\xi) - z_n(\xi')\| \leq C(1-\varrho)^n \|\xi - \xi'\|, \tag{6}$$

ahol $C > 0, 1 > \varrho > 0$ függetlenek az (x_n) sorozattól.

Bizonyos feltételek mellett ez teljesül a Baum-egyenletre, lásd 2.1 Állítás vagy [40].

Definiáljuk a (Z_n) folyamatot a következő módon:

$$Z_{n+1} = f(X_n, Z_n), \quad Z_0 = \xi, \tag{7}$$

ahol az (X_n) Markov lánc kielégíti a Doeblin feltételt. Jelöljük a Markov-lánc invariáns eloszlását π-vel. A (Z_n) folyamat M-korlátosságához kellenek a következő feltételek:

3.1. Feltétel. *Legyen az X_0 eloszlása π_0. Tegyük fel, hogy*

$$\frac{d\pi_0}{d\pi} \leq C_1. \tag{8}$$

3.2. Feltétel. *Minden $\xi \in \mathcal{Z}$ és $q \geq 1$-re tegyük fel, hogy*

$$E_\pi \|Z_1(\xi)\|^q \leq K_1(\xi) < \infty,$$

azaz

$$\int_{\mathcal{X}} \|f(x,\xi)\|^q d\pi(x) \leq K_1(\xi) < \infty, \tag{9}$$

ahol π az (X_n) stacionárius eloszlása és $K_1(\cdot)$ egy mérhető függvény.

3.2. Lemma. *Legyen $f(x,z)$ egy exponenciálisan stabil leképezés és tegyük fel, hogy a 3.1, 3.2 feltételek teljesülnek. Ekkor a (7) rekurzióval definiált (Z_n) folyamat tetszőleges $Z_0 = \xi$ kezdeti érték mellett M-korlátos.*

Legyen g egy mérhető függvény és tekintsük a $V_n = g(X_n, Z_n)$ folyamatot. Tegyük fel a következőt a g függvényre:

3.3. Feltétel. *Legyen $g(x,z)$ egy mérhető függvény az $\mathcal{X} \times \mathcal{Z}$ téren, mely Lipschitz-folytonos z-ben tetszőleges x esetén, és a Lipschitz konstans nem függ x-től.*

3.2. Tétel. *Tekintsük az (X_n, Z_n) folyamatot, ahol az (X_n) Markov láncra teljesül a Doeblin-feltétel, és Z_n-t a (7) rekurzió definiálja valamilyen exponenciálisan stabil f leképezéssel és ξ kezdeti értékkel. Tegyük fel, hogy a 3.1 és 3.2 feltételek teljesülnek. Továbbá legyen $g(x,z)$ egy korlátos, mérhető függvény, mely kielégíti a 3.3 feltételt. Ekkor a*

$$V_n = g(X_n, Z_n)$$

folyamat L-keverő.

Néhány alkalmazásban a 3.3 feltétel túl erős. Ezért tekintsük az alábbi gyengített verzióját:

3.4. Feltétel. *Legyen $g(x,z)$ egy mérhető függvény az $\mathcal{X} \times \mathcal{Z}$ téren, mely Lipschitz-folytonos z-ben tetszőleges x mellett olyan $L(x)$ x-től függő Lipschitz konstanssal, melynek Lipschitz konstans momentumai léteznek az (X_n) folyamat π stacionárius eloszlása szerint, azaz minden $q \geq 1$-re*

$$\int_{\mathcal{X}} |L(x)|^q d\pi(x) < L_q^q < \infty.$$

Kicserélve a 3.3 feltételt a 3.4 feltételre, 3.2 Tétel igaz marad.

3.4. Exponenciálisan stabil véletlen leképezések II.

Ebben a részben a 3.2 Tétel kiterjesztését vizsgáljuk nem-korlátos g függvényekre. A következő feltételre van szükségünk:

3.5. Feltétel. *Tegyük fel, hogy minden $q \geq 1$-re*

$$\int_{\mathcal{X}} \sup_{z \in \mathcal{Z}} \|g(x,z)\|^q d\pi(x) \leq M_q < \infty. \tag{10}$$

3.3. Tétel. *Tekintsük az (X_n, Z_n) folyamatot, ahol az (X_n) Markov láncra teljesül a Doeblin-feltétel, és Z_n-t a (7) rekurzió definiálja valamilyen exponenciálisan stabil f leképezéssel és ξ kezdeti értékkel. Tegyük fel, hogy a 3.1 és 3.2 feltételek teljesülnek. Továbbá tegyük fel, hogy a $g(x,z)$ függvényre teljesülnek a 3.3, 3.5 feltételek. Ekkor a*

$$V_n = g(X_n, Z_n)$$

folyamat L-keverő.

3.3 Tétel szintén igaz a gyengébb 3.4 feltétel mellett.

3.5. Exponenciálisan stabil véletlen leképezések III.

Erős approximációs tételekhez szükségünk lesz a $\frac{\partial}{\partial\theta}\log p(y_n|y_{n-1},\dots y_0,\theta)$ derivált folyamat L-keverőségére. Mivel 3.3 Tétel feltételei nem teljesülnek erre a folyamatra, szükségünk van a tétel további kiterjesztésére.

Tekintsük az alábbi feltételt a 3.3 feltétel helyett:

3.6. Feltétel. *Legyen $g(x,z)$ egy mérhető függvény az $\mathcal{X}\times\mathcal{Z}$ téren olyan, hogy minden x-re egy x-től függő $L(x)$ Lipschitz konstansra teljesül, hogy*

$$|g(x,z_1)-g(x,z_2)|\le L(x)\|z_1-z_2\|(\|z_1\|+\|z_2\|),$$

és

$$\left(\int_{\mathcal{X}}|L(x)|^q d\pi(x)\right)^{1/q}<L_q<\infty$$

minden $q\ge 1$-re, ahol $\pi(x)$ az (X_n) Markov lánc stacionárius eloszlása.

Gyengítsük 3.5 feltételt:

3.7. Feltétel. *Tegyük fel, hogy minden $q\ge 1$-re*

$$\int_{\mathcal{X}}\sup_{z\in\mathcal{Z}}\left(\frac{\|g(x,z)\|}{\|z\|+1}\right)^q d\pi(x)\le M_q<\infty. \tag{11}$$

Ekkor teljesül, hogy

3.4. Tétel. *Tekintsük az (X_n,Z_n) folyamatot, ahol az (X_n) Markov láncra teljesül a Doeblin-feltétel, és Z_n-t a (7) rekurzió definiálja valamilyen exponenciálisan stabil f leképezéssel és ξ kezdeti értékkel. Tegyük fel, hogy a 3.1 és 3.2 feltételek teljesülnek. Továbbá tegyük fel, hogy a $g(x,z)$ függvényre teljesülnek a 3.6, 3.7 feltételek. Ekkor a*

$$V_n=g(X_n,Z_n)$$

folyamat L-keverő.

3.6. On-line becslés

Ebben a fejezetben tárgyaljuk a Rejtett Markov Modellek rekurzív becslésének konvergenciájához szükséges általános modellt. Ehhez tekintünk egy Markov folyamatot, melyet egy exponenciálisan stabil leképezés hajt és használjuk Benveniste, Metivier és Priouret (BMP) ([6]) sztochasztikus approximációs algoritmusokra vonatkozó eredményeit. Az első alfejezetben vázoljuk a sztochasztikus approximációs algoritmus általános sémáját, a második alfejezetben pedig bebizonyítjuk, hogy a modellosztályunkra alkalmazható az algoritmus.

3.6.1. A BMP séma

Ebben a fejezetben vázoljuk a Benveniste, Metivier és Priouret (lásd a [6]-ban 2. Fejezet, II. Részt) szerzők sztochasztikus approximációs algoritmusokra vonatkozó alapvető eredményeit.

Legyen adott $\{\Pi_\theta,\ \theta \in D \subset R^d\}$ átmenetvalószínűségeknek egy családja az $\mathcal{U}$ téren, ahol $\mathcal{U}$ egy lengyel tér. Jelöljük a téren a metrikát d-vel. Megjegyezzük, hogy [6]-ban $\mathcal{U}$ azonosítva van R^n-nel, de az eredmények általánosíthatóak teljes szeparábilis metrikus térre. Legyen D egy nyílt halmaz. Tegyük fel, hogy minden $\theta \in D$-re létezik egy egyértelmű μ_θ invariáns mértéke a folyamatnak. Legyen $(U_n(\theta))$ egy olyan Markov lánc, melynek $U_0(\theta)$ kezdeti eloszlása μ_θ. Legyen $H(\theta, u)$ egy $R^d \times \mathcal{U} \to R^d$ leképezés. A BMP-elmélet alapfeladata a következő egyenlet megoldása:

$$E_{\mu_\theta} H(\theta, U(\theta)) = 0.$$

Tegyük fel, hogy egy $\theta^* \in D$ megoldás létezik.

A BMP-séma. A fenti egyenlet megoldására a következő rekurzív becslési eljárást definiáljuk

$$\theta_{n+1} = \theta_n + \frac{1}{n} H(\theta_n, U_n), \tag{12}$$

ahol U_n az időben változó paraméterrel definiált folyamatot jelenti, azaz

$$P(U_{n+1} \in A | \mathcal{F}_n) = \Pi_{\theta_n}(U_n, A),$$

ahol $\mathcal{F}_n$ az $U_0, \dots, U_n$ valószínűségi változók által generált σ-algebra és A egy Borel-mérhető részhalmaza az $\mathcal{X}$ térnek.

[6]-ban bizonyítják a következő konvergencia-tételt (13. Tétel, 236.old):

3.5. Tétel. *(Benveniste-Métivier-Priouret 1990, [6]) Tegyük fel, hogy a disszertációban szereplő* **A1 - A6** *feltételek teljesülnek és ϵ egy elegendően kicsi szám. Tegyük fel, hogy $\theta \in \mathrm{int} D_0$, $U_m = u \in \mathcal{U}$, és tekintsük a $\theta_n^\circ = \theta_{n \wedge \tau \wedge \sigma}$ megállított folyamatot. Ekkor tetszőleges $0 < \lambda < 1$ számhoz léteznek a B és s konstansok úgy, hogy minden $m \geq 0$-ra teljesül, hogy $\lim \theta_n^\circ = \theta^*$ legalább*

$$1 - B(1 + V(u)^s) \sum_{n=m+1}^{+\infty} n^{-1-\lambda}.$$

valószínűséggel.

Az **A1 - A6** feltételek megtalálhatóak a disszertáció 3.6.1 fejezetében, illetve [6]-ban.

3.6.2. Alkalmazás exponenciálisan stabil nem-lineáris rendszerekre

Ebben a fejezetben az (**A1**)-(**A3**) feltételek teljesülését ellenőrizzük exponenciálisan stabil nem-lineáris rendszerekre. Legyen $\mathcal{X}$ egy lengyel tér és $\mathcal{Z}$ egy szeparábilis Banach tér zárt részhalmaza. Jelölje $d_\mathcal{X}$ az $\mathcal{X}$ téren a metrikát.

Legyen f egy exponenciálisan stabil leképezés, lásd 3.1 Definíciót, és definiáljuk a (Z_n) folyamatot a

$$Z_{n+1} = f(X_n, Z_n, \theta), \quad Z_0 = \xi, \tag{13}$$

rekurzióval, ahol (X_n) egy olyan Markov lánc, mely kielégíti a Doeblin-feltételt. Legyen $U_n = (X_n, Z_n) \in \mathcal{X} \times \mathcal{Z} = \mathcal{U}$, és definiáljuk a metrikát $\mathcal{U}$-n a következőképpen:

$$d(u, u') = \|z - z'\| + d_{\mathcal{X}}(x, x'), \tag{14}$$

ahol $u = (x, z)$ és $u' = (x', z')$. Végül legyen

$$V(u) = \|z\|. \tag{15}$$

a Ljapunov függvény.

Jelöljük az (X_n) folyamat stacionárius eloszlását π-vel. (**A1**)-hez szükségünk lesz két feltételre: az első az, hogy ne legyenek egymástól "nagyon messze" állapotok, a második lényegében (**A1**) egy lépésben akkor, amikor az X_0 eloszlása stacionárius.

3.8. Feltétel. *Jelölje X_1 eloszlását π_1. Tegyük fel, hogy*

$$\frac{d\pi_1}{d\pi} \leq C_1. \tag{16}$$

3.9. Feltétel. *Tegyük fel minden $\xi \in \mathcal{Z}$-re és $p \geq 1$-re, hogy*

$$E_\pi \|Z_1(\xi)\|^p \leq K_1(1 + \|\xi\|^p),$$

azaz

$$\int_{\mathcal{X}} \|f(x, \xi)\|^p d\pi(x) \leq K_1(1 + \|\xi\|^p). \tag{17}$$

Megjegyezzük, hogy 3.8 Feltétel a 3.1 Feltétel egy módosított változata, míg 3.9 Feltétel a 3.2 Feltétel egy speciális esete.

3.6. Tétel. *Tekintsünk egy $U_n = (X_n, Z_n)$ folyamatot, ahol Z_n-t a (7) rekurzióval definiáljuk egy exponenciálisan stabil f leképezéssel és egy olyan (X_n) Markov lánccal, mely kielégíti a Doeblin-feltételt. Tegyük fel, hogy 3.8 és 3.9 Feltételek teljesülnek. Ekkor (**A1**) fennáll, azaz létezik olyan K pozitív konstans, hogy minden $n \geq 0$, $u \in \mathcal{U}$ és $\theta \in Q$-ra:*

$$E_{u,\theta}(|V(U_n)|^{p+1}) \leq K(1 + |V(u)|^{p+1}).$$

3.6 Tételben nem használjuk, hogy az $\mathcal{X}$ tér metrikus, azaz itt az $\mathcal{X}$ tér tetszőleges mérhető tér lehet. Másfelől a Doeblin tulajdonság csak az (X_n) Markov lánc stacionárius eloszlásának létezéséhez kell.

(**A2**)-höz további két feltételre lesz szükségünk az (X_n) folyamat stabilitásával kapcsolatban.

3.10. Feltétel. *Tegyük fel, hogy f Lipschitz folytonos x-ben, azaz*

$$\|f(x_1, z) - f(x_2, z)\| \leq L d_{\mathcal{X}}(x_1, x_2)$$

3.11. Feltétel. *Tegyük fel, hogy*

$$E d_{\mathcal{X}}(X_n, X_n') \leq K d_{\mathcal{X}}(X_0, X_0').$$

3.7. Tétel. *Tekintsünk egy $U_n = (X_n, Z_n)$ folyamatot, ahol Z_n-t a (7) rekurzió által definiáljuk egy exponenciálisan stabil f leképezéssel és egy olyan (X_n) Markov lánccal, mely kielégíti a Doeblin-feltételt. Tegyük fel, hogy 3.8, 3.9, 3.10 és 3.11 Feltételek teljesülnek. Ekkor (**A2**) fennáll, azaz léteznek olyan K, p pozitív konstansok és $0 < \rho < 1$, hogy minden $g \in Li(p)$, $\theta \in Q$, $n \geq 0$ és $u, u' \in \mathcal{U}$-ra:*

$$|\Pi_\theta^n g(u) - \Pi_\theta^n g(u')| \leq K \rho^n \|\Delta g\|_{V_p} (1 + |V(u)|^p + |V(u')|^p) d(u, u').$$

(**A3**)-hoz szükségünk van arra, hogy az f leképezés sima legyen a θ paraméterben, azaz $f : \mathcal{X} \times \mathcal{Z} \times \Theta \to \mathcal{Z}$ legyen egy olyan Borel-mérhető függvény, mely differenciálható θ-ban, és tetszőleges rögzített θ mellett az $f(\cdot, \cdot, \theta)$ exponenciálisan stabil leképezés.

3.8. Tétel. *Tekintsünk egy $U_n = (X_n, Z_n)$ folyamatot, ahol Z_n-t a (7) rekurzió által definiáljuk egy exponenciálisan stabil f leképezéssel és egy olyan (X_n) Markov lánccal, mely kielégíti a Doeblin-feltételt. Tegyük fel, hogy 3.8, 3.9 Feltételek teljesülnek. Ekkor (**A3**) fennáll, azaz léteznek olyan K, p pozitív konstansok, hogy minden $g \in Li(p)$, $u \in \mathcal{U}$, $n \geq 0$ és $\theta, \theta' \in Q$-ra:*

$$|\Pi_\theta^n g(u) - \Pi_{\theta'}^n g(u)| \leq K \|\Delta g\|_{V_p} (1 + |V(u)|^p) |\theta - \theta'|.$$

Összefoglalva a fentieket, kaptuk, hogy

3.9. Tétel. *Tekintsünk egy $U_n = (X_n, Z_n)$ folyamatot, ahol Z_n-t a (7) rekurzió által definiáljuk egy exponenciálisan stabil f leképezéssel és egy olyan (X_n) Markov lánccal, mely kielégíti a Doeblin-feltételt. Tegyük fel, hogy 3.8, 3.9, 3.10 és 3.11 Feltételek teljesülnek. Ekkor (**A1**)-(**A3**) fennállnak.*

Tehát ha (**A5**) fennáll valamilyen H függvényre és van olyan Ljapunov függvény, amely kielégíti (**A6**)-t, akkor a 3.5 Tétel szerinti konvergencia eredmény fennáll a (12) rekurzióval generált sorozatra.

3.9 Tétel eredményét használjuk Rejtett Markov Modellekre az 5. Fejezeteben.

4. Alkalmazás HMM-re

Ebben a fejezetben alkalmazzuk az előző fejezet eredményeit Rejtett Markov Modellekre. Legyen (X_n, Y_n) egy HMM, ahol az $\mathcal{X}$ állapottér véges és az $\mathcal{Y}$ kiolvasási tér folytonos, azaz legyen $\mathcal{Y}$ egy mérhető tér, $\mathcal{B}(\mathcal{Y})$ ezen a téren egy σ-algebra és λ egy σ-véges mérték. A gyakorlati alkalmazásokban az $\mathcal{Y}$ kiolvasási tér az $\mathbb{R}^d$ egy mérhető részhalmaza. Bár a fejezet eredményei igazak általános kiolvasási tér mellett is, mi a továbbiakban mindig feltesszük, hogy $\mathcal{Y}$ az $\mathbb{R}^d$ egy mérhető részhalmaza és λ a Lebesgue-mérték. Tegyük fel, hogy az átmenetvalószínűség-mátrix és a kiolvasási sűrűségek (valószínűségek) pozitívak, azaz $Q^* > 0$ és $b^{*i}(y) > 0$ minden i, y-ra. Ekkor a Doeblin-feltétel teljesül az (X_n, Y_n) folyamatra.

Legyen az (X_n) invariáns eloszlása ν, valamint az (X_n, Y_n) invariáns eloszlása π. Ekkor

$$\pi(\{i\}, dy) = \nu_i b^{*i}(y)\lambda(dy). \tag{18}$$

4.1. Rejtett Markov Modellek becslése

A becslési feladat egyik központi kérdése az, hogy a

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} g(y_k, p_k) \tag{19}$$

határérték létezik-e, ahol

$$g(y,p) = \log \sum_i b^i(y)p^i. \tag{20}$$

4.1. Tétel. *Tekintsünk egy (X_n, Y_n) Rejtett Markov Modellt, ahol az $\mathcal{X}$ állapottér véges és az $\mathcal{Y}$ kiolvasási tér az $\mathbb{R}^d$ tér egy mérhető részhalmaza. Tegyük fel, $Q, Q^* > 0$ és $b^i(y), b^{*i}(y) > 0$ minden i, y-ra. Indítsuk az (X_n, Y_n) folyamatot olyan π_0 kezdeti eloszlásból olyan, melynek Radon-Nikodym deriváltja a stacionárius π eloszlásra nézve korlátos, azaz*

$$\frac{d\pi_0}{d\pi} \le K. \tag{21}$$

Végül tegyük fel, hogy minden $i, j \in \mathcal{X}$-re és $q \ge 1$-re

$$\int |\log b^j(y)|^q b^{*i}(y)\lambda(dy) < \infty. \tag{22}$$

Ekkor a $g(Y_n, p_n)$ folyamat L-keverő.

Mivel a Q átmenetvalószínűségi mátrix pozitív, ezért az (X_n) folyamat stacionárius eloszlása szigorúan pozitív, így a pozitív kiolvasási sűrűségek (valószínűségek) miatt (21) feltétel nem egy erős feltétel. Például kezdeti eloszlásnak tekinthetjük azt, amikor az $\mathcal{X}$

téren veszünk egy állapotot egyenletes eloszlás szerint, majd egy mindenütt pozitív sűrűségfüggvénnyel generálunk hozzá egy kiolvasást.

(19) aszimptotikus viselkedését bizonyítjuk a következő tételben:

4.2. Tétel. *Tekintsünk egy (X_n, Y_n) Rejtett Markov Modellt, ahol az $\mathcal{X}$ állapottér véges és az $\mathcal{Y}$ kiolvasási tér az $\mathbb{R}^d$ tér egy mérhető részhalmaza. Tegyük fel, $Q, Q^* > 0$ és $b^i(y), b^{*i}(y) > 0$ minden i, y-ra. Indítsuk az (X_n, Y_n) folyamatot olyan π_0 kezdeti eloszlásból, melynek Radon-Nikodym deriváltja a stacionárius π eloszlásra nézve korlátos, azaz*

$$\frac{d\pi_0}{d\pi} \leq K. \tag{23}$$

Végül tegyük fel, hogy minden $i, j \in \mathcal{X}$-re és $q \geq 1$-re

$$\int |\log b^j(y)|^q b^{*i}(y)\lambda(dy) < \infty. \tag{24}$$

Ekkor a

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} g(Y_k, p_k)$$

határérték létezik 1-valószínűséggel.

Tekintsünk most egy véges állapotterű, véges kiolvasási terű HMM-t. Ekkor a 4.1 Tétel egy speciális esetét kapjuk:

4.3. Tétel. *Tekintsük az (X_n, Y_n) Rejtett Markov Modellt, ahol $\mathcal{X}$ és $\mathcal{Y}$ végesek. Tegyük fel, hogy $Q, Q^* > 0$ és $b^i(y), b^{*i}(y) > 0$ fennállnak minden i, y-ra. Ekkor a $g(Y_n, p_n)$ folyamat L-keverő.*

A fejezet végén összehasonlítjuk eredményeinket Legland és Mevel idevonatkozó eredményeivel, lásd [40] vagy a disszertáció 4.1.6 Állítását, és egy példát adunk olyan Rejtett Markov Modellre, ahol a 4.2 Tétel alkalmazható, de a 4.1.6 Állítás feltételei nem teljesülnek.

4.2. Kiterjesztés általános állapottérre

Ebben a fejezetben a 4.1 fejezet eredményeit terjesztjük ki arra az esetre, amikor az állapottér kompakt. Legyen (X_n) egy Markov lánc a $K \subset \mathcal{X}$ kompakt állapottéren, ahol $\mathcal{X}$ egy lengyel tér, és $\mathcal{B}(K)$ a megfelelő Borel σ-algebra. Rögzítsünk egy σ-véges domináló mértéket az $\mathcal{X}$ téren. Legyen $Q^*(x, A)$ ($x \in K$, $A \in \mathcal{B}(K)$) a Markov átmenet magja, lásd [44]. Az (Y_n) megfigyelések feltételesen függetlenek és azonos eloszlásúak feltéve az (X_n) folyamatot, ahol az $\mathcal{Y}$ kiolvasási tér lengyel tér. Jelölje $b^{*x_n}(y)$ a feltételes sűrűségfüggvényeket, lásd (1). Legyen az (X_n) folyamat kezdeti eloszlása P_0^*.

Tegyük fel, hogy a $b^x(y)$ sűrűségfüggvények ugyanazon λ σ-véges mérték szerint adottak, illetve a Q átmeneti magnak van egy q sűrűsége egy σ-véges μ domináló mérték szerint az $\mathcal{X}$ téren. Továbbá tegyük fel, hogy az (X_n) folyamat kezdeti eloszlásának is van egy p_0 sűrűsége a μ mértékre nézve.

Tekintsük a prediktív filter sűrűségfüggvényét, azaz az X_n változó sűrűségfüggvényét, ha az $(Y_i)_{i=0}^{n-1}$ múlt adott. A Baum-egyenlet alapján van egy rekurziónk a prediktív filter sűrűségfüggvényére (2):

$$p_{n+1}(x) = \frac{\int_u q(u,x) b^u(Y_n) p_n(u) d\mu(u)}{\int_u b^u(Y_n) p_n(u) d\mu(u)}.$$

A következő jelöléseket fogjuk használni: az $(K, \mathbf{B}(K), \mu)$ téren definiált tetszőleges f függvényre legyen

$$\operatorname{ess\,sup}(f) = \inf\{M \geq 0 : \mu(\{M < |f|\}) = 0\},$$

és ha f nem-negatív, akkor legyen

$$\operatorname{ess\,inf}(f) = \sup\{M \geq 0 : \mu(\{M > |f|\}) = 0\}.$$

Minden $y \in \mathcal{Y}$-ra legyen

$$\delta(y) = \frac{\operatorname{ess\,sup}_x b^x(y)}{\operatorname{ess\,inf}_x b^x(y)}. \tag{25}$$

Végül legyen

$$\epsilon = \frac{\operatorname{ess\,inf}_{x,x'} q(x,x')}{\operatorname{ess\,sup}_{x,x'} q(x,x')}. \tag{26}$$

A következő állítás a prediktív filter exponenciális stabilitását mutatja ([9]), mely a 2.1 Állítás általánosítása.

4.1. Állítás. *(Douc-Matias 2001, [9]) Tegyük fel, hogy $0 < \epsilon$. Legyenek p_0' és p_0'' az X_0 kezdeti eloszlás tetszőleges sűrűségfüggvényei a μ mértékre nézve. Ekkor*

$$\|p_n(p_0') - p_n(p_0'')\|_{L_1} \leq C(1-\epsilon)^n \|p_0' - p_0''\|_{L_1}. \tag{27}$$

4.2.1. HMM-k becslése: folytonos állapottér

Tegyük fel, hogy az (X_n) Markov lánc egy stacionárius eloszlása ν. Ekkor az (X_n, Y_n) folyamat stacionárius eloszlásának sűrűségfüggvénye

$$\pi(x,y) = b^x(y)\nu(x).$$

A likelihood függvény logaritmusa

$$\sum_{k=1}^{n-1} \log\left(\int_K b^x(Y_k) p_k \mu(dx)\right),$$

és definiáljuk a g függvényt úgy, hogy

$$g(y,p) = \log\left(\int_K b^x(y)p(x)\mu(dx)\right). \tag{28}$$

A következő tétel a 4.1 Tétel egy általánosítása.

4.4. Tétel. *Tekintsünk egy (X_n, Y_n) Rejtett Markov Modelt, ahol az állapottér $K \subset \mathcal{X}$ kompakt és az $\mathcal{Y}$ kiolvasási tér az $\mathbb{R}^d$ egy mérhető részhalmaza. Tegyük fel, hogy $\epsilon > 0$ és a Doeblin-feltétel teljesül az (X_n) Markov-láncra. Legyen az (X_n, Y_n) folyamat π_0 kezdeti eloszlása olyan, hogy a π stacionárius eloszlásra nézve a Radon-Nikodym derivált korlátos, azaz*

$$\frac{d\pi_0}{d\pi} \leq K. \tag{29}$$

Továbbá tegyük fel, hogy minden $q \geq 1$-re

$$\operatorname*{ess\,sup}_x \int |\log \operatorname*{ess\,sup}_{x'} b^{x'}(y)|^q b^{*x}(y)\lambda(dy) < \infty. \tag{30}$$

és

$$\operatorname*{ess\,sup}_x \int |\delta(y)|^q\, b^{*x}(y)\lambda(dy) < \infty. \tag{31}$$

Ekkor a $g(Y_n, p_n)$ folyamat L-keverő és a

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} g(Y_k, p_k)$$

határérték létezik 1-valószínűséggel.

5. Rejtett Markov Modellek rekurzív becslése

Legyen (X_n, Y_n) egy véges állapotterű, véges kiolvasási terű Rejtett Markov Modell. Tekintsük a következő becslési feladatot: legyenek Q és b a $\theta \in D$ paraméter függvényei, ahol D az $\mathbb{R}^r$ tér egy kompakt részhalmaza, továbbá legyenek

$$Q^* = Q(\theta^*), \quad b^* = b(\theta^*).$$

Tekintsük a paraméterfüggő Baum-egyenletet

$$\mathbf{p}_{n+1}(\theta) = \frac{Q^T(\theta)B(y_n,\theta)\mathbf{p}_n(\theta)}{\mathbf{b}(y_n,\theta)^T\mathbf{p}_n(\theta)} = \Phi_1(y_n, \mathbf{p}_n, \theta), \tag{32}$$

Az egyszerűség kedvéért a továbbiakban a θ paramétertől való függést nem jelöljük a kifejezésekben. Deriváljuk $\mathbf{p}_{n+1}$-t θ szerint:

$$W_{n+1} = Q^T \left(I - \frac{B(y_n)\mathbf{p}_n \mathbf{e}^T}{\mathbf{b}^T(y_n)\mathbf{p}_n} \right) \frac{B(y_n)W_n}{\mathbf{b}^T(y_n)\mathbf{p}_n} + F, \tag{33}$$

ahol

$$F = \frac{Q_\theta^T B(y_n)\mathbf{p}_n}{\mathbf{b}^T(y_n)\mathbf{p}_n} + Q^T \left(I - \frac{B(y_n)\mathbf{p}_n \mathbf{e}^T}{\mathbf{b}^T(y_n)\mathbf{p}_n} \right) \frac{\beta(y_n)\mathbf{p}_n}{\mathbf{b}^T(y_n)\mathbf{p}_n},$$

$W_n = \frac{\partial \mathbf{p}_n}{\partial \theta}$ és $\beta(y_n) = \frac{\partial B(y_n)}{\partial \theta}$.

Tömörebb formában kifejezve,

$$W_{n+1} = \Phi_2(y_n, \mathbf{p}_n, W_n, \theta).$$

azaz rögzített θ mellett az $u_n = (X_n, Y_n, \mathbf{p}_n, W_n, \theta)$ folyamat Markov lánc.

Legyen

$$\varphi_n(\theta) = \frac{\partial}{\partial \theta} \log p(y_n | y_{n-1}, \ldots, y_0, \theta)$$

a score-függvény. Felhasználva, hogy

$$\log p(y_n | y_{n-1}, \ldots, y_0, \theta) = \log \mathbf{b}^T(y)\mathbf{p}_n,$$

azt kapjuk, hogy

$$\varphi_n = \frac{\beta(y_n)\mathbf{p}_n + W_n \mathbf{b}(y_n)}{\mathbf{b}(y_n)^T \mathbf{p}_n}. \tag{34}$$

Legyen

$$H(\theta, u) = H(\theta, x, y, \mathbf{p}, W) = \frac{\beta(y,\theta)\mathbf{p} + W\mathbf{b}(y,\theta)}{\mathbf{b}(y,\theta)^T \mathbf{p}}, \tag{35}$$

és tekintsük a

$$\overline{\theta}_{n+1} = \overline{\theta}_n + \gamma_{n+1} H(\overline{\theta}_n, x_n, y_n, \overline{\mathbf{p}}_n, \overline{W}_n), \tag{36}$$

$$\overline{\mathbf{p}}_{n+1} = \Phi_1(y_n, \overline{\mathbf{p}}_n, \overline{\theta}_n), \tag{37}$$

$$\overline{W}_{n+1} = \Phi_2(y_n, \overline{\mathbf{p}}_n, \overline{W}_n, \overline{\theta}_n) \tag{38}$$

adaptív algoritmust.

Ennek az algoritmusnak a konveregenciájához használjuk fel a Benveniste, Metivier és Priouret féle megközelítést, lásd 3.6.1 fejezet, illetve [6]. A 3.9 Tétel feltételeit ellenőrizve kapjuk a következő eredményt véges állapotterű és véges kiolvasási terű HMM-ekre.

5.1. Tétel. *Legyen (X_n, Y_n) egy véges állapotterű és véges kiolvasási terű Rejtett Markov Modell. Tegyük fel, hogy $Q^* > 0$, $b^{*x}(y) > 0$, és $Q(\theta) > 0$, $b^x(y,\theta) > 0$ minden x, y-ra és $\theta \in D$-re, ahol D az $\mathbb{R}^d$ tér egy kompakt halmaza. Tegyük fel, hogy $Q(\theta)$ és $b(\theta)$ a θ paraméter sima függvényei, azaz a második deriváltak léteznek. Ekkor a 3.6.1 fejezet (***A1***)-(***A3***) és (***A5***) feltételei teljesülnek.*

Megjegyezzük, hogy ha az állapottér és a kiolvasási tér egyaránt véges, akkor az (**A4**) feltétel triviálisan teljesül.

(**A6**) teljesülését nehéz garantálni még lineáris sztochasztikus rendszerek esetén is. Legyen

$$h(\theta) = \lim_{n\to\infty} E\frac{\partial}{\partial\theta}\log p(Y_n|Y_{n-1},\dots Y_0,\theta). \tag{39}$$

Ez a határérték létezik (lásd 6.2 Tétel) és tegyük fel, hogy az alábbi identifikálhatósági feltétel teljesül:

5.1. Feltétel. *A $h(\theta)=0$ egyenletnek pontosan egy megoldása van D-ben, mégpedig θ^*.*

Az 5.1 feltételből következik, hogy (**A6**) teljesül egy kis tartományon. A fejezetet az alábbi tétellel zárjuk, mely a 3.5 Tétel egy alkalmazása.

5.2. Tétel. *Legyen (X_n, Y_n) egy véges állapotterű és véges kiolvasási terű Rejtett Markov Modell. Tegyük fel, hogy $Q^* > 0$, $b^{*x}(y) > 0$, és $Q(\theta) > 0$, $b^x(y,\theta) > 0$ minden x, y-ra és $\theta \in D$-re, ahol D az $\mathbb{R}^d$ tér egy kompakt halmaza. Tegyük fel, hogy $Q(\theta)$ és $b(\theta)$ a θ paraméter sima függvényei, azaz a második deriváltak léteznek. Végül tegyük fel, hogy az 5.1 identifikálhatósági feltétel teljesül. Ekkor a (36), (37) és (38) rekurziókkal definiált algoritmus konvergál a θ^* igazi paraméterhez $1-\epsilon$ valószínűséggel, ahol $\epsilon > 0$ tetszőleges pozitív szám.*

6. Rejtett Markov Modellek erős approximációja

6.1. Modell paraméterezése

Ebben a fejezetben azt vizsgáljuk, hogy a becsült paraméter milyen gyorsan tart a valódi paraméterértékhez. Legyen $G \subset \mathbb{R}^r$ egy nyílt halmaz, $D \subset G$ egy kompakt halmaz és $D^* \subset \mathrm{int}D$ egy másik kompakt halmaz, ahol $\mathrm{int}D$ jelöli a D halmaz belső pontjainak halmazát. Tegyük fel, hogy a θ^* igazi paraméterre teljesül, hogy $\theta^* \in D^*$, illetve a Rejtett Markov Modell θ becsült paraméterére fennáll, hogy $\theta \in D$. A továbbiakban D^* és D halmazokat kompakt tartományoknak nevezzük.

Tekintsük a következő becslési feladatot: legyenek Q és b a $\theta \in D$ paraméter függvényei. Továbbá legyen

$$Q^* = Q(\theta^*), \quad b^* = b(\theta^*).$$

Ebben a fejezetben mindig feltesszük, hogy az állapottér véges, a kiolvasási tér pedig folytonos. Bár a fejezet eredményei a 4. Fejezethez hasonlóan igazak általános kiolvasási tér mellett is, mi a továbbiakban mindig feltesszük, hogy $\mathcal{Y}$ az $\mathbb{R}^d$ egy mérhető részhalmaza és λ a Lebesgue-mérték.

6.2. A derivált folyamat L-keverősége

Az erős approximációs tételhez szükségünk van a $\frac{\partial^k}{\partial\theta^k}\log p(y_n, y_{n-1}, \ldots, y_0, \theta)$ derivált folyamatok L-keverőségére, $(k = 1, 2, 3)$.

Tetszőleges $y \in \mathcal{Y}$-ra legyen

$$\delta(y) = \frac{\max\limits_x b^x(y)}{\min\limits_x b^x(y)}, \tag{40}$$

valamint

$$\delta'(y) = \frac{\max\limits_x \|\partial b^x(y)/\partial\theta\|}{\min\limits_x b^x(y)}. \tag{41}$$

6.1. Tétel. *Tekintsük az (X_n, Y_n) Rejtett Markov Modellt, ahol az $\mathcal{X}$ állapottér véges és az $\mathcal{Y}$ kiolvasási tér az $\mathbb{R}^d$ egy mérhető részhalmaza. Tegyük fel, hogy $Q, Q^* > 0$ és $b^i(y), b^{*i}(y) > 0$ minden i, y-ra fennáll, valamint, hogy $Q(\theta)$ és $b(\theta)$ a θ paraméter sima függvényei. Legyen az (X_n, Y_n) folyamat π_0 kezdeti eloszlása olyan, hogy a π stacionárius eloszlásra vett Radon-Nikodym deriváltja korlátos, azaz*

$$\frac{d\pi_0}{d\pi} \leq K. \tag{42}$$

Tegyük fel továbbá, hogy

$$\int |\delta(y)|^q b^{*i}(y)\lambda(dy) < \infty, \tag{43}$$

$$\int |\delta(y)'|^q b^{*i}(y)\lambda(dy) < \infty. \tag{44}$$

Ekkor a

$$\frac{\partial}{\partial\theta}\log p(y_n|p_{n-1}, \ldots p_0, \theta)$$

folyamat L-keverő.

Alkalmazásokban szükségünk van arra, hogy a deriváltfolyamat várhatóértékének létezik a határértéke (lásd (39) illetve (49)).

6.2. Tétel. *A 6.1 Tétel feltételei mellett a*

$$\lim_{n\to\infty} E\frac{\partial}{\partial\theta}\log p(y_n|y_{n-1}, \ldots y_0, \theta)$$

határérték létezik.

Hasonló tételekre van szükség a második és harmadik deriváltakra.

6.3. A becslés hibájára vonatkozó karakterizációs tétel

Tekintsünk egy (X_n, Y_n) Rejtett Markov Modellt, ahol az $\mathcal{X}$ állapottér véges és az $\mathcal{Y}$ kiolvasási tér az $\mathbb{R}^d$ tér egy mérhető részhalmaza. Tegyük fel, hogy $Q(\theta), Q^* > 0$ és $b^i(y,\theta), b^{*i}(y) > 0$ minden i, y-ra, valamint azt, hogy az (X_n, Y_n) folyamat π_0 kezdeti eloszlásának Radon-Nikodym deriváltja a π stacionárius eloszlásra nézve véges, azaz

$$\frac{d\pi_0}{d\pi} \leq K. \tag{45}$$

Végezetül tegyük fel, hogy minden $i, j \in \mathcal{X}$, $\theta \in D$ és $q \geq 1$-re

$$\int |\log b^j(y,\theta)|^q b^{*i}(y)\lambda(dy) < \infty. \tag{46}$$

Az ismeretlen paraméter becslésére a maximum-likelihood (ML) becslést használjuk. Legyen

$$L_N = \sum_{n=1}^{N} \log p(Y_n | Y_{n-1}, \ldots, Y_0, \theta)$$

a log-likelihood függvény. Ezt a kifejezést tekintjük a θ paraméter szerinti költségfüggvényének. A kifejezés jobb oldala függ θ^*-tól is, hiszen az (Y_n) sorozatot θ^* szerint generáljuk. Az L_N függvény tehát függ mind a θ, mind a θ^* paraméterektől. A θ^* paraméter $\widehat{\theta}_N$ maximum-likelihood becslése legyen a

$$\frac{\partial}{\partial\theta} L_N(\theta, \theta^*) = L_{\theta N}(\theta, \theta^*) = 0 \tag{47}$$

egyenlet megoldása.

Legyen az aszimptotikus költségfüggvény

$$W(\theta, \theta^*) = \lim_{n\to\infty} E_{\theta^*} \log p(Y_n | Y_{n-1}, \ldots, Y_0, \theta). \tag{48}$$

Tegyük fel, hogy a $W(\theta, \theta^*)$ függvény sima a D tartomány belsejében, azaz a harmadik deriváltja létezik. A 6.2 Tétel feltételei mellett

$$W_\theta(\theta, \theta^*) = \lim_{n\to\infty} E_{\theta^*} \frac{\partial}{\partial\theta} \log p(Y_n | Y_{n-1}, \ldots, Y_0, \theta), \tag{49}$$

és a Fisher-információs mátrix

$$I^* = W_{\theta\theta}(\theta^*, \theta^*) =$$

$$\lim_{n\to\infty} E_{\theta^*} \left((\frac{\partial}{\partial\theta} \log p(y_n | y_{n-1}, \ldots, y_0, \theta^*))^T (\frac{\partial}{\partial\theta} \log p(y_n | y_{n-1}, \ldots, y_0, \theta^*)) \right).$$

6.1. Megjegyzés. *Megjegyezzük, hogy* $W_\theta(\theta^*, \theta^*) = 0$.

Tekintsük az alábbi identifikálhatósági feltételt:

6.1. Feltétel. *A*

$$W_\theta(\theta, \theta^*) = 0$$

egyenletnek pontosan egy megoldása van a D tartományban, mégpedig θ^*.

A [24] cikk gondolatmenetét követve az alábbi karakterizációs tételt bizonyítjuk a maximum-likelihood becslés hibájának nagyságára.

6.3. Tétel. *Tekintsünk egy (X_n, Y_n) Rejtett Markov Modellt, ahol az $\mathcal{X}$ állapottér véges és az $\mathcal{Y}$ kiolvasási tér az $\mathbb{R}^d$ tér egy mérhető részhalmaza. Tegyük fel, hogy $Q(\theta), Q^* > 0$ és $b^i(y,\theta), b^{*i}(y) > 0$ minden i, y-ra. Tegyük fel, hogy a 4.1 és a 6.1 Tétel feltételei, illetve a 6.1 identifikálhatósági feltétel teljesül. Legyen $\hat{\theta}_N$ a paraméter maximum-likelihood becslése. Ekkor*

$$\hat{\theta}_N - \theta^* = -(I^*)^{-1} \frac{1}{N} \sum_{n=1}^{N} \frac{\partial}{\partial \theta} \log p(Y_n | Y_{n-1}, \ldots, Y_0, \theta^*) + O_M(N^{-1}), \tag{50}$$

ahol I^* a Fisher-információs mátrix.

A tétel lényege az, hogy a főtag egy martingál, míg a hiba tag $O_M(N^{-1})$ nagyságrendű. Így minden határeloszlási tétel, ami teljesül a főtagra, igaz lesz a $\hat{\theta}_N - \theta^*$-re is.

Kimondjuk a tételt arra az esetre is, amikor a kiolvasási tér véges.

6.4. Tétel. *Tekintsünk egy (X_n, Y_n) Rejtett Markov Modellt, ahol mind az állapottér, mind a kiolvasási tér véges. Tegyük fel, hogy $Q(\theta), Q^* > 0$ és $b^i(y,\theta), b^{*i}(y) > 0$ minden i, y-ra, illetve a $Q(\theta)$, $b(\theta)$ függvények sima függvények θ-ban, azaz a harmadik deriváltak léteznek. Tegyük fel továbbá, hogy a 6.1 identifikálhatósági feltétel teljesül. Legyen $\hat{\theta}_N$ a maximum-likelihood becslés. Ekkor*

$$\hat{\theta}_N - \theta^* = -(I^*)^{-1} \frac{1}{N} \sum_{n=1}^{N} \frac{\partial}{\partial \theta} \log p(Y_n | Y_{n-1}, \ldots, Y_0, \theta^*) + O_M(N^{-1}), \tag{51}$$

ahol I^* a Fisher-információs mátrix.

7. Becslés felejtés esetén

Ha a rendszer dinamikája lassan változik, akkor mindig igazodnunk kell az aktuális állapothoz. Ekkor ahelyett, hogy minden múltbeli adatot egyformán veszünk számításba, a régi adatokat kisebb súllyal kell figyelembe vennünk. Tehát egyfajta felejtési faktorral kell módosítani a becslési eljárást. Technikailag ezt úgy valósítják meg, hogy beépítenek egy exponenciális felejtést az eredeti becslési eljárásba.

Tekintsük a következő módosított maximum-likelihood becslést: legyen $\widehat{\theta}_N(\lambda)$ a θ^* paraméternek az a becslése, amely minimalizálja a

$$\sum_{n=1}^{N}(1-\lambda)^{N-n}\lambda \log p(y_n|y_{n-1},\ldots,y_0;\theta), \tag{52}$$

függvényt, ahol $0 < \lambda < 1$. A λ-t nevezzük felejtési faktornak: minél kisebb a λ, annál lassabb a felejtés.

Legyen

$$L_N^{\lambda}(\theta,\theta^*) = \sum_{n=1}^{N}(1-\lambda)^{N-n}\lambda \log p(Y_n|Y_{n-1},\ldots,Y_0,\theta).$$

Ez utóbbi kifejezést nevezzük a módosított maximum-likelihood függvényhez tartozó költségfüggvénynek. A kifejezés jobb oldala függ θ^*-tól is, hiszen az (Y_n) sorozatot θ^* szerint generáljuk.

Könnyen látható, hogy a költségfüggvény rekurzívan is számolható:

$$L_N^{\lambda}(\theta,\theta^*) = (1-\lambda)L_{N-1}^{\lambda}(\theta,\theta^*) + \lambda \log p(Y_N|Y_{N-1},\ldots,Y_0,\theta).$$

Tehát az utolsó megfigyeléshez tartozó korrekciós tag mindig ugyanakkora súllyal szerepel. Ez a reprezentáció mutatja, hogy miért hívják ezt a típusú becslési eljárást "fixed-gain" becslésnek.

Tehát a θ^* paraméternek a módosított ML becsléssel kapott $\widehat{\theta}_N(\lambda)$ becslése a

$$\frac{\partial}{\partial\theta}L_N^{\lambda}(\theta,\theta^*) = L_{\theta N}^{\lambda}(\theta,\theta^*) = 0 \tag{53}$$

egyenlet megoldása.

A 4.1 Tétel és a 6.2 fejezet eredményeit felhasználva, és követve a [25] gondolatmenetét kapjuk az alábbi reprezentációs tételt, mely a 6.3 Tételnek egy módosított változata.

7.1. Tétel. *A 6.3 Tétel feltételei mellett*

$$\hat{\theta}_N(\lambda) - \theta^* = -I(\theta^*)^{-1}\sum_{n=1}^{N}(1-\lambda)^{N-n}\lambda\frac{\partial}{\partial\theta}\log p(Y_n|Y_{n-1},\ldots Y_0,\theta^*) + r_N,$$

ahol $0 < \alpha < 1$, $r_N = O_M(\lambda) + O_M(\alpha^N)$, *és* $I(\theta^*)$ *a Fischer-információs mátrix.*

7.1 Tétel következményeként kapjuk, hogy a kovariancia mátrixra fennáll, hogy

$$E(\widehat{\theta}_{n-1}-\theta^*)(\widehat{\theta}_{n-1}-\theta^*)^T = \frac{\lambda}{2}I(\theta^*)^{-1} + O(\lambda^{3/2}) + o(1). \tag{54}$$

8. Változás detektálása Rejtett Markov Modellekre

[3] alapján vizsgáljuk Rejtett Markov Modellek változás detektálásának problémáját. Először megjegyezzük, hogy a log-likelihood függvény -1-szeresét felfoghatjuk úgy is, mint egy kódhosszt modulo egy konstans, amit az $(y_N, \dots, y_1)$ sorozat kódolásával kapunk, amikor a kimenetelek együttes sűrűségfüggvénye $p(y_N, \dots, y_0; \theta)$.

A sztochasztikus komplexitás elméletében központi elem a likelihood függvénynek ez a felfogása. Legyen

$$C_n(y_n; \theta) \stackrel{\Delta}{=} -\log p(y_n | y_{n-1}, \dots, y_0; \theta),$$

a kódhossz.

[26] eredményeit terjesztjük ki Rejtett Markov Modellekre.

8.1. Tétel. *A 6.3 Tétel feltételei mellett*

$$\mathrm{E}(C_n(Y_n, \widehat{\theta}_{n-1}(\lambda)) - C_n(Y_n, \theta^*) = \frac{1}{2} r\lambda + O(\lambda^{3/2-c''}) + o(1),$$

ahol $c'' > 0$ egy tetszőleges kicsi konstans, és $r = \dim \theta$.

A tétel azt jelenti, hogy minél gyorsabb a felejtés, azaz minél közelebb van λ az 1-hez, annál többet vesztünk a kódoláskor. Egy egyszerű következménye a 8.1 Tételnek, hogy

8.1. Állítás. *Tekintsünk két különböző, $0 < \lambda_1 < \lambda_2 < 1$ felejtési faktort. Ekkor*

$$\mathrm{E}(C_n(y_n, \widehat{\theta}_{n-1}(\lambda_1)) - C_n(y_n, \widehat{\theta}_{n-1}(\lambda_2))) \simeq \frac{1}{2} r(\lambda_1 - \lambda_2) < 0.$$

8.1 Tétel hasznos a model kiválasztás szempontjából. Azonban ez az elméleti eredmény a konkrét realizációt tekintve nem mindig célravezető. Ezért vizsgáljuk trajektóriánként a prediktív hiba karakterizációját. Legyen a kumulatív hiba

$$S_N(\lambda) = \sum_{n=1}^{N} (C_n(y_n, \widehat{\theta}_{n-1}(\lambda)) - C_n(y_n, \theta^*))$$

8.2. Tétel. *A 6.3 Tétel feltételei mellett*

$$\limsup_{N \to \infty} |\frac{1}{N} S_N(\lambda) - \frac{\lambda}{2} r| \le C\lambda^{3/2}.$$

Hasonlóan a korábbiakhoz, egy egyszerű következménye az előző tételnek:

8.2. Állítás. *Legyenek $0 < \lambda_1 < \lambda_2 < 1$ különböző felejtési faktorok. Ekkor*

$$\limsup_{N \to \infty} |\frac{1}{N} S_N(\lambda_1) - \frac{1}{N} S_N(\lambda_2) - \frac{\lambda_1 - \lambda_2}{2} r| \le C\lambda_2^{3/2}.$$

Végül tegyük fel, hogy a τ időpillanatban változás következik be a paraméterben : az igazi paraméter legyen θ_1 amikor $n \leq \tau$ és θ_2 amikor $n \geq \tau + 1$, azaz

$$\theta^* := \begin{cases} \theta_1, & \text{if} \quad n \leq \tau \\ \theta_2, & \text{if} \quad n \geq \tau \end{cases}$$

Legyenek $0 < \lambda_1 < \lambda_2 < 1$. Ekkor a 8.2 Állítás alapján $N \leq \tau$-ra

$$S_N(\lambda_1) - S_N(\lambda_2) \approx \frac{\lambda_1 - \lambda_2}{2} Nr.$$

Másfelől egy kicsivel a változás után a gyorsabb felejtés lesz a jobb. Tekintsük az alábbi algoritmust a változás detektálására:

Algoritmus: Legyen $d(N) := S_N(\lambda_1) - S_N(\lambda_2)$ és

$$d_N^* = \min_{n \leq N} d(n).$$

Jelzünk, ha $d(N) - d_N^* > \epsilon$, ahol $\epsilon > 0$ egy előre lerögzített konstans. Az ilyen típusú algoritmusokat az irodalomban Hinkley detektornak nevezik, lásd [12].

Publikációk

- Gerencsér, L., Molnár-Sáska, G,. A New Method for the Analysis of Hidden Markov Model Estimates, Proceedings of the 15th Triennial World Congress of the International Federation of Automatic Control, Barcelona, 2002., T-Fr-M03
- Gerencsér L., Molnár-Sáska G., Michaletzky Gy., Tusnády G., Vágó Zs., New methods for the statistical analysis of Hidden Markov Models, Proceedings of the 41th IEEE Conference on Decision & Control, Las Vegas, 2002., WeP09-6 2272-2277.
- Gerencsér L., Molnár-Sáska G., Adaptive encoding and prediction of Hidden Markov processes, In proceedings of the European Control Conference, ECC2003, Cambridge, 2003.,
- Gerencsér L., Molnár-Sáska G., Estimation error in adaptive prediction of Hidden Markov Processes, In proceeding of the 11th Mediterranean Conference on Control and Automation MED03, Rhodes, 2003.
- Gerencsér L., Molnár-Sáska G., Estimation and Strong Approximation of Hidden Markov Models, Lecture Notes in Control and Information Sciences, Springer, vol. 294., 313-320., 2003.
- Gerencsér L., Molnár-Sáska G., Change detection of Hidden Markov Models, Proceedings of the 43th IEEE Conference on Decision & Control, 1754-1758, 2004.
- Gerencsér L., Molnár-Sáska G., Michaletzky Gy., Tusnády G., A new approach for the statistical analysis of Hidden Markov Models, IEEE Transactions on Automatic Control, benyújtva

Hivatkozások

[1] A. Arapostathis and S.I. Marcus. Analysis of an Identification Algorithm Arising in the Adaptive Estimation of Markov Chains. *Math. Control Signals Systems*, 3:1–29., 1990.

[2] R. Atar and O. Zeitouni. Exponential stability for nonlinear filtering. *Ann. Inst. H. Poincaré Probab. Statist.*, 33 (6):697–725, 1997.

[3] J. Baikovicius and L. Gerencsér. Change point detection in a stochastic complexity framework. In *Proc. of the 29-th IEEE CDC*, volume 6, pages 3554–3555, 1990.

[4] A. R. Barron. The strong ergodic theorem for densities: Generalized Shannon-McMillan-Breiman theorem. *The Annals of Probability*, 13:1292–1303, 1985.

[5] L.E. Baum and T. Petrie. Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Stat.*, 37:1559–1563, 1966.

[6] A. Benveniste, M. Métivier, and P. Priouret. *Adaptive algorithms and stochastic approximations*. Springer-Verlag, Berlin, 1990.

[7] R. Bhattacharya and E. C. Waymire. An approach to the existence of unique invariant probabilities for Markov processes. *Limit theorems in probability and statistics, János Bolyai Math. Soc.*, I (Balatonlelle 1999):181–200, 2002.

[8] V. S. Borkar. On white noise representations in stochastic realization theory. *SIAM J. Control Optim.*, 31:1093–1102, 1993.

[9] R. Douc and C. Matias. Asymptotics of the Maximum likelihood estimator for general Hidden Markov Models. *Bernoulli*, 7:381–420, 2001.

[10] R. Douc, É. Moulines, and T. Ryden. Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *Annals of Statistics*, 32:2254–2304, 2004.

[11] T.E. Duncan, B. Pasik-Duncan, and L. Stettner. Some Results on Ergodic and Adaptive Control of Hidden Markov Models. In *Proceedings of the 41th IEEE Conference on Decision & Control, Las Vegas*, pages WeA07–1, 1369–1374, 2002.

[12] D.V.Hinkley. Inference about the change-point from Cumulative Sum Tests. *Biometrika*, 58 (3):509–523., 1971.

[13] R. J. Elliott, W. P. Malcolm, and A. Tsoi. HMM Volatility Estimation. In *Proceedings of the 41th IEEE Conference on Decision & Control, Las Vegas*, pages TuA 12–6, 398–404, 2002.

[14] R.J. Elliott and J.B. Moore. Almost sure parameter estimation and convergence rates for Hidden Markov models. *Systems and Control Letters*, 32:203–207., 1997.

[15] Y. Ephraim and N. Merhav. Hidden Markov Processes. *IEEE Transactions on Information Theory*, 48:1508–1569., 2002.

[16] X. Feng, K.A. Loparo, Y. Ji, and H.J. Chizeck. Stochastic stability properties of jump linear systems. *IEEE Transactions on Automatic Control*, 37:38–53., 1992.

[17] L. Finesso, L. Gerencsér, and I. Kmecs. Estimation of parameters from quantized noisy observations. In *Proceedings of the European Control Conference, ECC99, Karlsruhe*, pages AM–3, F589, 6p., 1999.

[18] L. Finesso, L. Gerencsér, and I. Kmecs. A randomized EM-algorithm for estimating quantized linear Gaussian regression. In *Proceedings of the 38th IEEE Conference on Decision and Control, Phoenix*, pages 5100–5101., 1999.

[19] L. Finesso, C.C. Liu, and P. Narayan. The optimal error exponent for Markov order estimation. *IEEE Trans. Inform. Theory*, 42:1488–1497, 1996.

[20] C. Francq and M. Roussignol. Ergodicity of autoregressive processes with Markov-switching and consistency of the maximum-likelihood estimator. *Statistics*, 32:151–173., 1998.

[21] H. Furstenberg and H. Kesten. Products of random matrices. *Ann. Math. Statist.*, 31:457–469., 1960.

[22] S. Geman. Some averaging and stability results for random differential equations. *SIAM Journal of Applied Mathematics*, 36:87–105, 1979.

[23] L. Gerencsér. On a class of Mixing Processes. *Stochastics*, 26:165–191, 1989.

[24] L. Gerencsér. On the martingale approximation of the estimation error of ARMA parameters. *Systems & Control Letters*, 15:417–423, 1990.

[25] L. Gerencsér. Fixed gain off-line estimators of ARMA parameters. *Journal of Mathematical Systems, Estimation and Control*, 4(2):249–252., 1994.

[26] L. Gerencsér. On Rissanen's Predictive Stochastic Complexity for Stationary ARMA Processes. *Statistical Planning and Inference*, 41:303–325, 1994.

[27] L. Gerencsér and J. Baikovicius. A computable criterion for model selection for linear stochastic systems. In L. Keviczky and Cs. Bányász, editors, *Identification and System Parameter Estimation, Selected papers from the 9th IFAC-IFORS Symposium, Budapest*, volume 1, pages 389–394, Pergamon Press,Oxford, 1991.

[28] L. Gerencsér and J. Rissanen. A prediction bound for Gaussian ARMA processes. *Proc. of the 25th Conference on Decision and Control, Athens*, 3:1487–1490., 1986.

[29] P. Hall and C.C. Heyde. *Martingale Limit Theory and Its Application.* Academic Press, 1980.

[30] X.D. Huang, Y. Ariki, and M.A. Jack. *Hidden Markov models for speech recognition.* Edinburgh University Press, 1990.

[31] I.W. Hunter, L.A. Jones, M. Sagar, S.R. Lafontaine, and P.J. Hunter. Opthalmic microsurgical robot and associated virtual environment. *Computers in Biology and Medicine*, 25:173–182., 1995.

[32] I. Ibragimov and R. Khasminskii. *Statistical Estimation. Asymptotic Theory.* Springer Verlag, Berlin, 1981.

[33] Y. Kifer. Ergodic Theory of Random Transformation. *Progress in Probability and Statistics*, 10, 1986.

[34] V. Krishnamurthy and T. Rydén. Consistent estimation of linear and nonlinear autoregressive models with Markov regime. *J. Time Ser. Anal.*, 19 (3):291–307., 1998.

[35] V. Krishnamurthy and G. Yin. Recursive algorithms for estimation of hidden Markov models and autoregressive models with Markov regime. *IEEE Trans. Inform. Theory*, 48(2):458–476, 2002.

[36] H.J. Kushner and G. Yin. *Stochastic Approximation Algorithms and Applications.* Springer-Verlag, New York, 1997.

[37] F. LeGland and L. Mevel. Recursive Identification of HMM's with Observation in a Finite Set. In *Proc. of the 34th IEEE CDC*, pages 216–221, 1995.

[38] F. LeGland and L. Mevel. Recursive Estimation in Hidden Markov Models. In *Proc. of the 36th IEEE CDC*, pages 3468–3473, 1997.

[39] F. LeGland and L. Mevel. Basic Properties of the Projective Product with Application to Products of Column-Allowable Nonnegative Matrices. *Mathematics of Control, Signals and Systems*, 13:41–62, 2000.

[40] F. LeGland and L. Mevel. Exponential Forgetting and Geometric Ergodicity in Hidden Markov Models. *Mathematics of Control, Signals and Systems*, 13:63–93, 2000.

[41] B.G. Leroux. Maximum-likelihood estimation for Hidden Markov-models. *Stochastic Processes and their Applications*, 40:127–143, 1992.

[42] L. Ljung. On consistency and identifiability. *Mathematical Programming Study*, 5:169–190., 1976.

[43] L. Mevel. *Statistique asymptotique pour les modéles de Markov cachés.* Doctoral Thesis, Université de Rennes, 1997.

[44] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability.* Springer Verlag, London, 1993.

[45] J. Neveu. *Discrete-Parameter Martingales.* North-Holland Publishing Company, 1975.

[46] J. Rissanen. Stochastic complexity and predictive modelling. *Annals of Statistics*, 14(3):1080–1100, 1986.

[47] J. Rissanen. *Stochastic complexity in statistical inquiry.* World Scientific Publisher, 1989.

[48] J. Rissanen and P.E. Caines. The strong consistency of maximum likelihood estimators for ARMA processes. *Ann. Statist.*, 7:297 – 315., 1979.

[49] J. Rissanen and S. Forchhammer. Partially Hidden Markov Models. *IEEE Trans. on Information Theory*, 42:1253–1256., 1996.

[50] T. Rydén. On recursive estimation for hidden Markov models. *Stochastic Process. Appl.*, 66 (1):79–96, 1997.

[51] E. Seneta. *Non-negative Matrices and Markov Chains.* Springer Verlag, New York, 1981.

[52] L. Shue, S. Dey, B.D.O. Anderson, and F. De Bruyne. Remarks on Filtering Error due to Quantisation of a 2-state Hidden Markov Model. In *Proceedings of the 40th IEEE Conference on Decision & Control*, pages FrA05, 4123–4124., 1999.

[53] G.E. Tusnady and I. Simon. Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J Mol Biol.*, 283(2):489–506., 1998.