Eötvös Loránd Tudományegyetem

Természettudományi Kar

Ferenczi Dóra

Loglineáris és grafikus modellek

Szakdolgozat

Alkalmazott Matematikus MSc

Témavezető:

Dr. Michaletzky György
egyetemi tanár

Valószínűségelméleti és Statisztika Tanszék

Budapest, 2017

Tartalomjegyzék

Köszönetnyilvánítás

Hálával tartozom Michaletzky György Tanár Úrnak, aki temérdek elfoglaltsága mellett is mindig szakított rám időt, segítette munkámat hasznos észrevételeivel és tanácsaival. Köszönöm az egyetem minden oktatójának, akikkel a tanulmányaim során megismerkedhettem, hogy tudásukkal hozzájárultak szakmai fejlődésemhez. Nagyon hálás vagyok, amiért fantasztikus tanároktól tanulhattam izgalmas matematikát.

Végül szeretnék köszönetet mondani a családomnak a sok támogatásért és bátorításért, amit a tanulmányaim során kaptam tőlük.

Bevezetés

Szakdolgozatom fő célja olyan módszerek bemutatása, amelyek lehetőséget adnak valószínűségi változók közötti függőségi viszonyok feltárására. Amikor a megfigyeléseink több különböző kategorikus változóból származnak, akkor a változók értékeinek együttes előfordulási gyakoriságait kontingencia tábla segítségével ábrázolhatjuk. Kontingencia táblák elemzése loglineáris modellben vett maximum likelihood becslés segítségével az egyik leghatékonyabb statisztikai eszköz,melyet gyakran alkalmaznak orvosi kutatásokban, illetve társadalomtudományokban egyaránt.
A második fejezettől kezdve grafikus modellekkel foglalkozunk. A grafikus modellek előnye, hogy könnyen vizualizálható és jól interpretálható módon jelenítenek meg komplex rendszereket. Elsőként a loglineáris modell grafikus felépítését tekintjük át és explicit formulát adunk a maximum likelihood becslésre felbontható hipergráfok tulajdonságait kihasználva. Végül megismerkedünk irányítatlan gráfok segítségével definiált Markov-mezőkkel és ezek alkalmazásával a képfeldolgozás területén.

1. fejezet

A loglineáris modell

1.1. Maximum likelihood becslés speciális esetekben

Legyen $A = A_1 \times \cdots \times A_k$ véges szorzathalmaz az alaphalmaz, amely az $X = (X_1, \ldots, X_k)$ valószínűségi változók lehetséges értékeit tartalmazza, továbbá P egy k-dimenziós, ismeretlen eloszlás A-n.

Célunk a maximum-likelihood becslés meghatározása egy függőségi struktúra mellett.
Legyen $a \in A$, ahol $a = (a_1, \ldots, a_k)$ és $a_j \in A_j$, valamint jelölje $\mid A_j \mid = n_j$ a j-edik valószínűségi változók által felvehető értékek halmazát. Ekkor $\mid A \mid = \prod_{j=1}^{k} n_j = n$.
Tekintsünk egy N elemű független mintát a k-dimenziós eloszlásból, legyen ez
$Z = (Z_1, \ldots, Z_N)$. A két dimenziós esetet vizsgálva,adott két valószínűségi változónk, melyek felvesznek bizonyos értékeket,így az együttesen felvett értékek felsorolhatóak egy táblázatban. Az eloszlás megfigyelésekor tehát egyetlen megfigyelést tekintve ezen táblázat egy cellájában kapunk egy pontot. Ennek következtében a $Z_1, Z_2, \ldots, Z_N$ valószínűségi változók értékei a A celláiban N elemnek felelnek meg.

Mi lesz ekkor a maximum likelihood becslés?
Az ismeretlen paraméter most maga az eloszlás, nézzük a megfelelő cellák valószínűségeit,ahol minden cella annyiszor fordul elő,ahány megfigyelés esik bele. Jelölje $Z(a)$ az a cella gyakoriságát, amely megadja, hogy a szóban forgó cella $p(a)$ valószínűsége hányszor szerepel.
Ekkor a $Z_1, \ldots, Z_N$ valószínűségi változók együttes eloszlását a $\prod_{a \in A} p(a)^{Z(a)}$ képlet adja, ahol

- $\sum_{a \in A} Z(a) = N$
- $p(a) > 0 \quad \forall a \in A$, vagyis csak azon cellákat vesszük figyelembe, amelyek biztosan előfordulnak a megfigyelések között, tehát egy szigorított osztályon belül nézzük a maximum likelihood becslést.

A maximalizálandó mennyiség tehát a

$$\prod_{a\in A} p(a)^{Z(a)} \to \max_{\substack{p(a)>0, a\in A \\ \sum\limits_{a\in A} p(a)=1}},$$

melyről a loglikelihood függvényre áttérve az alábbi kifejezés adódik:

$$\sum_{a\in A} Z(a)\cdot logp(a) \to \max_{\substack{p(a)>0, a\in A \\ \sum\limits_{a\in A} p(a)=1}}$$

Feltételes szélsőértékfeladatként tekintve a problémára alkalmazhatjuk a Lagrange-féle multiplikátor elvet, ahol a nemnegativitás peremfeltétel elhagyásával egy bővebb tartományon keresünk megoldást:

$$\sum_{a\in A} Z(a)\cdot log\ p(a) + \lambda(1-\sum_{a\in A} p(a))$$

Deriválva $p(a)$ szerint

$$\frac{\partial}{\partial p(a)} = \frac{z(a)}{p(a)} - \lambda = 0 \Rightarrow \lambda\cdot p(a) = Z(a), \sum_{a\in A} p(a) = 1.$$

Összegezve és átrendezve:

$$\widehat{p(a)} = \frac{Z(a)}{N}$$

vagyis a megfelelő becslés az "a" cella relatív gyakorisága.
Vegyük észre, hogy ezzel az eredeti feladatot nem oldottunk meg teljesen, ugyanis ha valamely cellába nem esik megfigyelés, annak relatív gyakorisága nulla, így $\widehat{p(a)} = 0$, azaz a bővebb tartományon belüli maximum nem esik egybe a szűkebb tartomány maximumával.
Egy kontingencia táblát ritkának mondnunk, ha tartalmaz legalább egy üres cellát. A nulla értékű celláknak két csoportját különböztethetjük meg: mintavételből adódó (sampling) nullák és szerkezeti (structural) nullák. Előbbi esetben a minta elemszámának bővítésével van esély az esetleges nulla cellák kitöltésére, míg a második esetben nem áll módunkban megfigyelést tenni. Ha valamely cellára $z(a) = 0$ teljesül, akkor nem létezik maximum likelihood becslés,de ha minden $a \in A$ esetén $z(a) > 0$, akkor $\widehat{p(a)} = \frac{Z(a)}{N}$.

A feladat megfogalmazható információelméleti mennyiségek segítségével is.

Definíció 1.1.1
A q, p valószínűségi eloszlások ***divergenciája*** $D(p \parallel q) = \sum\limits_{a\in A} p(a)\ log\ \frac{p(a)}{q(a)}$.

Ha van olyan $a \in A$, melyre $p(a) = 0$ és $q(a) > 0$, akkor $D(p \parallel q) = \infty$.
*A q eloszlás **tartója** $T(q) = \{a \in A : q(a) > 0\}$.*

Ekkor

$$\sum_{a\in A}\left[\frac{Z(a)}{N}\right] log\ \frac{[\frac{Z(a)}{N}]}{p(a)} = \sum_{a\in A}\left[\frac{Z(a)}{N}\right] log\left[\frac{Z(a)}{N}\right] - \sum_{a\in A}\frac{Z(a)}{N} log p(a),$$

ahol a $\sum\limits_{a\in A}\frac{Z(a)}{N} log p(a)$ kifejezést szeretnénk maximalizálni (előjeltől eltekintve), azaz a tapasztalati eloszlás és az elméleti eloszlás divergenciáját kell minimalizálnunk. Összefoglalva tehát a maximum likelihood becslés keresése ekvivalens a

$$D(\widehat{P_N} \parallel P) \to \min_{\substack{p(a)>0, a\in A\\ \sum\limits_{a\in A} p(a)=1}}$$

feladat megoldásával.

Igaz-e, hogy $D(p \parallel q) \geq 0$, tetszőleges p és q véges halmazon értelmezett eloszlások esetén? A Jensen-egyenlőtlenséget kihasználva:

$$-D(q \parallel p) = \sum_{a\in A} q(a)\ log\ \frac{p(a)}{q(a)} \leq log \sum_{a\in A} q(a)\ \frac{p(a)}{q(a)} = log \sum_{a\in A} p(a) = log\ 1 = 0.$$

Tehát $-D(q \parallel p) \leq 0$, így $D(q \parallel p) \geq 0$.
A logaritmus függvény szigorúan konkáv tulajdonsága miatt $\dfrac{p(a)}{q(a)} = c\ \forall a \in A$. Mivel $\sum\limits_{a\in A} p(a) = 1$ és $\sum\limits_{a\in A} q(a) = 1$, így $c = 1$ esetén $p = q$.
Azt kaptuk tehát, hogy $D(\widehat{P_N} \parallel P)$ akkor minimális, ha $D(\widehat{P_N} \parallel P) = 0$, és az egyenlőség feltétele $\widehat{P_N} = P$, azaz ha a tapasztalati eloszlás benne van a megengedett mértékcsaládban, akkor ő lesz a maximum likelihood becslés. Fontos megjegyezni, hogy a divergencia nem szimmetrikus mennyiség, az elméleti eloszlás mindig a második helyen áll.

Tegyük fel , hogy $X_1, \dots X_k$ valószínűségi változók függetlenek, $a = (a_1, \dots, a_k)$ és $p(a) = p_1(a_1) \cdots p_k(a_k)$ vagyis a megfelelő együttes eloszlás szorzatmérték.
Mivel az együttes eloszlás csak a cellagyakoriságoktól függ, a sűrűségfüggvényt felírva

$$\prod_{a\in A} p(a)^{Z(a)} = \prod_{a\in A}\left[\prod_{j=1}^{k} p_j(a_j)\right]^{Z(a)} = \prod_{a\in A}\prod_{j=1}^{k} p_j(a_j)^{Z(a)} = \prod_{(a_1,\dots,a_k)\in A_1\times\dots\times A_k}\prod_{j=1}^{k} p_j(a_j)^{Z(a)} =$$

$$= \prod_{j=1}^{k} \prod_{a_j \in A_j} p_j(a_j)^{\sum\limits_{a'_j = a_j} Z(a')},$$

ahol $\frac{1}{N} \sum\limits_{a'_j = a_j} Z(a')$ a j-edik koordináta lefixálásával adódó mennyiség, a tapasztalati eloszlás marginálisa.
Jelölje $Z_j(a_j) = \sum\limits_{a'_j = a_j} Z(a')$ a marginális gyakoriságot. Az összesítést most olyan a_j-kre végezzük, ahol a j-edik koordináta rögzített. Ekkor a kétdimenziós kontingencia táblának a j-edik sorát kell vizsgálnunk, és itt számoljuk meg, hogy egy sávba hány mintaelem esik.
A j-edik marginális összes celláján végighaladva a likelihood függvény a következő alakot ölti:

$$\prod_{j=1}^{k} \prod_{a_j \in A_j} p_j(a_j)^{Z_j(a_j} \to \max_{\substack{p_j(a_j)>0, a_j \in A_j \\ j=1,\dots,k}}$$

Most tehát k darab tényező szorzatát kell maximalizálni, ahol a szorzatban szereplő összes tényező külön-külön maximalizálható a $\sum\limits_{a_j \in A_j} p_j(a_j) = 1$ peremfeltétel mellett:

$$\prod_{a_j \in A_j} p_j(a_j)^{Z_j(a_j)} \to \max_{\substack{p_j(a_j)>0 \\ \sum\limits_{a_j \in A_j} p_j(a_j)=1}}$$

Ebből azt kapjuk, hogy a maximum éppen a tapasztalati eloszlás megfelelő marginálisa, azaz

$$\widehat{p_j}(a_j) = \frac{Z_j(a_j)}{N},$$

ahol $\widehat{p} = \widehat{p_1} \otimes \ldots \otimes \widehat{p_k}$, vagyis a komponensek tapasztalati eloszlásainak szorzatmértéke adja a megfelelő maximum likelihood becslést.

Legyen most a dimenzió három, és tegyük fel, hogy X_1 és X_3 valószínűségi változók az X_2-re nézve feltételesen függetlenek, továbbá $p(a) > 0 \quad \forall a \in A = A_1 \times A_2 \times A_3$, azaz minden cella valószínűsége pozitív.
N darab független megfigyelést tekintve keresünk maximum likelihood becslést ebben a családban, a feltételes függetlenségből kiindulva:

$$P(X_1 = a_1, X_3 = a_3 \mid X_2 = a_2) = P(X_1 = a_1 \mid X_2 = a_2) \cdot P(X_3 = a_3 \mid X_2 = a_2)$$

$$P(X_1 = a_1, X_2 = a_2, X_3 = a_3) = P(X_2 = a_2) \cdot \frac{P(X_1 = a_1, X_2 = a_2)}{P(X_2 = a_2)} \cdot \frac{P(X_2 = a_2, X_3 = a_3)}{P(X_2 = a_2)} =$$

$$= \frac{P(X_1 = a_1, X_2 = a_2) \cdot P(X_2 = a_2, X_3 = a_3)}{P(X_2 = a_2)}.$$

Tehát minden valószínűség kevesebb változótól függ, mint az együttes eloszlás változóinak száma. Teljesül-e, hogy a feltételes függetlenség ekvivalens azzal, hogy az együttes eloszlás a fenti alakba írható?

Állítás 1.1
Legyen $P(a) = exp\{f_{12}(a_1, a_2) + f_{23}(a_2, a_3) + f_2(a_2)\}$ *alakú loglikelihood függvény. Ekkor a megfelelő valószínűségi változók feltételesen függetlenek.*

Bizonyítás

$$\frac{P(X_1 = a_1, X_2 = a_2) \cdot P(X_2 = a_2, X_3 = a_3)}{P(X_2 = a_2)} =$$

$$= \frac{\left(\sum\limits_{a_3 \in A_3} exp\{f_{12}(a_1, a_2) + f_{23}(a_2, a_3) + f_2(a_2)\}\right) \cdot \left(\sum\limits_{a_1 \in A_1} exp\{f_{12}(a_1, a_2) + f_{23}(a_2, a_3) + f_2(a_2)\}\right)}{\sum\limits_{a_1 \in A_1} \sum\limits_{a_3 \in A_3} exp\{f_{12}(a_1, a_2) + f_{23}(a_2, a_3) + f_2(a_2)\}} =$$

$$= exp\{f_{12}(a_1, a_2) + f_{23}(a_2, a_3) + f_2(a_2)\} = P(a)$$

.

$\square$

1.2. Maximum likelihood becslés loglineáris modellben

Adott $A = A_1 \times \ldots \times A_k$ véges szorzathalmaz esetén legyen $\Gamma \subset 2^{\{1,\ldots,k\}}$ részhalmazrendszer; $\gamma \in \Gamma$ és a_γ $(a_j, j \in \gamma)$ marginális cella. Tegyük fel, hogy $p(a) = exp\Big\{\sum\limits_{\gamma \in \Gamma} f_\gamma(a_\gamma)\Big\}$ alakú.

Definíció 1.2.1
$\mathcal{P}_\Gamma = \left\{p \text{ eloszlás } A\text{-n} \mid \exists\ f_\gamma, \gamma \in \Gamma \text{ függvények: } A_\gamma \to \mathbb{R} \text{ és } p(a) = exp\left(\sum\limits_{\gamma \in \Gamma} f_\gamma(a_\gamma)\right)\right\}$
mértékcsalád az úgynevezett loglineáris eloszláscsalád.

A maximum likelihood becslés ekvivalens az úgynevezett L-vetület megkeresésével, azaz a $D(P_N \parallel P)$ divergenciát kell minimalizálnunk a második változóban, egy adott logkonvex eloszláscsaládon ($P \in \mathcal{P}_\Gamma$).

A fentebb említett speciális esetek a loglineáris eloszláscsaládban:

- $k = 2$: $\Gamma = \Big\{\{1, 2\}\Big\}$ függetlenség

- $k=3$: $\Gamma = \Big\{\{1,2\},\{2,3\},\{2\}\Big\}$ feltételes függetlenség

A következő kérdést vizsgáljuk: loglineáris modell marginális eloszlásai loglineáris modellt alkotnak-e?

Legyen $M \subset \{1,\ldots,k\}$ rögzített részhalmaz, továbbá $P_{\Gamma,M} = \{P_M \mid P \in \mathcal{P}_\Gamma\}$ azon mértékek halmaza, melyek indexe M-beli.

Tetszőleges a_M-el jelölt marginális cella esetén

$$P_M(a_M) = \sum_{\substack{a'_M = a_M \\ a' \in A}} P(a') = \sum_{\substack{a'_M = a_M \\ a' \in A}} exp\left(\sum_{\gamma\in\Gamma} f_\gamma(a_\gamma)\right) = exp\left(\sum_{\substack{\gamma\in\Gamma \\ \gamma\subseteq M}} f_\gamma(a_\gamma)\right) \cdot \sum_{\substack{a'_M = a_M \\ a' \in A}} exp\left(\sum_{\substack{\gamma\in\Gamma \\ |\gamma\setminus M|\geq 1}} f_\gamma(a_\gamma)\right)$$

Lemma

Legyen $M \subset \{1,\ldots,k\}$ és tekintsük a loglineáris modell marginális eloszlásait:

$\{P_M \mid P \in \mathcal{P}_\Gamma\}$. Tegyük fel speciálisan, hogy van olyan $\gamma_0 \in \Gamma$, hogy minden $\gamma \in \Gamma$ esetén, melyre $\mid \gamma \setminus M \mid \geq 1$, teljesül, hogy $\gamma \cap M \subseteq \gamma_0 \cap M$, vagyis $\gamma \cap M$ lefedhető egyetlen részhalmazzal.Ekkor

$$P_M(a_M) = exp\left(\sum_{\substack{\gamma\in\Gamma \\ \gamma\subseteq M}} f_\gamma(a_\gamma)\right) \cdot exp\left(f_{\gamma_0\cap M}(a_{\gamma_0\cap M})\right),$$

vagyis $P_{\Gamma,M}$ is loglineáris modell a $\Gamma_M = \{\gamma \in \Gamma, \gamma \subseteq M, \gamma_0 \cap M\}$ halmazrendszeren.

□

Mi történik feltételes függetlenség esetén? Igaz-e, hogy a maximum likelihood becslés bizonyos speciális marginálisai egybeesnek a marginális loglineáris modellben adódó maximum likelihood becsléssel?

Láttuk, hogy a maximum likelihood becslés megkeresése ekvivalens egy divergencia minimalizálási feladattal. Ha adott egy halmazon értelmezett mérték, továbbá adott az alaphalmaz egy partíciója, akkor a partíciókra átörökítve egy mértéket, az egyes partíciók összemértékből egy származtatott mértékcsalád adódik.

A következő állítás megmutatja, hogy a teljes halmazon mért divergencia kapcsolatba hozható a származtatott divergenciával, méghozzá a partíciókon vett divergencia és a feltételes eloszlások divergenciáinak súlyozott összege szolgáltatja a megfelelő divergenciafelbontási összefüggést.

Állítás 1.2

Legyen S véges alaphalmaz (nem feltétlenül szorzathalmaz) és tegyük fel, hogy $S = \bigcup_{j \in J} C_j$ partícionált J indexhalmaz esetén.

Legyen P, Q eloszlások S-en, melyekre $\widetilde{P}(j) = \sum_{s \in C_j} P(s) = P(c_j)$ és $\widetilde{Q}(j) = Q(c_j)$ marginálisok. Ekkor:

$$D(Q \parallel P) = D(\widetilde{Q} \parallel \widetilde{P}) + \sum_{j \in J} \widetilde{Q}(j) \cdot D(Q_j \parallel P_j)$$

Bizonyítás

$$D(Q \parallel P) = \sum_{s \in S} Q(s) \cdot log \frac{Q(s)}{P(s)} = \sum_{j \in J} \left(\sum_{s \in C_j} Q(s) \cdot log \frac{Q(s)}{P(s)} \right) = \sum_{j \in J} \left(\sum_{s \in C_j} Q(s) \cdot log \frac{\frac{Q(s)}{\widetilde{Q}(j)}}{\frac{P(s)}{\widetilde{P}(j)}} + log \frac{\widetilde{Q}(j)}{\widetilde{P}(j)} \right) =$$

$$= \sum_{j \in J} \widetilde{Q}(j) \cdot log \frac{\widetilde{Q}(j)}{\widetilde{P}(j)} + \sum_{j \in J} \widetilde{Q}(j) \sum_{s \in C_j} \frac{Q(s)}{\widetilde{Q}(j)} \cdot log \frac{\frac{Q(s)}{\widetilde{Q}(j)}}{\frac{P(s)}{\widetilde{P}(j)}} = D(\widetilde{Q} \parallel \widetilde{P}) + \sum_{j \in J} \widetilde{Q}(j) \cdot D(Q_j \parallel P_j).$$

Adott N megfigyelés esetén létrehoztuk a tapasztalati eloszlást: $\widehat{P}(a) = \frac{Z(a)}{N}$, ahol $Z(a)$ az "a" cella gyakorisága.

□

Állítás 1.3

A P_Γ eloszláscsalád maximum likelihood becslés marginálisát jelölje $(P_{ML}^{\Gamma})_M$ és legyen $P_{ML}^{\Gamma_M}$ a marginális eloszlásokból elkészített loglineáris modell maximum likelihood becslése. Ha teljesülnek a fenti lemma feltételei, akkor

$$(P_{ML}^{\Gamma})_M = P_{ML}^{\Gamma_M},$$

tehát a maximum likelihood becslés bizonyos marginálisai egybeesnek a marginális loglineáris *modellben adódó maximum likelihood becsléssel, ha a marginális család is loglineáris modellt alkot.*

Bizonyítás

Legyen P^{Γ}_{ML} a maximum likelihood becslés és indirekt tegyük fel, hogy $(P^{\Gamma}_{ML})_M \neq P^{\Gamma_M}_{ML}$. Láttuk, hogy a maximum likelihood becslés meghatározása divergencia minimalizálással ekvivalens, így ha ezek a mennyiség nem egyenlőek, akkor van olyan mérték a marginálisok között a marginális cellarendszeren, amely "közelebb van" a tapasztalati eloszlás marginálisához.

Legyen tehát q_M mérték A_M-en, melyre $D(\widehat{P}_M \parallel q_M) < D(\widehat{P}_M \parallel (P^{\Gamma}_{ML})_M)$.

Azt kell megmutatnunk, hogy ez a marginális cellákon értelmezett eloszlás nem lehet maximum likelihood becslés a teljes cellarendszeren.

Legyen

$$P^{'}(a) = P^{\Gamma}_{ML}(a) \cdot \frac{q_M(a_M)}{(P^{\Gamma}_{ML})_M(a_M)}$$

eloszlás A-n. Így a feltételes eloszlásokon nem változtatunk, csak az 1.2 állításból adódó divergenciafelbontási összefüggésben szereplő első tag csökken.

Legyen továbbá

$$P^{'}_M(a_M) = \sum_{a^{'}_M = a_M} P^{'}(a^{'}) = \sum_{a^{'}_M = a_M} P^{\Gamma}_{ML}(a^{'}) \cdot \frac{q_M(a_M)}{(P^{\Gamma}_{ML})_M(a_M)} = q_M(a_M),$$

tehát az átskálázás hatására egy másik marginális eloszlás adódik.

A feltételes eloszlások így az egyes partíciókon belül nem változnak, ugyanis

$$\frac{P^{'}(a)}{P^{'}_M(a_M)} = \frac{P^{\Gamma}_{ML}(a)}{(P^{\Gamma}_{ML})_M(a_M)}.$$

A divergenciafelbontást felírva a teljes mértékre:

$$(1)\quad D(\widehat{P} \parallel P^{'}) = D(\widehat{P}_M \parallel q_M) + \sum_{a_M \in A_M} \widehat{P}_M(a_M) \cdot D(\widetilde{\widehat{P}}_{a_M} \parallel \widetilde{P^{'}}_{a_M})$$

$$(2)\quad D(\widehat{P} \parallel P^{\Gamma}_{ML}) = D(\widehat{P}_M \parallel (P^{\Gamma}_{ML})_M) + \sum_{a_M \in A_M} \widehat{P}_M(a_M) \cdot D(\widetilde{\widehat{P}}_{a_M} \parallel \widetilde{(P^{\Gamma}_{ML})}_{a_M})$$

Mivel a súlyok és a feltételes eloszlások megegyeznek, ezért a feltétel szerint $D(\widehat{P} \parallel P') < D(\widehat{P} \parallel P^{\Gamma}_{ML})$, így az általunk feltételezett maximum likelihood becslésnél P' jobb, ahol a P'-t definiáló reláció miatt P' benne van a megfelelő loglineáris modellben. Ezzel ellentmondásra jutottunk.

$\square$

2. fejezet

A loglineáris modell grafikus megközelítése

A grafikus modellek két nagy előnye, hogy a kontingencia táblákat dimenziójuknak megfelelően ábrázolhatjuk véges, irányítatlan gráfokon, továbbá ezekről a gráfokról a valószínűségi változók közti interakciók közvetlenül leolvashatók. Ebben a fejezetben a grafikus modellek egy speciális osztályával, a felbontható hipergráfokkal reprezentálható loglineáris grafikus modellekkel foglalkozunk.

2.1. Felbontható hipergráfok jellemzése

Definíció 2.1.1

Legyen $\Gamma \subset 2^{\{1,\ldots k\}}$ részhalmazrendszer. Ekkor a $\gamma \in \Gamma$: $\gamma \subseteq \{1, \ldots, k\}$ részhalmazok hipergráfot alkotnak, melynek csúcsai az 1-től k-ig terjedő számok, élei pedig γ részhalmazok.

Definíció 2.1.2

Azt mondjuk, hogy $H = (V_H, E_H)$ és $K = (V_K, E_K)$ hipergráfok felbontják a $G = (V_G, E_G)$ hipergráfot, ha

i) $V_H \cup V_K = V_G$, *azaz nem feltétlenül diszjunktak, de csúcshalmazaik egyesítése kiadja G csúcshalmazát.*

ii) $\exists h^* \in E_H$ *és* $k^* \in E_K$*:* $\forall \gamma \in E_K : \gamma \cap V_H \subseteq h^*$, $\forall \gamma \in E_H : \gamma \cap V_K \subseteq k^*$,

azaz amennyire H *belemetsz* K*-ba, azt* h^* *lefedi és fordítva.*

iii) $E_H \cap E_K = \emptyset$ *és* $E_H \cup E_K = \Gamma$, *vagyis az élhalmazok uniója kiadja a teljes élhalmazt.*

Definíció 2.1.3

Azt mondjuk, hogy $G = (V_G, E_G)$ *hipergráf felbontható, ha felbontásokkal visszavezethető egyetlen hiperélt tartalmazó hipergráfokra.*

A felbontásban szereplő H és K hipergráfok bizonyos tulajdonságai szükséges és elégséges feltételt biztosítanak G hipergráf felbonthatóságára.

Szükségünk lesz a következő definíciókra:

Definíció 2.1.4

Azt mondjuk, hogy $G_2 = (V_G, E)$ *kettő-metszet, ha* $\overline{ij} \Leftrightarrow \exists g \in E_G : \{i, j\} \subseteq g$, *vagyis* i *és* j *pontosan akkor van összekötve* G_2*-ben, ha van olyan* g *hiperél, melyre* $i, j \in g$.

Definíció 2.1.5

A hipergráf triangularizálható, ha minden legalább négy hosszú körbe behúzható él.

Definíció 2.1.6

A hipergráfot grafikusnak mondjuk, ha a hozzá tartozó G_2*-ben a maximális teljes részgráfok éppen a maximális hiperélek.*

Tétel 2.1.1

Egy $G = (V_G, E_G)$ *hipergráf pontosan akkor felbontható, ha grafikus és triangularizálható.*

Bizonyítás.

1. Lemma. Ha H és K felbontását adják G-nek, ahol H és K grafikus, akkor G is grafikus.

Bizonyítás:

Legyen $g \in E_G$ hiperél és tegyük fel, hogy $g \in E_H$. Ez megtehető, hiszen H a G felbontásában szereplő hipergráf. Mivel H a feltétel szerint grafikus, így a hozzá tartozó H_2-ben g maximális teljes részgráf. Azt kell megmutatnunk, hogy g a G_2-ben is maximális teljes részgráf. Indirekt tegyük fel, hogy ez nem teljesül, azaz van olyan extra csúcs G-ben, mellyel G minden további éle össze van kötve, formálisan: $\exists\, j \in V_G :\ \overline{ij} \in E_{G_2}\ \forall i \in g$. Három esetet különböztethetünk meg:

1. eset Ha minden ilyen "kritikus" él H_2-ben lenne, akkor g a H_2-ben nem lehetne maximális teljes részgráf, de H grafikus, így ez az eset nem fordulhat elő, mert ekkor g a H-ban bővíthető lenne.

2. eset Tegyük fel most, hogy az új élek K-hoz tartoznak, azaz $\forall i \in g : \overline{ij} \in E_K$. Mivel az (i,j) párok mindegyikét lefedi egy K-beli részhalmaz, vehetjük ezek unióját:
$\bigcup\limits_{i\in g}\{i,j\} \subseteq \bigcup\limits_{k\in E_K} k$.
A felbontás szerint ezt az uniót elmetszve H-beli hiperélek uniójával, a kapott metszet lefedhető $h^* \cap k^*$-gal.
Legyen $g \in E_H \subseteq \bigcup\limits_{h\in E_H} h$ H-beli hiperél. Ekkor $g \bigcap \bigcup\limits_{i\in g}\{i,j\} \subseteq h^* \cap k^*$, vagyis $g \subseteq h^* \cap k^*$, ugyanis j a g-n kívüli pont, így $\bigcup\limits_{i\in g}\{i,j\} = j \bigcup g$, vagyis a metszet éppen g. Ekkor azonban $g \subseteq k^*$ is teljesül, tehát g benne van K-ban is. Ez viszont nem lehetséges, ugyanis a feltétel szerint K grafikus, és a felbontás miatt a hiperéleket szétosztottuk H és K között, másrészt maximális részeket választottunk, így nem lehet az egyik része a másiknak.

3. eset

$$\left(g \bigcup_{i\in g} \bigcup_{(i,j)\in E_{H_2}} \{i,j\}\right) \bigcap \left(\bigcup_{(i,j)\in E_{K_2}} \{i,j\}\right) \subseteq h^* \bigcap k^*,$$

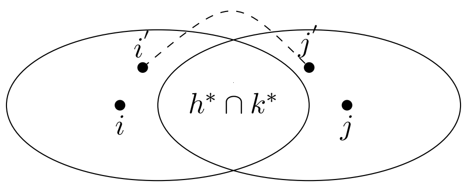
ahol a metszet első tagja lefedhető H-beli, míg második tagja lefedhető K-beli hiperélekkel. A $h^* \bigcap k^*$-ban j benne van, mert egyik sem üres, továbbá $\forall i \in g : \{i,j\} \subseteq h^*$ és $\overline{ij} \in E_{K_2}$, így $\overline{ij} \in E_{H_2}$, ezzel azonban az 1. esethez jutunk, melyről láttuk, hogy nem lehetséges.

$\square$

2. Lemma. Ha H és K felbontása G-nek, ahol H és K triangularizálható, akkor G is triangularizálható.

Bizonyítás:

Ha olyan kört vennénk, melynek minden pontja H-beli, akkor ott az élek is H-ból származnak, mert egy K-ból jövő él $h^* \cap k^*$-ban is benne lenne. Ugyanez elmondható a csupa K-beli élet tartalmazó körről is. Tekintsünk tehát egy olyan kört a gráfban, melyre $i \in V_H \setminus (h^* \cap k^*)$ és $j \in V_K \setminus (h^* \cap k^*)$.



Ha i' jelöli az utolsó pontot, amely még nincs benne K-ban és j' az utolsó olyat, amely még nem eleme H-nak, akkor ezen két pont $h^* \cup k^*$ kikerülésével nem köthető össze, ugyanis ha $i'j'$ valóban él, akkor az lefedhető egy H-beli vagy K-beli hiperéllel.

□

Azt kell még megmutatunk, hogy a grafikus és triangularizálható tulajdonságokból következik a felbonthatóság. Ehhez konstruálnunk kell egy felbontást: diszjunkt módon két részre osztva a hiperéleket, a csúcsok is két részre osztódnak (nem feltétlenül diszjunkt módon), továbbá amennyire az egyik belemetsz a másikba, az mindkét halmazból hiperélekkel lefedhető.
Legyen tehát $G = (V_G, E_G)$ hipergráf és $h^* \cap k^* = \left(\bigcup_{h \in E_H} h\right) \bigcap \left(\bigcup_{k \in E_K} k\right)$. Célunk a feltételeknek megfelelő h^* és k^* megtalálása.
Induljunk ki tetszőleges h és k hiperélekből, majd tekintsük a $h \cap k$ kihagyásával keletkező hipergráf összefüggő komponenseit:$h \setminus (h \cap k)$ kerüljön H-ba, $k \setminus (h \cap k)$ kerüljön K-ba és az egyikhez sem kapcsolódó összefüggő komponensek pedig szintén K-ba. Ezzel a módszerrel elkészítettünk egy olyan felbontást, melyben csak $h \cap k$ közös.
Felmerül a kérdés, hogy bármely két hiperélből kiindulva elkészíthető-e ilyen felbontás.
Előfordulhat-e olyan szituáció, melyben $h \setminus (h \cap k)$ és $k \setminus (h \cap k)$ ugyanazok az összefüggő komponensek?

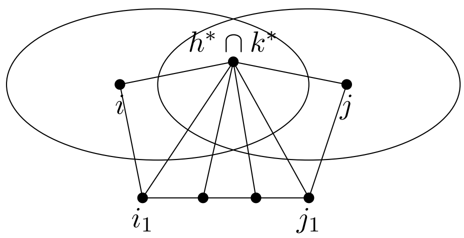
Tegyük fel, hogy ez a helyzet. Ekkor $i \in h \setminus (h \cap k)$ és $j \in k \setminus (h \cap k)$ összeköthető $(h \cap k)$ kihagyásával, kettőmetszet-beli hiperélek sorozatával. Tekintsük az ilyen összekötő utak közül a legrövidebbet, ahol a minimalizálás pontpárra és úthosszra értendő.

Tegyük fel, hogy $j_1 \notin k \setminus (h \cap k)$ a j utáni első pont. ($j_1 \in k \setminus (h \cap k)$ esetén található lenne rövidebb út).

1. eset: $j_1 = i$

   Vegyük észre, hogy i és j közvetlenül össze vannak kötve egymással és a metszet minden elemével is, ezért i, j, valamint a metszet együtt teljes részgráfot alkotnak. Vegyük G_2-n belül az ezt tartalmazó maximális teljes részgráfot, amely G grafikussága miatt hiperél lesz. Jelölje ezt az i-t, j-t és a metszetet is tartalmazó hiperélt g. Ekkor $h \cap g \supseteq h \cap k \cup \{j\}$, azaz ilyenkor van olyan hiperéle G-nek, amely jobban belemetsz h-ba, mint k-ba, ami ellentmond a feltételnek.

2. eset: $j_1 \neq i$ és $i_1 \neq j$.



   Mivel i és j között a legrövidebb utat választottuk, a további élek a metszetben lévő pontból indulnak, melyből a hipergráf triangularizálható. Az i_1 pont a metszet bármely pontjával összeköttetésben áll, és ez i-re is igaz (h lefedi), így i_1, i és $h \cap k$ tejes részgráfot alkotnak. Jelölje a szóban forgó teljes részgráfot tartalmazó maximális teljes részgráfot g.Ekkor $(h \cap k) \cup \{i\} \subseteq g \cap h$, azaz van olyan hiperél, amely jobban belemetsz h-ba, mint k. Feltéve tehát, hogy $h \setminus (h \cap k)$ és $k \setminus (h \cap k)$ ugyanazok az összefüggő komponensek, van olyan g hiperél, melyre $|g \cap h| > |h \cap k|$.

   Összefoglalva tehát tetszőleges $h \in E_H$ hiperélhez azt a k hiperélt választva, amely legjobban belemetsz h-ba, azaz melyre $|h \cap k| \to max$, tetszőleges hiperélből kiindulva konstruálható felbontás.

2.2. Adott felbontás esetén történő maximum likelihood becslés előállítása

G hipergráf felbontását adó H és K segítségével a marginálisokra vonatkozó maximum likelihood becslésből szeretnénk létrehozni az együttes eloszlásra vonatkozó maximum likelihood becslést visszavezetés segítségével.

Készítsünk el egy megfeleltetést:
jelölje $\mathcal{P}_G$ a G hipergráfhoz tartozó, $\mathcal{P}_\Gamma$-nak megfelelő loglineáris modellt; valamint legyenek $\mathcal{P}_H$ és $\mathcal{P}_K$ a H és K felbontásban szereplő hipergráfokban lévő részhalmazrendszerekhez tartozó loglineáris modellek, melyek $\mathcal{P}_{\Gamma,M}$, $M \subset \{1,\ldots,k\}$ marginális loglineáris modell megfelelői.
Láttuk, hogy bizonyos feltételek mellett a loglineáris modell marginálisa is loglineáris modell, amely feltételek a felbonthatóság definíciójából adódóan a megfeleltetés után is teljesülnek.
Jelölje $\widehat{P_G}$ a megfelelő maximum likelihood becslést és tekintsük ennek marginálisait;
$\left(\widehat{P_G}\right)_{V_H} = \widehat{P_H}$ és $\left(\widehat{P_G}\right)_{V_K} = \widehat{P_K}$ melyek szintén loglineáris modellt alkotnak.
Mi lesz az együttes eloszlás?
A közös részt h^* és k^* lefedi, továbbá bizonyos valószínűségi változók indexei V_H-ban és V_K-ban is szerepelnek, a többi pedig vagy az egyikben, vagy a másikban:

$$p(a) = exp\left(\sum_{\gamma \in max\Gamma} f_\gamma(a_\gamma)\right) = exp\left(\sum_{\gamma \in E_H} f_\gamma(a_\gamma)\right) \cdot exp\left(\sum_{\gamma \in E_K} f_\gamma(a_\gamma)\right)$$

Szétbontva aszerint, hogy a hiperélek melyik élhalmazhoz tartoznak, az együttes eloszlás szorzat alakba írható. A függetlenség automatikusan ugyan nem teljesül az esetleges közös valószínűségi változók miatt, de a közös részt lerögzítve feltételes függetlenségre következtethetünk.
Tegyük fel, hogy $X_j, j \in V_H$ és $X_l, l \in V_K$ valószínűségi változók feltételesen függetlenek $X_i, i \in h^* \cap k^*$ közös részre nézve. Ekkor az együttes eloszlás

$$P(X_{V_H} = a_{V_H}, X_{V_K} = a_{V_K} \mid X_{h^* \cap k^*} = a_{h^* \cap k^*}) =$$

$$= P(X_{V_H} = a_{V_H} \mid X_{h^* \cap k^*} = a_{h^* \cap k^*}) \cdot P(X_{V_K} = a_{V_K} \mid X_{h^* \cap k^*} = a_{h^* \cap k^*})$$

Ebből

$$P(X = a) = \frac{P(X_{V_H} = a_{V_H}) \cdot P(X_{V_K} = a_{V_K})}{P(X_{h^* \cap k^*} = a_{h^* \cap k^*})},$$

így

$$p(a) = \frac{p_{V_H}(a_{V_H}) \cdot p_{V_K}(a_{V_K})}{p_{h^* \cap k^*}(a_{h^* \cap k^*})}.$$

Tehát a maximum likelihood becslés:

$$\widehat{P_G} = \frac{\widehat{P_H}(a_H) \cdot \widehat{P_K}(a_K)}{\widehat{P_{h^* \cap k^*}}(a_{h^* \cap k^*})}.$$

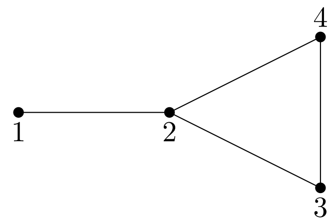
2.3. Explicit maximum likelihood becslés

Felbontható hipergráfokkal definiált loglineáris modellben a maximum likelihood becslés explicit formulájának megtalálásához definiálnunk kell az úgynevezett partíciófüggvényt vagy más néven indexet.

Ehhez legyen G összefüggő és felbontható hipergráf, továbbá $D \subset V_G$ olyan, melyhez teljes részgráf tartozik. Ekkor létezik $g \in E_G$ hiperél, melyre $D \subset g$.

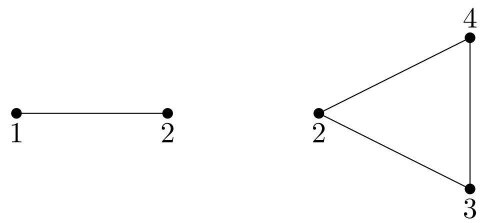
Első lépésként hagyjuk el a gráfból D-t, vagyis a D-beli csúcsokat és minden olyan élt, melynek legalább egy pontja D-ben van. Jelölje a maradék gráf összefüggő komponenseit G_i^D, majd ezen összefüggő komponensekhez adjuk vissza D-t a megfelelő élek behúzásával. Legyen ez a gráf G_D^i, melyben $D \subset G_D^i$ és hozzá teljes részgráf tartozik.

A fenti folyamatot illusztráljuk egy egyszerű példával:



Ha most $D = \{2\}$ akkor D elhagyása után az összefüggő komponensek

D visszaadása után pedig



A továbbiakban minden $D \subset V_G$ teljes részgráf esetén legyen $\mu_G(D)$ azon indexek számossága, melyekre D nem maximális teljes részgráf G^i_D-ben és $\nu_G(D) = 1 - \mu_G(D)$ a partíciófüggvény. Ha D teljes részgráf volt az eredeti gráfban, akkor ez a tulajdonság az összefüggő komponensekben is érvényesül, de a maximalitás nem feltétlenül teljesül. A fenti példában $\nu_G(\{2\}) = -1$ és $\{2\}$ nem maximális.

Definíció 2.3.1

Azt mondjuk, hogy $D \subset V_G$ a G hipergráf teljes elvágó részgráfja, ha D teljes részgráf és $G \setminus D$ már nem összefüggő.

Lemma.

$\nu_G(D) = 1$, ha D maximális teljes részgráf

$\nu_G(D) = 0$, ha D nem maximális teljes részgráf és nem elvágó részgráf

$\nu_G(D) < 0$, ha D elvágó részgráf.

Bizonyítás. Ha D maximális teljes részgráf volt az eredeti gráfban, akkor az összefüggő komponensekhez visszaadva is megőrzi ezt a tulajdonságát, így $\nu_G(D) = 1 - \mu_G(D) = 1 - 0 = 1$.
Ha D nem maximális és nem is elvágó halmaz, azaz egyetlen komponensünk van, akkor visszatéve sem lesz maximális, így $\nu_G(D) = 1 - \mu_G(D) = 1 - 1 = 0$.
Ha D elvágó halmaz, akkor biztosan van legalább kettő összefüggő komponens, így $\mu_G(D) \geq 2$, vagyis $\nu_G(D) \leq 1 - 2 < 0$.

□

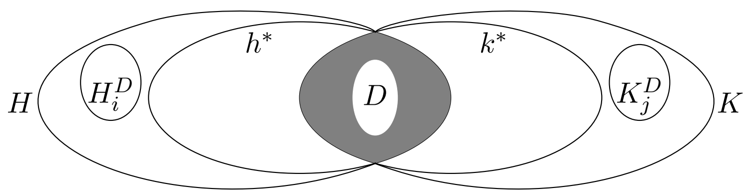
Állítás 2.3.1

i) Ha $D \neq h^* \cap k^*$, *akkor* $\nu_G(D) = \nu_H(D) + \nu_K(D)$.

ii) Ha $D = h^* \cap k^*$, *akkor* $\nu_G(D) = \nu_H(D) + \nu_K(D) - 1$.

Bizonyítás.

1. eset: $D \subsetneq h^* \cap k^*$

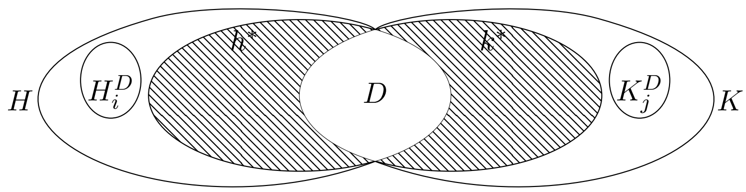


D-t elhagyva G-ből, H-ból és K-ból, majd vizsgálva a megmaradó összefüggő komponenseket, vegyük észre, hogy H_i^D-ben és K_j^D-ben a $h^* \cap k^* \setminus D$-t tartalmazó összefüggő komponens G-ben összeolvad. Így D-t visszarakva D pontosan akkor lesz maximális teljes részgráf H-ban, amikor G-ben. Továbbá a $h^* \cap k^* \setminus D$-t tartalmazó komponens D visszarakása után nem lehet maximális sem H-ban, sem K-ban, így ez a komponens $\mu_H(D)$-hez és $\mu_K(D)$-hez is hozzájárul egy elemmel, melyek G-ben összeolvadnak, azaz

$$\mu_G(D) = \mu_H(D) + \mu_K(D) - 1$$

$$\nu_G(D) = 1 - \mu_H(D) - \mu_K(D) + 1 = \nu_H(D) + \nu_K(D)$$

2. eset: $D = h^* \cap k^*$

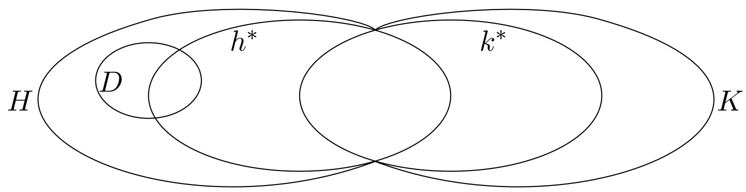


Ebben az esetben G-ben a $h^* \setminus h^* \cap k^*$ és $k^* \setminus h^* \cap k^*$ összefüggő komponenseket $h^* \cap k^*$ elkerülése nélkül nem lehet összekötni, így ezek a komponensek most nem olvadhatnak össze G-ben, azaz D visszarakásával ennek $\mu_H(D)$-ben és $\mu_K(D)$-ben is lesz hozadéka.Tehát:

$$\mu_G(D) = \mu_H(D) + \mu_K(D)$$

$$\nu_G(D) = 1 - \mu_H(D) - \mu_K(D) = \nu_H(D) + \nu_K(D) - 1.$$

3. eset: $D \subset E_H$, $D \nsubseteq E_K$



Ekkor K-beli út behúzásával nem keletkezik új teljes részgráf sem H-ban, sem G-ben, vagyis D elhagyásával most a H-beli és G-beli viszonyok megegyeznek, azaz

$$\mu_G(D) = \mu_H(D).$$

$\square$

Állítás 2.3.2

Legyen G hipergráf és jelölje $\mathcal{D}$ a maximális teljes részgráfok halmazát. Ekkor

$$\text{ha } G \text{ felbontható} \Rightarrow \sum_{D \in \mathcal{D}} \nu_G(D) = 1.$$

Bizonyítás. Láttuk, hogy tetszőleges hiperélből kiindulva konsturálható felbontás mindig a legjobban belemetsző hiperélt választva. Tegyük fel, hogy H és K a G hipergráf felbontását adják. Ekkor a 2.3.1. állítás alapján

$$\sum_{D \in \mathcal{D}} \nu_G(D) = \sum_{D \in \mathcal{D}} \nu_H(D) + \nu_K(D) + \nu_H(D) + \nu_K(D) - 1 = 1 + 1 - 1 = 1.$$

$\square$

Az eddigi konstrukció során olyan $D \subseteq V_G$ részgráfokat vizsgáltunk, melyek teljesek voltak G-ben.

Legyen most $D^* \subseteq V_G$ tetszőleges részgráfja a G hipergráfnak. Nevezzünk egy G hipergráfot szétbonthatónak egy tetszőleges $D^* \subseteq V_G$ részhalmazra nézve, ha $E_G = \cup\{G^i_{D^*} : i \in I\}$ és $\forall i \neq j$ esetén $G^i_{D^*} \cap G^j_{D^*} = D^*$ [3]. Speciálisan, ha D^* egy teljes elvágó részgráf és $G^i_{D^*}$ jelöli azon összefüggő kompenenseit G-nek, melyeket úgy kapunk, hogy D^*-ot elhagyjuk a hipergráfból, majd a maradék összefüggő kompensekhez visszaadjuk, akkor $G^i_{D^*}$ részgráfok szétbontják a G hipergráfot.

Ekkor minden $G^i_{D^*}$ részgráfra, melyben $D^* \subseteq G^i_{D^*}$ tartozik egy $\nu_i(D^*)$-gal jelölt partíciófüggvény. Megállapodás szerint $\nu_i(D^*) = 0$, ha $D^* \nsubseteq G^i_{D^*}$.

A 2.3.1 állítás általánosabb alakja:

Állítás 2.3.3

Legyen G hipergráf felbontható $D^* \subseteq V_G$ teljes elvágó részgráfra nézve és $D \subseteq V_G$ teljes részgráf. Ekkor

$$\nu_G(D) = \begin{cases} \sum\limits_{i \in I} \nu_i(D) & D \neq D^*, \\ \sum\limits_{i \in I} \nu_i(D^*) - |I| + 1 & D = D^*. \end{cases}$$

Állítás 2.3.4

Legyen G összefüggő hipergráf. Jelölje $\mathcal{D}$ a G-ben található maximális teljes részgráfok halmazát. Ekkor

$$\sum_{\mathcal{D}} \nu_G(D) \geq 1$$

Bizonyítás. A bizonyítás $\mathcal{D}$ halmaz elemszáma szerinti indukcióval történik.

Egyetlen ilyen halmaz esetén az állítás a partíciófüggvény tulajdonságai miatt azonnal következik, így tegyük fel, hogy $|\mathcal{D}| > 1$, továbbá az indukciós feltevésünk szerint az állítás teljesül azon összefüggő hipergráfokra, melyek kevesebb, mint $|\mathcal{D}|$ maximális teljes részgráffal rendelkeznek.

Ekkor vagy $\forall D \in \mathcal{D}$ esetén $\nu(D) \geq 0$, így a lemma szerint az állítás teljesül, vagy $\exists D^*$, melyre $\nu(D^*) < 0$. Ismét a fenti lemmát használva ez csak úgy fordulhat elő, ha D^* elvágó részgráf.

Hagyjuk el D^*-ot a gráfból, majd a megmaradt összefüggő komponensekhez adjuk vissza. Láttuk, hogy az így keletkező részgráfok szétbontják a kiindulási gráfot, így $|G_i(D)| < |V(G)|$ esetén az indukciós feltevés és a 2.3.3 állítás felhasználásával adódik, hogy

$$\sum \nu_G(D) = \sum_{D \neq D^*} \left(\sum_{i \in I} \nu_i(D) \right) + \sum_{i \in I} \nu_i(D^*) - |I| + 1 = \sum_{i \in I} \left(\sum_D \nu_i(D) \right) - |I| + 1 \geq |I| - |I| + 1 = 1.$$

□

A vizsgált partíciófüggvény jelentősége abban rejlik, hogy segítségével explicit formula adható a felbontható hipergráfokkal definiált grafikus loglineáris modell maximum likelihood becslésére. A 2.2 bekezdésben az együttes eloszlásra kapott maximum likelihood becslésre vonatkozó képletet és a 2.3.1. állítást összevetve kapjuk a következő eredményt:

$$\widehat{P_G}(a) = \prod_{\mathcal{D}} \widehat{P}^{\nu_G(D)}(a_D),$$

ahol $\mathcal{D}$ a maximális teljes részgráfok halmaza. Tehát a megfelelő marginálisokon vett becslések és a partíciófüggvény ismeretében $\widehat{P_G}$ kiszámolható, továbbá ha a maximum likelihood becslés már ismert, akkor ebből következtethetünk ν_G index további tulajdonságaira.

Vegyük azt a speciális esetet, amikor a mintaelemszám N és minden cellába egyetlen megfigyelés esik, azaz $Z(A) = 1$. Felhasználva a maximum likelihood becslésre kapott formulát:

$$\widehat{P_G}(a) = \frac{1}{N} \prod_D p(a_D)^{\nu_G(D)} = \frac{1}{N} \prod_D \left(\prod_{\gamma \notin D} p(a_\gamma) \right)^{\nu_G(D)} =$$

$$= \frac{1}{N} \prod_{\gamma \in \Gamma} \prod_{\gamma \notin D} p(a_\gamma)^{\nu_G(D)} = \frac{1}{N} \left(\prod_{\gamma \in \Gamma} p(a_\gamma) \right)^{\sum\limits_{D \subseteq \Gamma \setminus \{\gamma\}} \nu_G(D)}.$$

A fenti összefüggés csak akkor lehet igaz, ha minden (a modellhez tartozó) összefüggő, felbontható hipergráf esetén

$$\sum_{D \subseteq \Gamma \setminus \{\gamma\}} \nu_G(D) = 0 \quad \forall \gamma \in \Gamma.$$

A felbonthatóság miatt

$$\sum_{D} \nu_G(D) = \sum_{\gamma \notin D} \nu_G(D) + \sum_{\gamma \in D} \nu_G(D),$$

így

$$\sum_{\gamma \in D} \nu_G(D) = 1$$

minden a modellt leíró $G = (V_G, \Gamma)$ összefüggő, felbontható hipergráfra.

Összegezve a fenti eredményt $\gamma \in \Gamma$ esetén

$$\mid \Gamma \mid = \sum_{\gamma \in \Gamma} \sum_{\gamma \in D} = \sum_{D} \mid D \mid \nu(D)$$

egyenlőség is teljesül.

3. fejezet

A hierarchikus loglineáris modell

A megfelelő loglineáris modellhez tartozó hipergráf definiálásakor a csúcshalmazt az egytől k-ig terjedő számok halmazának választottuk, az élhalmazt pedig Γ halmaz bizonyos részhalmazai határozták meg. Γ összes elemének beválasztása helyett érdemesebb a maximális részhalmazokból álló halmazrendszert tekinteni élhalmaznak, ugyanis az egymást tartalmazó halmazok közül a kisebb redundáns és ezzel a módszerrel is ugyanaz a loglineáris modell kerül meghatározásra. Láttuk, hogy minden $\gamma \in \Gamma$ részhalmazhoz tartozik egy alkalmas függvény, melyre

$$p(a) = exp\left\{\sum_{\gamma\in\Gamma} f_\gamma(a_\gamma)\right\},$$

azonban a paraméterezés nem lesz egyértelmű, ha Γ halmazra semmilyen kikötést nem teszünk. Hogy lehet a függvénykijelölést egyértelműsíteni?
Az eredeti Γ halmaz helyett tekintsük

$$\overline{\Gamma} = \{\gamma \subseteq \{1, \ldots, k\}| \ \ \exists\overline{\gamma} \in \Gamma : \gamma \subseteq \overline{\gamma}\}$$

halmazt, azaz válasszuk be a meglévő részhalmazok összes részhalmazát. Ezzel látszólag növeljük a modell nem-egyértelműségét, ugyanis több függvényre van szükség, azonban megköveteljük a következő feltétel teljesülését

$$\sum_{a_j \in A_j} \overline{f_{\overline{\gamma}}}(a_{\overline{\gamma}\setminus\{j\}}, a_j) = 0$$

$\forall\, \emptyset \neq \overline{\gamma} \in \overline{\Gamma}$, $\forall j \in \gamma$ és $\forall a_{\overline{\gamma}\setminus\{j\}} \in A_{\gamma\setminus\{j\}}$ esetén.

Ez azt jelenti, hogy $\overline{f}_{\overline{\gamma}}$ marginálisai nullák,így nem befolyásolják a $\overline{\Gamma}$ részhalmazaihoz tartozó függvények által felvett értéket. Az így kapott modell az úgynevezett hierarchikus loglineáris modell.

Vizsgáljuk meg azt a kérdést, hogy lehet ugyanazt az eloszlást előállítani hierarchikus loglineáris modell segítségével?

Célunk annak megmutatása, hogy $\forall \overline{\gamma} \in \overline{\Gamma}$ esetén léteznek olyan $\overline{f_{\overline{\gamma}}}$ függvények, melyekre

$$\sum_{\overline{\gamma} \in \overline{\Gamma}} \overline{f_{\overline{\gamma}}}(a_{\overline{\gamma}}) = \sum_{\gamma \in \Gamma} f_\gamma(a_\gamma),$$

azon feltétel mellett, hogy bármely koordináta szerint összegezve az összeg 0.

Ehhez rögzített $\gamma \in \Gamma$ részhalmaz esetén alkalmasan legyártjuk az $\overline{f_{\overline{\gamma}}}^{\gamma}$, $\overline{\gamma} \subseteq \gamma$ függvényeket rekurzív módon:

1. $\overline{f}_{\emptyset}^{\gamma} = \dfrac{1}{\mid A_\gamma \mid} \sum_{a_\gamma \in A_\gamma} f_\gamma(a_\gamma)$
2. $\overline{f}_{\overline{\gamma}}^{\gamma}(a_{\overline{\gamma}}) = \dfrac{1}{\mid A_\gamma \setminus \overline{\gamma} \mid} \sum_{a_{\overline{\gamma}} = a'_{\overline{\gamma}}} f_\gamma(a'_\gamma) - \sum_{\gamma' \subset \overline{\gamma}} \overline{f}_{\gamma'}^{\gamma}(a_{\gamma'})$

Átrendezés után

$$\frac{1}{\mid A_\gamma \setminus \overline{\gamma} \mid} \sum_{a_{\overline{\gamma}} = a'_{\overline{\gamma}}} f_\gamma(a'_\gamma) = \overline{f}_{\overline{\gamma}}^{\gamma}(a_{\overline{\gamma}}) + \sum_{\gamma' \subset \overline{\gamma}} \overline{f}_{\gamma'}^{\gamma}(a_{\gamma'})$$

adódik, vagyis $\gamma = \overline{\gamma}$ esetén

$$\overline{f}_{\gamma}^{\gamma}(a_\gamma) = f_\gamma(a_\gamma) - \sum_{\gamma' \subset \overline{\gamma}} \overline{f}_{\gamma'}^{\gamma}(a_{\gamma'}) \quad \text{és} \quad \overline{f}_{\overline{\gamma}} = \sum_{\overline{\gamma} \subseteq \gamma} \overline{f}_{\overline{\gamma}}^{\gamma}\ .$$

3.1. Maximum likelihood becslés hierarchikus modellben

Mi lesz a maximum likelihood becslés hierarchikus modellben?

A megfelelő loglikelihood függvény a következő alakba írható:

$$\sum_{a\in A} Z(a) log\ p(a) = \sum_{a\in A} Z(a) \sum_{\gamma\in\Gamma} f_\gamma(a_\gamma)$$

Hierarchikus modellt feltételezve a maximalizálandó mennyiség:

$$\sum_{\overline{\gamma}\in\Gamma} Z(a_{\overline{\gamma}})\overline{f}_{\overline{\gamma}}(a_{\overline{\gamma}}) \to \max_{\substack{\sum_{a\in A} exp\left(\sum_{\overline{\gamma}\in\Gamma}\overline{f}_{\overline{\gamma}}(a_{\overline{\gamma}})\right)=1 \\ \sum_{\substack{a \\ a_{\gamma'}=a'_{\gamma'}}} f_{\overline{\gamma}}(a)=0}},$$

minden rögzített $a'_{\gamma'}$ értékre, ahol az egyszerűség kedvéért jelölje $\gamma' := \overline{\gamma}\setminus\{i\}$, $i\in\overline{\gamma}$, $\overline{\gamma}\in\overline{\Gamma}$.

A Lagrange-féle multiplikátor elvet felírva rögzített γ' és $a_{\gamma'}$ esetén

$$\frac{\partial}{\partial f_{\gamma'}(a_{\gamma'})}\left[\sum_{a\in A} Z(a_{\overline{\gamma}}) \sum_{\overline{\gamma}\in\overline{\Gamma}} \overline{f_{\overline{\gamma}}}(a) - \lambda\cdot\left(\sum_{a\in A} exp\left[\sum_{\overline{\gamma}\in\overline{\Gamma}}\overline{f_{\overline{\gamma}}}(a_{\overline{\gamma}})\right]-1\right) - \sum_{\substack{a_{\gamma'}\in A_{\gamma'} \\ \gamma'\subsetneq\overline{\gamma}\in\overline{\Gamma} \\ |\overline{\gamma}\setminus\gamma'|=1}} \lambda_{\overline{\gamma},\gamma'}(a_{\gamma'}) \sum_{a_i\in A_i}\overline{f_{\overline{\gamma}}}(a_{\gamma'},a_i)\right] = 0.$$

Ebből:

$$Z(a_{\gamma'}) - \lambda\ p(a'_\gamma) - \sum_{\substack{\gamma'\subsetneq\overline{\gamma} \\ |\overline{\gamma}\setminus\gamma'|=1}} \lambda_{\overline{\gamma},\gamma'}(a_{\gamma'}) = 0$$

Állítás 3.1.1

$$\widehat{P}_{\overline{\gamma}}(a_{\overline{\gamma}}) = \frac{Z(a_{\overline{\gamma}})}{\mid A_{\overline{\gamma}}\mid}$$

Bizonyítás.

1. eset: $\gamma' = \emptyset \Rightarrow N - \lambda = 0$, ahol $N = \mid A_{\overline{\gamma}}\mid$

2. eset: $|\gamma'| = 1 \Rightarrow Z(a_{\gamma'}) - N\ p(a_{\gamma'}) - \lambda_{\gamma'} = 0$, így $\lambda_{\gamma'} = 0$.

Tegyük fel, hogy rögzített $\gamma' \in \overline{\Gamma}$ esetén az állítás igaz.

Indukcióval megmutatjuk, hogy ekkor $\widetilde{\gamma} \subsetneq \gamma'$, $|\gamma' \setminus \widetilde{\gamma}| = 1$ részhalmazok mellett is fennáll. Legyen $\widetilde{a}_{\widetilde{\gamma}}$ rögzített. Ekkor

$$Z(\widetilde{a}_{\widetilde{\gamma}}) - N\ p(\widetilde{a}_{\widetilde{\gamma}}) - |A_{\gamma'\setminus\widetilde{\gamma}}|\ \lambda_{\gamma',\widetilde{\gamma}}(\widetilde{a}_{\widetilde{\gamma}}) - \sum_{\substack{\widehat{\gamma}\subset\widetilde{\gamma} \\ |\widetilde{\gamma}\setminus\widehat{\gamma}|=1 \\ a_{\gamma'\setminus\widetilde{\gamma}}}} \lambda_{\widehat{\gamma}\cup(\gamma'\setminus\widetilde{\gamma}),\gamma'}(\widetilde{a}_{\widehat{\gamma}} \cap a'_{\gamma'\setminus\widetilde{\gamma}}) = 0.$$

Ebből

$$\lambda_{\gamma',\widetilde{\gamma}}(\widetilde{a}_{\widetilde{\gamma}}) = -\frac{1}{|A_{\gamma'\setminus\widetilde{\gamma}}|} \sum_{\substack{\widehat{\gamma}\subset\widetilde{\gamma} \\ |\widetilde{\gamma}\setminus\widehat{\gamma}|=1 \\ a_{\gamma'\setminus\widetilde{\gamma}}}} \lambda_{\widehat{\gamma}\cup(\gamma'\setminus\widetilde{\gamma}),\gamma'}(\widetilde{a}_{\widehat{\gamma}} \cap a'_{\gamma'\setminus\widetilde{\gamma}}).$$

Feltéve, hogy $\gamma' = \{i, j\}$

$$\lambda_{\gamma',i}(\widetilde{a}_i) = -\frac{1}{|A_j|} \sum_{a'_j} \lambda_{j,\gamma'}(a'_j)$$

$$\lambda_{\gamma',i}(\widetilde{a}_i) = \lambda_{i,\gamma'} \quad \text{és} \quad \lambda_{j,\gamma'}(a'_j) = \lambda_{j,\gamma'}$$

$$-|A_j|\ \lambda_{i,\gamma'} - |A_j|\ \lambda_{j,\gamma'} = 0.$$

Tehát valóban

$$\widehat{P}_{\gamma'}(a_{\gamma'}) = \frac{Z(a_{\gamma'})}{|A_{\gamma'}|}, \quad \forall\ \gamma' \subseteq \overline{\gamma}.$$

$\square$

4. fejezet

Markov-hálózatok

Ebben a fejezetben a grafikus modellek egy másik gyakori csoportjáról, az úgynevezett Markov-hálózatokról (Markov-mezőkről) lesz szó.

4.1. Markov-tulajdonságok

A Markov-hálózatok speciális irányítatlan gráfok, melyek valószínűségi változók egy halmazának együttes eloszlását reprezentálják. A szóban forgó gráf csúcsai felelnek meg a valószínűségi változóknak, az élek pedig a köztük lévő feltételes függőségi struktúrát jellemzik.

$G = (V, E)$ irányítatlan gráf esetén vezessük be a következő jelöléseket:

- $\alpha \in A \subseteq V$ esetén $n(\alpha)$ jelentse α szomszédainak halmazát és legyen $n(A) = \bigcup\limits_{\alpha \in A} n(\alpha) \setminus A$.
- $bd(A) = \{v \in V \setminus A : n(v) \in A\}$
- $cl(A) = A \cup bd(A)$

Legyen $X_V = (X_\alpha)_{\alpha \in V}$ valószínűségi változók halmaza az $\mathcal{X}$ véges és diszkrét szorzathalmazon. $A \subseteq V$ esetén legyen $X_A = (X_\alpha)_{\alpha \in A}$ és az egyszerűség kedvéért minden $A, B, C \subset V$ esetén $A \perp\!\!\!\perp B \mid C$ jelentse X_A és X_B valószínűségi változók feltételes függetlenségét X_C-re nézve. Legyen adott $G = (V, E)$ irányítatlan gráf, melynek segítségével definiálhatjuk a $(X_\alpha)_{\alpha \in V}$ valószínűségi változók közötti feltételes függetlenségi viszonyokat. Olyan $\mathcal{X}$-en értelmezett valószínűségi mértékeket tekintünk, melyek az alábbi Markov-tulajdonságok valamelyikével rendelkeznek:

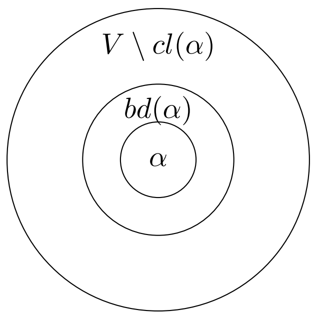
(P) *páronkénti Markov-tulajdonság*: $\forall\ (\alpha, \beta)$ nem-szomszédos csúcspár esetén

$$\alpha \perp\!\!\!\perp \beta \mid V \setminus \{\alpha, \beta\},$$

azaz minden nem szomszédos csúcspár független az összes többi csúcstól.

(L) *lokális Markov-tulajdonság*: $\forall\ \alpha \in V$ esetén

$$\alpha \perp\!\!\!\perp V \setminus cl(\alpha) \mid bd(\alpha)$$



(G) *globális Markov-tulajdonság:* $\forall A, B, C \subset V$ diszjunkt részgráfok esetén, ahol C szeparáló részgráf

$$A \perp\!\!\!\perp B \mid C.$$

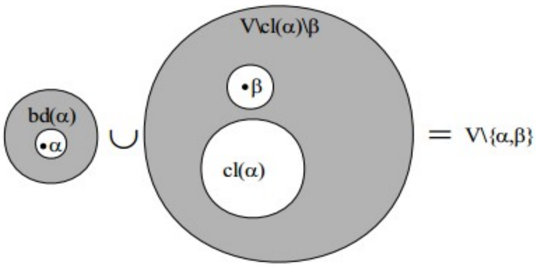
Megjegyzés. $C \subset V$ szeparáló részgráfja $A, B \subset V$ részgráfoknak, ha minden A-ban kezdődő és B-ben végződő út C halmazon keresztül vezet.

Állítás 4.1.1

Tetszőleges $G = (V, E)$ irányítatlan gráf és valószínűségi eloszlás esetén

$$\textbf{(G)} \Rightarrow \textbf{(L)} \Rightarrow \textbf{(P)}.$$

Bizonyítás. A globális Markov-tulajdonságból **(G)** következik a lokális Markov-tulajdonság **(L)**, ugyanis $bd(\alpha)$ szeparáló részgráf α és $V \setminus cl(\alpha)$ között. Mivel α és β nem szomszédosak, ezért $\beta \in V \setminus cl(\alpha)$. Azt kell még megmutatnunk, hogy a lokális Markov-tulajdonságból következik a páronkénti Markov-tulajdonság. Valóban, ugyanis $bd(\alpha) \cup ((V \setminus cl(\alpha)) \setminus \{\beta\}) = V \setminus \{\alpha, \beta\}$.



Alkalmazzuk a következő, valószínűségszámításból jól ismert állításokat, melyeket a szemléletesség kedvéért külön lemmában fogalmazunk meg.

Lemma. X, Y, Z, U valószínűségi változók és tetszőleges $\mathcal{X}$-en értelmezett h mérhető függvény esetén

i)

$$X \perp\!\!\!\perp Y \mid Z;\ U = h(X) \Rightarrow X \perp\!\!\!\perp Y \mid (Z, U)$$

ii)

$$X \perp\!\!\!\perp Y \mid Z;\ U = h(X) \Rightarrow U \perp\!\!\!\perp Y \mid Z.$$

Felhasználva, hogy $\alpha \perp\!\!\!\perp V \setminus cl(\alpha) \mid bd(\alpha)$ és alkalmazva a fenti lemmát $X = \alpha$, $Y = V \setminus cl(\alpha)$, $Z = bd(\alpha)$, $U = h(Y) = Y \setminus \{\beta\}$ szereposztással kapjuk következőt

$$\alpha \perp\!\!\!\perp (V \setminus cl(\alpha)) \mid (bd(\alpha) \cup V \setminus cl(\alpha)) \setminus \{\beta\},$$

melyből

$$\alpha \perp\!\!\!\perp (V \setminus cl(\alpha)) \mid V \setminus \{\alpha, \beta\}$$

adódik. Mivel $\beta \in V \setminus cl(\alpha)$, így

$$\alpha \perp\!\!\!\perp \beta \mid V \setminus \{\alpha, \beta\}.$$

$\square$

A globális Markov-tulajdonság nagy jelentőséggel bír, ugyanis általános feltételt ad bizonyos valószínűségi változó halmazok közötti feltételes függetlenség eldöntésére, bonyolult gráfok esetén azonban nehéz ezt a feltételt ellenőrizni. Létezik-e a globális Markov-tulajdonságot karakterizáló egyéb eljárás?

Definíció 4.1.1

*Azt mondjuk, hogy az $\mathcal{X}$ alaphalmazon értelmezett P valószínűségi mérték faktorizálható**(F)** $G = (V, E)$ gráf felett, ha minden $a \subseteq V$ teljes részgráf esetén létezik olyan ψ_a függvény, amely csak X_A valószínűségi változóktól függ és melynek sűrűségfüggvényére*

$$f(x) = \prod_{a \text{ teljes részgráf}} \psi_a(x)$$

teljesül $\mu = \otimes_{\alpha \in V} \mu_\alpha$ $\mathcal{X}$-en értelmezett szorzatmértékre nézve, továbbá a vele ekvivalens szorzatmértékekre egyaránt.

Megjegyzés

Legyen $\mathcal{P}$ azon loglineáris eloszláscsalád, amelyben a megfelelő hipergráfokból származtatott 2-metszet grafikus hipergráfot ad. Ekkor $\forall P \in \mathcal{P}$ előáll a fenti alakban.

Tétel 4.1.1 (Hammersley-Clifford [1])

Egy P valószínűségi eloszlás pozitív és folytonos f sűrűségfüggvénnyel pontosan akkor globális Markov-tulajdonságú egy $G = (V, E)$ irányítatlan gráfra nézve, ha faktorizálható G felett.

Bizonyítás.

Láttuk, hogy a globális Markov-tulajdonságból következik a páronkénti Markov-tulajdonság. A bizonyítás során először belátjuk, hogy a páronkénti Markov-tulajdonság implikálja a faktorizálhatóságot,majd a faktorizálhatóságból következtetünk a globális Markov-tulajdonságra.

A bizonyításhoz szükségünk lesz a következő lemmára:

Lemma (*Möbius-féle inverziós lemma*)

Legyenek ψ, $\phi : \mathcal{P}(V) \to G$ függvények, ahol V véges halmaz és G Abel-csoport. Ekkor a következők ekvivalensek:

$$\forall a \subseteq V : \psi(a) = \sum_{b:b\subseteq a} \phi(b) \tag{4.1}$$

$$\forall a \subseteq V : \phi(a) = \sum_{b:b\subseteq a} (-1)^{|a\setminus b|}\psi(b) \tag{4.2}$$

Lemma bizonyítása:

Behelyettesítve *(4.2)*-t az első egyenletbe:

$$\sum_{b:b\subseteq a} \phi(b) = \sum_{b:b\subseteq a} \sum_{c:c\subseteq b} (-1)^{|b\setminus c|}\psi(c) = \sum_{c:c\subseteq b} \psi(c) \sum_{\substack{b:c\subseteq b\\ b\subseteq a}} (-1)^{|b\setminus c|} = \sum_{c:c\subseteq b} \psi(c) \sum_{h:h\subseteq a\setminus c} (-1)^{|h|}$$

Az utolsó átalakítás megtehető, ugyanis $b : c \subseteq b; b \subseteq a$ és $a \setminus c$ részhalmazok halmazai megegyeznek abban az értelemben, hogy $c \subseteq b$ megkövetelése mellett a szóban forgó részhalmazok száma $2^{|c|}$ részére csökken, így a lehetséges részhalmazokból mindkét esetben $2^{|a|-|c|} = 2^{|a-c|}$ van, melyek számossága is egyenlő.

Teljesül továbbá, hogy

$$\sum_{h:h\subseteq a\setminus c} (-1)^{|h|} = 0$$

$\forall a \setminus c \neq \emptyset$ esetén, ahol $a \setminus c = \emptyset$ akkor lehetséges, ha $a = c$, így

$$\sum_{c:c\subseteq b} \psi(c) \sum_{h:h\subseteq a\setminus c} (-1)^{|h|} = \phi(a).$$

$\square$

Visszatérve az állítás bizonyítására; legyen X_A az $A \subseteq V$ részgráfhoz tartozó valószínűségi változók halmaza, ahol V a $G = (V, E)$ irányítatlan gráf csúcshalmaza.

Láttuk, hogy **(G)** $\Rightarrow$ **(L)** $\Rightarrow$ **(P)**, ezért elegendő megmutatnunk, hogy **(P)** $\Rightarrow$ **(F)** és **(F)** $\Rightarrow$ **(G)**. Induljunk ki a faktorizálhatóság definíciójából, és vegyük mindkét oldal logaritmusát. Ez megtehető, mert f-ről feltettük, hogy pozitív.

$$log\ f(x) = \sum_{a:a\subseteq V} \phi_a(x),$$

ahol $\phi_a(x) = log\ \psi_a(x)$ és $\phi_a \equiv 0$, ha $a \subseteq V$ nem teljes.

Tegyük fel, hogy P páronkénti Markov-tulajdonságú és legyen $x^* \in X$ tetszőleges és rögzített. Vezessük be a

$$H_a(x) = log f(x_a, x^*_{a^c})$$

függvényt $\mathcal{X}$-en.

Ekkor azt mondhatjuk, hogy rögzített x^* esetén $H_a(x)$ csak x_a koordinátákon keresztül függ x-től. A továbbiakban definiáljuk $\forall a \subseteq V$ esetén

$$\phi_a(x) = \sum_{b:b\subseteq a} (-1)^{|a\setminus b|} H_b(x)$$

kifejezést, ahol $\phi_a(x)$ csak x_a-n keresztül függ x-től, mivel $H_b$csak x_b-től függ és b végigfutja a-t. A Möbius inverziós lemma segítségével megmutatjuk, hogy

$$H_V(x) = \sum_{a:a\subseteq V} \phi_a(x).$$

(A lemma szerint ez V minden részhalmazára teljesül, de most elegendő $a \subseteq V$ esetén.)

$H_a(x)$ definíciójából következik, hogy

$$log f(x) = H_V(x),$$

ezért

$$log f(x) = H_V(x) = \sum_{a:a\subseteq V} \phi_a(x).$$

A faktorizálhatósághoz még meg kell mutatnunk, hogy $\phi_a(x) \equiv 0$, ha $a \subseteq V$ nem teljes.

Tegyük fel, hogy $\alpha, \beta \in a$, ahol α és β nem szomszédosak, továbbá legyen $c = a \setminus \{\alpha, \beta\}$.

Ekkor

$$\phi_a(x) = \sum_{b:b\subseteq a} (-1)^{|a\setminus b|} H_b(x) = \sum_{b:b\subseteq(c\cup\{\alpha,\beta\})} (-1)^{|a\setminus b|} H_b(x) =$$

$$= \sum_{b:b\subseteq c} (-1)^{|a\setminus b|} H_b(x) + \sum_{b:b\subseteq c} (-1)^{|a\setminus b\cup\{\alpha\}|} H_{b\cup\{\alpha\}}(x) +$$

$$+ \sum_{b:b\subseteq c} (-1)^{|a\setminus b\cup\{\beta\}|} H_{b\cup\{\beta\}}(x) + \sum_{b:b\subseteq c} (-1)^{|a\setminus(b\cup\{\alpha,\beta\})|} H_{b\cup\{\alpha,\beta\}}(x) =$$

$$= \sum_{b:b\subseteq c} (-1)^{|a\setminus b|} H_b(x) - \sum_{b:b\subseteq c} (-1)^{|a\setminus b|} H_{b\cap\{\alpha\}}(x) -$$

$$- \sum_{b:b\subseteq c} (-1)^{|a\setminus b|} H_{b\cup\{\beta\}}(x) + \sum_{b:b\subseteq c} (-1)^{|a\setminus b|} H_{b\cup\{\alpha,\beta\}}(x) =$$

$$= \sum_{b:b\subseteq c} (-1)^{|a\setminus b|} (H_b(x) - H_{b\cup\{\alpha\}}(x) - H_{b\cup\{\beta\}}(x) + H_{b\cup\{\alpha,\beta\}}(x)) =$$

$$= \sum_{b:b\subseteq c} (-1)^{|c\setminus b|} (H_b(x) - H_{b\cup\{\alpha\}}(x) - H_{b\cup\{\beta\}}(x) + H_{b\cup\{\alpha,\beta\}}(x)).$$

Ha sikerül megmutatni, hogy $H_b(x) - H_{b\cup\{\alpha\}}(x) - H_{b\cup\{\beta\}}(x) + H_{b\cup\{\alpha,\beta\}}(x) = 0$, akkor készen vagyunk.

Legyen $d = V \setminus \{\alpha, \beta\}$.A pozitivitást és folytonosságot kihasználva kapjuk, hogy

$$H_{b\cup\{\alpha,\beta\}}(x) - H_{b\cup\{\alpha\}}(x) = log \frac{f(x_b, x_\alpha, x_\beta, x^*_{d\setminus b})}{f(x_b, x_\alpha, x^*_\beta, x^*_{d\setminus b})} =$$

$$= log \frac{f(x_\alpha \mid x_\beta, x_b, x^*_{d\setminus b}) f(x_\beta, x_b, x^*_{d\setminus b})}{f(x_\alpha \mid x^*_\beta, x_b, x^*_{d\setminus b}) f(x^*_\beta, x_b, x^*_{d\setminus b})} = log \frac{f(x_\alpha \mid x_b, x^*_{d\setminus b}) f(x_\beta, x_b, x^*_{d\setminus b})}{f(x_\alpha \mid x_b, x^*_{d\setminus b}) f(x^*_\beta, x_b, x^*_{d\setminus b})} =$$

$$= log \frac{f(x^*_\alpha \mid x_b, x^*_{d\setminus b}) f(x_\beta, x_b, x^*_{d\setminus b})}{f(x^*_\alpha \mid x_b, x^*_{d\setminus b}) f(x^*_\beta, x_b, x^*_{d\setminus b})} = log \frac{f(x_b, x_\beta, x^*_\alpha, x^*_{d\setminus b})}{f(x_b, x^*_\alpha, x^*_\beta, x^*_{d\setminus b})} = H_{b\cup\{\beta\}}(x) - H_b(x).$$

Végezetül belátjuk, hogy **(F)** $\Rightarrow$ **(G)**.

Legyenek $A, B, S \subseteq V$ diszjunkt részgráfok, melyekre teljesül, hogy S szeparáló részgráf A és B között, továbbá jelölje $\tilde{A}$ a $G_{V\setminus S}$ azon összefüggő komponensét, melyek tartalmaznak A-ból csúcsot, azaz

$$\tilde{A} = \bigcup_{\alpha \in A} [\alpha]_{V \setminus S},$$

ahol $[\alpha]_{V \setminus S}$ $V \setminus S$ azon komponense, amely A-t tartalmazza.

Vezessük be $\tilde{B} = V \setminus (\tilde{A} \cup S)$ halmazt. Ekkor $B \in \tilde{B}$.

Mivel S az A és B részgráfok szeparáló részgráfja,továbbá A-nak és B-nek a $G_{V \setminus S}$ különböző összefüggő komponenseiben van közös elemük, ezért G gráf maximális teljes részgráfjai csak $\tilde{A} \cup S$ *vagy* $\tilde{B} \cup S$ részgráfjai lehetnek. Ellenkező esetben S kikerülésével is eljuthatnánk $\tilde{A}$-ból $\tilde{B}$-ba, de ez S részgráf tulajdonsága miatt nem lehetséges.

Jelölje $\mathcal{C}$ a maximális teljes részgráfok halmazát G-ben. Ekkor:

$$\mathcal{C} = \mathcal{C}_A \cup \mathcal{C}_B,$$

ahol $\mathcal{C}_A$ az $\tilde{A} \cup S$ maximális teljes részgráfjait jelöli és $C_B = \widetilde{B} \cup S$. Abbanaz esetben, ha S maximális teljes részgráf lenne, akkor tetszőlegesen eldöntjük, hogy melyikbe rakjuk bele. Kihasználva a faktorizálhatóságot, ekkor a következőt írhatjuk:

$$f(x) = \prod_{c \in \mathcal{C}} \phi_c(x) = \prod_{c \in \mathcal{C}_A} \phi_c(x) \prod_{c \in \mathcal{C} \setminus \mathcal{C}_A} \phi_c(x) = h(x_{\tilde{A} \cup S}) k(x_{\tilde{B} \cup S})$$

alkalmas h és k mérhető függvények esetén. Ebből azonban következik, hogy $\tilde{A} \perp\!\!\!\perp \tilde{B} \mid S$, így

$$A \perp\!\!\!\perp B \mid S,$$

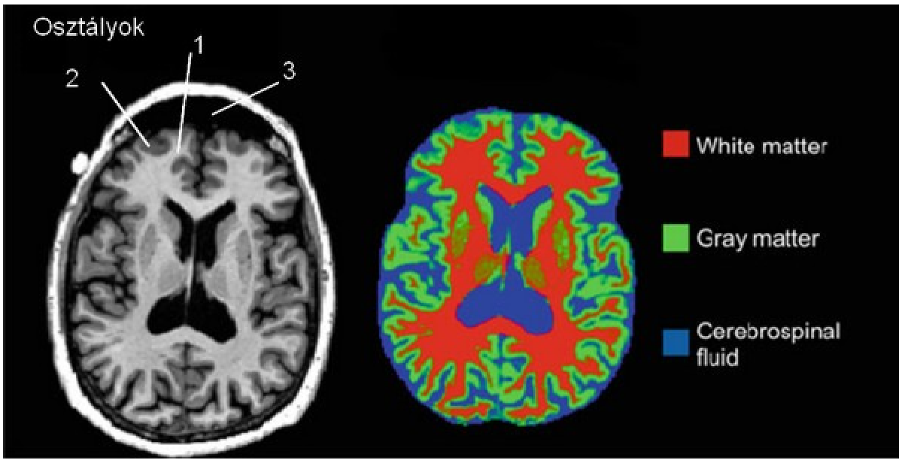
ami éppen a globális Markov-tulajdonság.

$\square$

4.2. Markov-mezők a képfeldolgozásban

A képszegmentálás a képfeldolgozás egyik legfontosabb alapproblémája, mely a hasonló tulajdonságú pixelek homogén területekbe történő csoportosításával foglalkozik. A probléma igazi nehézsége abban rejlik, hogy maga a szegmentálási folyamat függ a képről meglévő előzetes ismereteinktől is. Másra vagyunk kíváncsiak például egy meteorológiai műholdfelvétel és egy megfigyelőrendszer kamerájából érkező kép szegmentálása során. Ezért a gyakorlati alkalmazás során előtérbe kerülnek azok a megoldások, amelyekben lehetőségünk van ezeknek az előzetes információk kódolására. Ilyen tulajdonságokkal rendelkeznek a Markov mezőket alkalmazó képmodellek, amelyek egy véletlen mező segítségével írják le a bemeneti kép és a szegmentált kép közötti összefüggést. A képszegmentálás kimagaslóan nagy jelentőséggel bír MRI-felvételek kiértékelése során, ugyanis lehetőséget biztosít az esetleges elváltozások felismerésére és elkülönítésére. Az egyik legkritikusabb terület az agydaganatok azonosítása, melyek a legkülönbözőbb alaki és szövettani tulajdonságokkal rendelkeznek, megnehezítve az egészséges agyterületektől való elkülönítést.

Az azonosítás során voxelenként (háromdimenziós kép legkisebb megkülönböztethető egysége) vizsgálják a szürkeállományban lévő eltéréseket. Elsőként a szövettípusok szegmentációjával foglalkoznak, vagyis az agy különböző részeinek szövettípusok mentén való szétválasztásával, melynek során az agyi (GM-grey matter, WM-white matter) és nem-agyi (CSF-agy-gerincvelői folyadék) szöveteket csoportosítják és kategóriákba sorolják. Ehhez az MRI-felvételen látható agyterületek intenzitásait használják, mellyel feltérképezik az esetleges elváltozásokat. A szegmentálás fő feladata, hogy a felvételeket megfelelően partícionálják az előre meghatározott kritériumok alapján. A Markov-mezők segítségével kihasználható, hogy a tapasztalatok szerint a térben szomszédos agyterületeknek megfelelő voxelek homogenitást mutatnak, vagyis az egyes voxelek ugyanazon kategóriához tartoznak, mint a környezetükben lévő szomszédos voxelek. Tehát az egyes voxelek kategóriája csak az őket körülvevő szomszédosak kategóriájától függ.



Jelölje $s = s_i = (s_{ix}, s_{iy}, s_{iz})\ \ i = 1, \ldots, m$ az adott felvétel voxeleit, ahol m a voxelek száma és legyen S a voxelek halmaza. Legyen N_i az s_i voxel szomszédait tartalmazó halmaz. Minden voxelhez tartozik egy $f = \{f_i : i = 1, \ldots, m\}$ megfigyelésvektor, amely a megfigyelt intenzitásokat tartalmazza. Adott továbbá $\Omega = \{1, \ldots, K\}$ a szövet osztályainak halmaza, melyet anatómiai ismeretekre támaszkodva határoznak meg. Minden voxelhez tarozik egy kijelölt osztály, jelölje ω_i az s_i-nek megfelelőt és legyen $\omega = \{\omega_i \in \Omega : i = 1, \ldots, m\}$. Az első feladat egy olyan $\hat{\omega}$ megtalálása, amely maximalizálja a $P(\omega \mid X)$ valószínűséget, azaz egy olyan osztály keresése, amely legjobban közelíti a megfigyeléseket:

$$\hat{\omega} = arg \max_{\omega \in \Omega} \ P(\omega \mid f)$$

Ez az úgynevezett maximum a posterioiri (MAP) becslés. Általában minden $\omega \in \Omega$ osztályt normális eloszlás segítségével reprezentálnak, melynek paramétereit a tapasztalati adatok alapján becslik

$$\mu_\omega = \frac{1}{\mid S_\omega \mid} \sum_{s \in S_\omega} f_s$$

$$\sigma^2_\omega = \frac{1}{\mid S_\omega \mid} \sum_{s \in S_\omega} (f_s - \mu_\omega)^2,$$

ahol S_ω jelöli azon voxeleket, melyek a ω osztály egy kiválasztott részhalmazában szerepelnek, vagy a felvétel alapján készült hisztogramról leolvasott adatok alapján számolják. Ezt az eljárást gyakran *FGMM* (*Finite Gaussian Mixture Model*) modellként emlegetik, amely nem használja ki a voxelek térbeli elhelyezkedését, így gyakran a koponyaüreg egyes részeit is agyi szövetekként osztályozza. Ennek kiküszöbölésében segít a Markov-mezők alkalmazása. Ahhoz, hogy Markov-mezőt definiáljunk, két tulajdonságnak kell teljesülnie

1. $\forall \omega \in \Omega : \; P(X = \omega) > 0$

2. $P(\omega_i \mid \omega_{S-\{i\}}) = P(\omega_i \mid \omega_{N_i})$

A második tulajdonság szerint tehát elegendő a voxelek lokális tulajdonságait vizsgálni.

A továbbiakban szükség lesz a következő definícióra:

Definíció 4.2.1

Legyen adott G gráf, melynek csúcshalmaza valószínűségi változókat, élhalmaza pedig a közöttük lévő kapcsolatokat reprezentálja. Jelöle C a gráf maximális teljes részgráfjait, röviden klikkjeit. Az $X = \{X_1, \ldots, X_n\}$ valószínűségi változók együttes eloszlása Gibbs-eloszlás a G gráf felett, ha

$$P(X) = \frac{1}{Z} \prod_{c_i \in C} \phi_i(c_i),$$

ahol $\phi_i(c_i)$ az úgynevezett klikk-potenciál, melynek értéke csak az i-edik klikkbe tartozó valószínűségi változók értékeitől függ, továbbá $Z = \sum_x \prod_{c_i \in C} \phi_i(c_i)$ normalizáló konstans, amely biztosítja, hogy valószínűségi eloszlást kapjunk. Ekkor azt mondjuk, hogy X Gibbs-mezőt alkot a megfelelő alaphalmazon.

A klikk-potenciál esetünkben mindig pozitív és gyakran a következő alakot ölti:

$$\phi_i(c_i) = exp(-V(c_i)),$$

ahol V a c_i-től függő energiafüggvény. Ebben az esetben

$$P(X) = \frac{1}{Z} exp\left(-\sum_{c_i \in C} V(c_i)\right).$$

Mi a kapcsolat a Gibbs-eloszlás és a Markov-mezők között? A Hammersley-Clifford tétel egy másik alakja szerint([2]):

Tétel 4.2.1

P valószínűségi eloszlás Gibbs-eloszlás pontosan akkor, ha a hozzátartozó valószínűségi változók Markov-mezőt definiálnak.

Visszatérve a képszegmentálási feladatra, a fenti tétel szerint

$$P(\omega \mid f) = \frac{1}{Z}\ exp\left(-U(\omega)\right) = \frac{1}{Z} exp\left(-\sum_{c \in C} V_c(\omega)\right),$$

ahol $C \subseteq S$ klikkek halmaza, amelyben bármely két voxel szomszédos. Egy n voxelből álló klikket n-edrendű klikknek hívunk, ezek halmazát jelölje C_n, az összes klikk halmaza pedig legyen $C = C_1 \cup C_2 \cup \ldots \cup C_n$. Megjegyezzük, hogy a képszegmentálás során általában elegendő az egy-vagy kételemű klikkek halmazát vizsgálni. Minden klikkhez tartozik egy, az ω osztályozásnak megfelelő $V_c(\omega)$ klikkpotenciál, melynek segítségével definiálható az energiafüggvény:

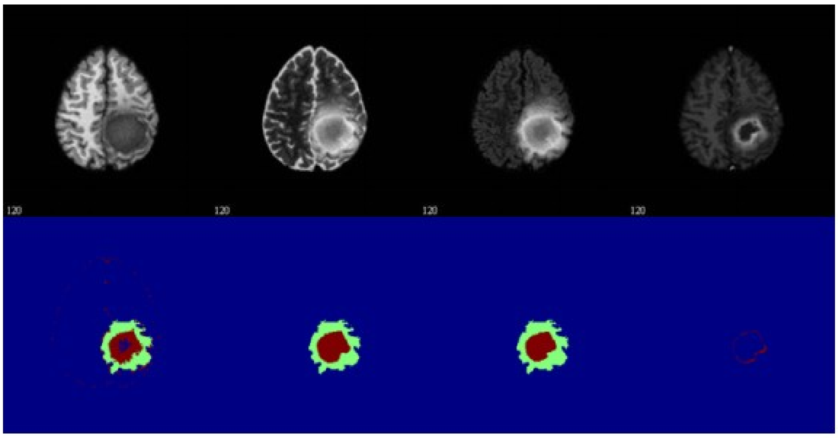
$$U(\omega) = \sum_{c \in C} V_c(\omega) = \sum_{i \in C_1} V_{C_1}(\omega_i) + \sum_{(i,j) \in C_2} V_{C_2}(\omega_i, \omega_j)$$

Ekkor a feladat a követkő módon fogalmazható át

$$\hat{\omega} = arg \max_{\omega \in \Omega}\ P(\omega \mid f) = arg \min_{\omega \in \Omega} U(\omega).$$

A problémát tehát a Markov-mezők segítségével visszavezettük egy nem-konvex energiafüggvény minimalizálására, melynek számos lokális minimuma van, így megoldása nehéz, de lokális optimumok keresésére alkalmas iteratív optimalizálási algoritmusokkal kiszámolható ([7]).

Az alábbi ábra egy képszegmentálási eljárás eredményét szemlélteti. Az első sorban az agyról készült MRI felvételek láthatóak az eltérő spin-rács és a spin-spin relaxációs időket figyelembe véve, melyeket változtatva az MRI-felvétel kontrasztja mesterségesen befolyásolható. Az első két oszlop mutatja a tumor szegmentációját a Markov-mezőket alkalmazó módszer segítségével, a harmadik oszlopban a kezdeti megfigyelésekből származó feltételezett daganat látható, míg az utolsó oszlop a szegmentálási eljárás és a feltételezés közötti differenciát demonstrálja.



Irodalomjegyzék

[1] S.L. Lauritzen, *Graphical Models*, NY: Oxford University Press, pp. 279, 1996.

[2] J.N. Darroch, S.L. Lauritzen, T.P. Speed, *Markov Fields And Log-Linear Interaction Models For Contingency Tables*, The Annals Of Statistics, 1980., Vol. 8, No.3, 522-539.

[3] S.L. Lauritzen, T.P. Speed,K. Vijayan, *Decomposable Graphs And Hypergraphs*, J.Austral. Math, Soc. Series A 36, 1984., 12-29.

[4] Stephen E. Fienberg and Alessandro Rinaldo, *Maximum Likelihood Estimation In Log-Linear Models*, The Annals of Statistics, 2012., Vol. 40, No. 2, 996–1023.

[5] Tianming Zhan, Shenghua Gu, Can Feng, Yongzhao Zhan, Jin Wang, *Brain Tumor Segmentation from multispectral MRIs Using Sparse Representation Classification and MRF Regularization* Advanced Science and Technology Letters, Vol.81, CST 2015., 130-134.

[6] D. Ferrario, M. Bach Cuadra, M. Schaer, N. Houhou, D. Zosso, S. Eliez, L. Guibaud and J.-Ph. Thiran, *Brain Surface Segmentation of Magnetic Resonance Images Of The Fetus*, 2008., http://www.eurasip.org/Proceedings/Eusipco/Eusipco2008/papers/1569103142.pdf

[7] Stan Z. Li, *Markov Random Field Modeling in Image Analysis* Computer Science Workbench, Springer 2001.

[8] Zoltan Kato, Josiane Zerubia, *Markov Random Fields in Image Segmentation* Foundations and Trend in Signal Processing Vol. 5, Nos. 1–2, 2011., 1–155.