Antal Péter

Jelölések*

Felhasznált jelölések

x,$\underline{x}$,$\underline{\underline{x}}$	skalár, (oszlop)vektor vagy halmaz, mátrix
$X, x, p(X)$	véletlen változó X, érték x, valószínűségi tömegfüggvény/sűrűségfüggvény X
$\mathrm{E}_{X,p(X)}[f(X)]$	$f(X)$ várható értéke $p(X)$ szerint
$\mathrm{var}_{p(X)}[f(X)]$	$f(X)$ varianciája $p(X)$ szerint
$I_p(\underline{X}\|\underline{Z}\|\underline{Y})$	$\underline{X}$ és $\underline{Y}$ megfigyelési függetlensége $\underline{Z}$ feltétellel p esetében
$(X \perp\!\!\!\perp Y\|Z)_p$	$I_p(\underline{X}\|\underline{Z}\|\underline{Y})$
$(X \not\!\perp\!\!\!\perp Y\|Z)_p)$	$\neg I_p(\underline{X}\|\underline{Z}\|\underline{Y})$
$CI_p(\underline{X};\underline{Y}\|\underline{Z})$	$\underline{X}$ és $\underline{Y}$ beavatkozási függetlensége $\underline{Z}$ feltétellel p esetében
$\prec$	(részleges) sorrendezés
$\prec^c$	a változók egy teljes sorrendezése
$\prec^G$	adott G irányított körmentes gráffal kompatibilis sorrendek halmaza
$\prec(n)$	n objektum sorrendjeinek (permutációinak) a halmaza
$G, \underline{\theta}$	Bayes-háló struktúrája és paraméterei
$G^{\sim}$	G irányított körmentes gráf esszenciális gráfja
$\mathcal{G}(n)/\mathcal{G}^k(n)$	n csomópontú maximum k szülőjű DAG-ok halmaza
$\mathcal{G}^{\prec}$	adott $\prec$ sorrenddel kompatibilis DAG-ok halmaza
$\mathcal{G}^G$	adott G DAG-gal megfigyelési ekvivalens DAG-ok halmaza
$\sim$	kompatibilitási reláció
$\mathrm{pa}(X_i, G) \sim\prec$	$\mathrm{pa}(X_i, G)$ szülői halmaz kompatibilis $\prec$ sorrendezéssel
$\mathrm{MB}_p(X_i)$	Markov-takarója X_i-nek p-ben
$\mathrm{pa}, \mathrm{pa}(X_i, G)$	szülői változók halmaza, X_i szüleinek halmaza G-ben
pa_{ij}	a j. konfigurációja a szülői értékeknek egy sorrendben
$\mathrm{bd}(X_i, G)$	X_i szüleinek, gyerekeinek és gyerekei egyéb szüleinek halmaza G-ben
$\mathrm{MBG}(X_i, G)$	a Markov-takaró algráfja X_i-nek G-ben
$\mathrm{MBM}(X_i, X_j, G)$	a Markov-takaróbeliség relációja
n	valószínűségi változók száma
k	maximális szülőszám DAG-okban
N	mintaszám
V	összes valószínűségi változók száma
Y	válasz, kimeneteli, függő változó

*További konvenciók az egyes fejezetekben jelöltek.

$N_+/N_{\ldots,+,\ldots}$	$N_i/N_{\ldots,i,\ldots}$ megfelelő összegei
$D\|X$	X változóhalmazra szűkített adathalmaz
$\|\|$	kardinalitás
$1()$	indikátorfüggvény
f', f''	f függvény első és második deriváltjai
A^T	A mátrix transzponáltja
$\underline{x}\cdot\underline{y}$	$\underline{x}$ és $\underline{y}$ vektorok skalárszorzata
ξ^+/ξ^-	informatív/nem informatív információs kontextus
$\neg,\wedge,\vee,\neq,\rightarrow$	standard logikai operátorok
$\cap,\cup,\setminus,\Delta$	standard halmazműveletek
$KB\vdash_i\alpha$	α bizonyíthatósága KB-ből
Γ	a Gamma függvény
$\mathrm{Beta}(x\|\alpha,\beta)$	a Béta eloszlás sűrűségfüggvénye (pdf)
$\mathrm{Dir}(x\|\underline{\alpha})$	a Dirichlet eloszlás sűrűségfüggvénye
$\mathrm{N}(x\|\mu,\sigma)$	az egyváltozós normál eloszlás sűrűségfüggvénye
$\mathrm{N}(x\|\underline{\mu},\underline{\underline{\Sigma}})$	a többváltozós normál eloszlás sűrűségfüggvénye
BD,BD_e	Bayesian Dirichlet prior, megfigyelési ekvivalens BD prior
BD_{CH}	Bayesian Dirichlet (BD) prior 1 hiperparaméterekkel
$\mathrm{BD}_{\mathrm{eu}}$	megfigyelési ekvivalens és uniform BD prior
$L(\underline{\theta};D_N)$	$p(D_N\|\underline{\theta})$ likelihood függvénye
$H(X,Y)$	X és Y entrópiája
$I(X;Y)$	X és Y kölcsönös információja
$\mathrm{KL}(X\\|Y)$	X és Y Kullback–Leibler divergenciája
$H(X\\|Y)$	X és Y keresztentrópiája
$L_1(,),\ L_2(,)$	az abszolútértékbeli (Manhattan) négyzetes (euklidészi) távolságok
$L_0(,)$	0-1 veszteség
$\mathcal{O}()/\Theta()$	aszimptotikus, nagyságrendi felső és alsó határ

Rövidítések

ROC	Receiver Operating Characteristic (ROC) görbe
AUC	ROC-görbe alatti terület
BMA	bayesi modell átlagolás
BN	Bayes-háló
DAG	irányított körmentes gráf
FSS	jegykiválasztási probléma
MAP	maximum a posteriori
MI	kölcsönös információ
ML	maximum likelihood
MBG	Markov-határ gráf
MB	Markov-takaró
MBM	Markov-takaróbeliség
(MC)MC	(Markov-láncos) Monte Carlo
NBN	naiv Bayes-háló

Tartalomjegyzék

1. fejezet

Valószínűségi gráfos modellek

A fejezetben összefoglaljuk a Bayes-statisztikai keretet, különösen a megközelítés által igényelt bizonytalanságokat kezelő modell tulajdonságait. Ezt követően a valószínűségi eloszlások strukturális jellemzőit vizsgáljuk meg, ami a strukturális jellemzők explicit reprezentálásán át elvezet a valószínűségi gráfos modellosztályok használatához. A valószínűségi gráfos modellosztályon belül elsőként az egyszerű Naív Bayes-háló, Markov-lánc és rejtett Markov modell modelltípusokat foglaljuk össze, majd a Bayes-hálók és a Markov-hálók általános definícióit adjuk meg és tekintjük át. Megvizsgáljuk a reprezentációk pontosságát, kiterjesztéseiket és azok tudásmérnöki vonatkozásait.

1.1. Bevezetés

A jegyzetben a bizonytalanság különböző megjelenési formáit kizárólagosan a valószínűségelmélet keretében formalizáljuk. A valószínűségi alapokon történő megközelítés mellett, különösen a szubjektív valószínűségi értelmezéshez szorosan kapcsolódó bayesi megközelítés mellett számos érvet fel lehet sorakoztatni. A terjedelmi korlátokon belül ezek közül többet is bemutatunk, amelyek közül kiemelkednek az axiomatikus (avagy meta-axiomatikusnak is nevezhető) megközelítések, mint amilyen a döntéselméleti alapú érvelés [4]. A bayesi megközelítés induktív következtetésben, statisztikai következtetésben való tárgyalása a jelen jegyzet Bayesi döntés- és becsléselmélet című fejezetében és az Intelligens adatelemzés című jegyzet több fejezetében található. A szakirodalomból a következő alapművek ajánlhatóak [2, 20, 46] (magyar nyelven lásd [26]), mesterséges intelligencián belüli tárgyalása megtalálható például [7, 47]; történeti áttekintésre és trendek felmérésére pedig a következőek ajánlhatóak [3, 38, 1, 14].

A bayesi értelmezéshez szorosan kapcsolódó mintavételi technikákat röviden a jelen jegyzet Következtetési módszerek című fejezetében foglaltuk össze. Ezeket részletesebben az Intelligen adatelemzés jegyzet több fejezetében is tárgyaljuk. (Általános referenciaként lásd [20, 21, 40, 34, 18, 22, 8, 32]).

1.1.1. Racionális bizonytalanságoktól a valószínűség szubjektív értelmezéséig

Az adat- és tudáselemzésben megjelenő események sokfélesége miatt túlzó egyszerűsítésnek tűnhet a több szinten jelen lévő bizonytalanságot — az adat, a tudás, az elemző, a modell, az elemzés, az értelmezés bizonytalanságát — ugyanazon formalizmus keretében és ráadásul akár egyetlen keretben kezelni (a rövidség és egyszerűség kedvéért véges, diszkrét eseményrendszerekkel foglalkozunk általában). A keretrendszer univerzális volta miatt feltehető, hogy a bizonytalanság nem az eseményrendszerhez, hanem azon belül az eseményekhez kapcsolódik (eltérően egyéb módszerektől, mint a fuzzy vagy a Dempster-Shaffer megközelítés). Egy ilyen teljes eseményrendszer definiálása, kölcsönösen kizáró és teljes atomi eseményekkel, különösen nehéz olyan szakterületeken, amelyeken eleve több, autonóm, de bizonytalanul is kapcsolódó szint van. Ilyen az orvosbiológa is. Ennek gyakorlására a tárgyhoz kapcsolódó laborgyakorlatok és anyagok adnak lehetőséget.

Egy helyes eseményrendszert feltételezve az események bizonytalanságának értelmezésére több megközelítés is népszerűvé vált a valószínűségszámítás történetének évszázadai során. Ilyen többek között a szerencsejátékokhoz köthető kombinatorikus értelmezés, a fizikalista avagy propensity alapú megközelítés [44], a frekventista, sorozatok határértékeként való értelmezés (az eredeti, von Misses-hez köthető megközelítés áttekintésére és annak számításelméleti aspektusainak kiterjesztésére lásd [54]). A bizonytalanság formalizálására és a valószínűségek értelmezésére unikális lehetőséget nyújt az axiomatikus megközelítés, amely leegyszerűsítve az események bizonytalanságai feletti logikai (preferencia) reláció feltevéséből bizonyítja be a létezését és egyértelmű voltát a valószínűségszámítás Kolmogorov-féle halmazelméleti megalapozásával kompatibilis „pontszámnak”, racionalitási axiómák (meta-axiómák) feltevésével. Az így származtatott pontszám valószínűségi mértékk;nt használható, amelynek értelmezése a racionalitási axiómákból következően döntéselméleti alapú, de nevezett szubjektív, bayesi, személyes valószínűségi értelmezésnek is (a teljeskörű tárgyalásért lásd [4]). Az axiomatikus megközelítéshez közelinek mondható a pragmatista vagy eszközhasználati értelmezés, amely a valószínűségi megközelítés modellezésben való felhasználását helyezi az értelmezés középpontjába [10, 20, 12]. Megközelítésünkben a szubjektív-pragmatista értelmezést követjük, amely ismételt, valamiféle állandóságot mutató megfigyelések esetén a fizikalista vagy frekventista értelmezéssel is összhangba hozható, azok ontológiai elkötelezettsége nélkül.

A bayesi értelmezés bemutatására sorra vesszük a döntéselméleti alapú axiomatikus származtatás főbb lépéseit. A feltevések egy racionalista szubjektum modellezésének szemszögéből is értelmezhető .

1.1.1. Definíció. [[4]] A döntési problémát $\mathcal{E}, \mathcal{C}, \mathcal{A}, \leq$, definálja, ahol:

(i) $\mathcal{E}$ algebrája az eseményeknek, E_j;

(ii) $\mathcal{C}$ a halmaza a lehetséges következményeknek, c_j;

(iii) $\mathcal{A}$ a halmaza a lehetséges cselekedeteknek, amelyek események partícióit feleltetik meg következményeknek;

(iv) $\leq$ egy bináris preferenciareláció $\mathcal{A}$ elemei felett.

Amint a definíció mutatja az eseményrendszer feletti bizonytalanságot csupán egy preferencia reláció testesíti meg. A preferencia reláció értelmezése kulcsfontosságú: definíció szerint aktív beavatkozáshoz kötődik, de szándékolt (és később bebizonyosodó) jelentése „valószínűbb”. A preferenciarelációból összehasonlíthatósági, tranzitivitás és mennyiségi-folytonossági feltevésekkel az alábbi eredmény bizonyítható.

1. Propozíció. [[4]] Adott $\leq$ bizonytalansági reláció esetén egyértelműen létezik egy valós szám P(E) minden eseményhez, s ezek a számok kompatibilisek $\leq$-val (azaz $E \leq F \Leftrightarrow P(E) \leq P(F)$) és véges, additív valószínűségi mértéket alkotnak.

Egy párhuzamos eredmény korlátos következmények esetén származtatja a preferencia relációval kompatibilis (azaz azt tökélesen modellező) hasznosságok (veszteségek) egyértelmű létét.

2. Propozíció. [[4]] Egy döntési problémában $\mathcal{E}, \mathcal{C}, \mathcal{A}, \leq$ korlátos következmények esetén $c_* < c^*$,

(i) minden c-re a *hasznosság függvény* $u : \mathcal{C} \to \mathcal{R}$ $u(c) = u(c|c_*, c^*)$ létezik és egyértelmű;

(ii) az érték $u(c|c_*, c^*)$ független valamely esemény G feltételezett bekövetkeztétől;

(iii) $0 = u(c_*|c_*, c^*) \leq u(c|c_*, c^*) \leq u(c^*|c_*, c^*) = 1$;

(iv) teljesül az úgynevezett *maximális hasznosság elve* :

$$a_1 \leq_G a_2 \;\Leftrightarrow\; \sum_j u(c_{a_1(\mathrm{E}_j)}) P(\mathrm{E}_j\,|G) \leq \sum_j u(c_{a_2(\mathrm{E}_j)}) P(\mathrm{E}_j\,|G) \tag{1.1}$$

Bár a döntéselméleti keret és a származtatáshoz használt „racionalitási” axiómák megkérdőjelezhetők, más axiomatikus alapok is hasonlóan a valószínűségelméletben megszokott mértékeket származtatják, amelyek kompatibilisek, pontosabban egybeesnek az eseményekhez tartozó bizonytalanságok közötti relációkkal, így végeredményként normatívan írják elő a valószínűségi modellezés használatát. A fenti eredmények kiterjeszthetőek komplex események feletti szigma-algebrákra, amely esetben szigma-additív valószínűségi mértékek adódnak eredményként, a valószínűségelmélet Kolmogorov-féle felépítésének megfelelően [45].

1.1.2. Felcserélhetőségtől a bayesi modellátlagolásig

A bizonytalanságok reprezentálásának és értelmezésének axiomatikus megközelítése megmutatta, hogy a racionalitási axiómákat elfogadva az események feletti bizonytalanságok, sőt döntések feletti preferenciák modellezésére is normatívan adódik a valószínűségszámítás és valószínűségi döntéselmélet. Az autonóm, racionális entitás pillanatnyi állapotának a reprezentálását meghaladó induktív esethez azonban tovább kell lépni, ami egy adott stabilitású megfigyelések esetében a valószínűségi keret szükségszerű kiterjesztéséhez vezet. Ennek bemutatásához idézzük a de Finetti-től származó, bináris eseményekre kimondott reprezentációs tételt, amely általánosabb esetekre is kiterjeszthető.

3. Propozíció. [[4]] Ha $x_1, x_2, \dots$ egy végtelen felcserélhetőségű 0-1 véletlen sorozat P valószínűségi mérték szerint, azaz bármely n-re és permutációra $\pi(1), \dots, \pi(n)$ az együttes valószínűségi tömegfüggvény $p(x_1, \dots, x_n) = p(x_{\pi(1)}, \dots, x_{\pi(n)})$, akkor létezik egy olyan eloszlásfüggvény $\mathcal{Q}$, amelynek segítségével $p(x_1, \dots, x_n)$ felírható a következő alakban

$$p(x_1, \dots, x_n) = \int_0^1 \prod_{i=1}^{n} \theta^{x_i}(1 - \theta^{1-x_i}) \, \mathrm{d}\mathcal{Q}(\theta),$$

Itt

$$\mathcal{Q}(\theta) = \lim_{n \to \infty} P[y_n/n \leq \theta],$$

illetve $y_n = x_1 + \dots + x_n$ és $\theta = \lim_{n \to \infty} y_n/n$.

Az úgynevezett reprezentációs tétel szerint a végtelen felcserélhetőség feltevése, nevezetesen, hogy a szubjektív, bayesi valószínűségek a megfigyelés sorrendjétől függetlenek, azt vonja maga után *mintha* a megfigyelések egymástól feltételesen függetlenek lennének egy hipotetikus mintavételi eloszlás paraméterével vett feltétel esetében. Ezen paraméter felett is, mintegy magasabb szinten, megjelenik egy eloszlás, amely a lehetséges paraméterekre mint aszimptotikus határértékekre vonatkozó elvárásokat reprezentálja. Ámbár a végtelen felcserélhetőség az eredmény aszimptotikus volta miatt (a bayesi megközelítésben éppen olyan fontos) véges esetben kritizálható, analóg reprezentációs eredmények szolgáltatják az axiomatikus megközelítés induktív kiterjesztéséből származó Bayes-statisztikai keretet, amelyben az események feletti bayesi, szubjetív valószínűségek egy modellparaméterezés feletti bayesi, szubjetív valószínűségek által indukáltak. (Ennek véges megfigyelések esetén való relevanciájával kapcsolatban észrevehető, hogy bizonyos véges felcserélhetőségű sorozatoknak nincsen keverékreprezentációja, lásd 226.o.,[4]).

Ezen eredmények összekapcsolása szerint tehát az események (kimenetelek) feletti bizonytalanság valószínűségekkel reprezentálható, s ez a valószínűségi eloszlás maga is parametrikus eloszlások keverékével reprezentáható (ahol az eloszlások feletti eloszlást az eloszlások paraméterei feletti eloszlás definiálja). Máshogyan fogalmazva: az axiomatikus megközelítések eredményei (például de Finetti, Cox–Jaynes, Bernardo eredményei) arra utalnak, hogy a bizonytalanság kezelésénél a valószínűségszámítás standard, additív elmélete szükségszerűen alkalmazandó, különben a rendszer, ágens, szubjektum egy játékelméleti szituációban veszteséget szenved el -[47]. Pontosabban, ahogyan de Finetti „mintha” tételei mutatják (lásd ,, tételeket):

I. : A bizonytalanságok feletti racionális (konzisztens) preferencia rendszer valószínűségekkel leírható.

II. : A kimenetelek (akciók) feletti racionális (konzisztens) preferencia rendszer hasznosságokkal (és a maximális hasznosság elvével) leírható.

III. : A felcserélhetőség feltevése maga után vonja a modellátlagolással való leírhatóságot.

Fontos hangsúlyozni, hogy az általunk követett bayesi-pragmatista megközelítésben a valószínűségi keret univerzális alkalmazása mint modellezési eszköz jelenik meg, s az értelmezés semmiben nem érinti a valószínűségszámítás megszokott axiómáit.

1.2. A Bayes-statisztikai keretrendszer általános sémája

Az axiomatikus-pragmatista megközelítésben adódó, megfigyeléseket is kezelő statisztikai keretrendszerben a bizonytalanságok modellezése technikailig két szintre bontható feladatként fogalmazható meg: az események feletti bizonytalanságokat kifejező valószínűségi modell megalkotása, illetve ezen modell feletti bizonytalanságok modellezése egy másik valószínűségi modellel. Fontos hangsúlyozni, hogy csak technikailag elválasztott a két modell és modellezési szint, a kettő egyetlen, egységes valószínűségszámítási keretben jelenik meg (a „valószínűségek valószínűségeivel" kapcsolatos filozófiai vagy pszichológiai megfontolások matematikai oldaról nem jelennek meg). Mielőtt megvizsgálnánk a valószínűségi gráfos modellek felhasználását ezen technikailag kettős szintű keretrendszerben, összefoglaljuk a Bayes-statisztikai keret általános sémáját. A séma gyakorlati aspektusait fogjuk hangsúlyozni, a lényegének bemutatása végett nem tárgyaljuk a döntéselméleti kapcsolatát (azaz itt csak a Bayes-statisztikai keretet vázoljuk, az általános bayesi döntéselméleti keretet a Becslés- és döntéselmélet című fejezetben tárgyaljuk).

Az axiomatikus származtatás jogosságának elfogadásától függetlenül a Bayes-statisztikai keretrendszer koncepcionálisan nagyon egyszerű, s ez a gyakorlati alkalmazásához szükséges számítási erőforrások megjelenése mellett a népszerűségére is magyarázat. Ebben a statisztikai megközelítésben, parametrikus modelleket feltételezve, egy adott információs ellátottságú ξ szituációban a megfigyelések feletti $p(x|\xi)$ bizonytalan elvárásokat úgy állítjuk elő, hogy első lépésként meghatározzuk a releváns, θ paraméterezésű $p(x|\theta)$ modelleket, majd ezen θ paraméterezés felett egy $p(\theta|\xi)$ valószínűség eloszlást (az x_i mennyiségek a megfigyelhető, a θ paraméter a tipikusan nem megfigyelhető kategóriába esnek). A ξ információs kontextus és a valószínűségek feltételeiben való szerepeltetése a valószínűségek szubjektív értelmezését hivatott hangsúlyozni. Gyakran használt jelölés a ξ^+ és ξ^-, amelyek a neminformatív és informatív szituációkat jelölik. A $p(x, \theta|\xi)$ együttes eloszlás megkonstruálása után a valószínűségszámítás szabályai szerint tetszőleges következtetések lehetségesek uniform módon használva a megfigyelhető x_i mennyiségeket és a nem megfigyelhető θ paraméreket. A következtetések általános formában egy $p(\alpha(x, \theta)|\xi)$ közvetlen valószínűségi állítást jelentenek, amely felfogható személyes elvárásnak* A Bayes-statisztikai keret erejét sok tekintetben a valószínűségszámítás ezen uniform, megfigyeléseket és modellparamétereket egyöntetűen kezelő, koherens volta adja.

A jegyzetben a valószínűségszámításban megszokott jelölésrendszert használjuk. Nagybetűs változók jelölik a véletlen változókat és kisbetűs megfelelőik az értékeiket, például $P_X(X = x)$, ahonnan egyértelműség esetén az eloszlás jelölését és a véletlen változót magát

*Az angolban szinte kizárólagosan használt „belief" bizonytalanságokkal kapcsolatos neutrális volta a magyarban a meggyőződés és elvárás szavakkal adható talán legjobban vissza, de a hit, hiedelem szavakat is használjuk, esetlegesen érzett mellékjelentéseiktől elvonatkoztatva.

is elhagyjuk ($P(x)$). Bináris, propozicionális változók esetében sajnos a konvenció szerint a nagybetű gyakran használt a ponált érték jelölésére is, ezért azon esetekben gyakran a teljes kiírást használjuk ($P(A = igaz)$). Ugyanazt a $p(.)$ jelölést használjuk a valószínűségi tömegfüggvényre és sűrűségfüggvényre, és a megnevezéseiket is megkülönböztetés nélkül használjuk. Ha lehetséges, X jelöli a magyarázó, bemeneti, avagy független változót, és Y jelöli a kimeneteli, függő avagy válasz változót. Általában $D_N = \{x^{(1)}, \ldots, x^{(N)}\}$ jelöli N teljes adatot, azaz amikor a változók teljes $\underline{V}$ halmazában minden változó értéke ismert. Ha szükséges, akkor a vektorokat aláhúzás, a mátrixokat dupla aláhúzás jelöli.

1.2.1. A modell specifikálása a bayesi keretben

Egy idealizált bayesi megközelítésben a felhasznált modellek körét a lehető legtágabbra lehetne választani, természetesen az információs kontextusban jelen lévő strukturális kényszerek, például szimmetriák figyelembevételével. Azonban három aspektust mindenképpen érdemes megfontolni:az Ockham–elv potenciális megsértését, a számítási komplexitást és a komplex modell specifikációjának gyakorlati nehézségeit. Az Ockham elv („Ochkham borotvája") szembeállítása a bayesi keretrendszerrel sokat vitatott kérdés. A még mindig parázsló vita a gyakorlatban úgy oldódik meg, hogy a bayesi megközelítésben a komplexebb, általánosabb modellek kevesebb megerősítést kapnak, mint a megfigyelésekhez hasonlóan illeszkedő, de egyszerűbb modellek [28, 33, 40]. A második, számítási komplexitással kapcsolatos ellenvetést a 2000-es évektől megfigyelhető hardverfejlesztési trendek részben megválaszolják, mivel az egyre elérhetőbb párhuzamos architektúrák, általános célú grafikus kártyák (GPGPU-k), vagy akár a „közüzemi szolgáltatásként" igénybe vehető felhőinfrastruktúrák, nagyon jól kihasználhatóak a bayesi következtetésben a Monte Carlo eljárásokon belül. A harmadik, a komplex modellek specifikálásának nehézségéhez kapcsolódó ellenvetés a fejezet fő témája, nevezetesen, hogy hogyan is lehet hatékonyan komplex valószínűségi modelleket formalizálni. A keretek kijelölése és a fogalmak bevezetése érdekében most csak a hierarchikus modellezés koncepcióját foglaljuk össze.

Hierarchikus modellek axiomatikus származtatásához szintén a felcserélhetőségi megfontolások használhatóak fel, de ekkor már a θ paraméterek szintjén, ami analóg módon vezet egy újabb (technikai) „hiper" szint megjelenéséhez modellek együttese (keveréke) és a hozzájuk tartozó ϕ hiperparaméterek által:

$$p(\theta, \phi) = p(\phi)p(\theta|\phi). \tag{1.2}$$

A gyakorlatban elterjedt megközelítés szerint a hierarchikus specifikációban a releváns $\mathcal{M}^i$ modellosztályok specifikációjával, majd az azokon belüli $\mathcal{S}_k^i$ vagy M_k^i model struktúrák specifikációjával, és végül a modellstruktúrákhoz tartozó θ_k^i paraméterek specifikációjával történik. Ennek megfelelően egy adott i modellosztálybeli k struktúra θ_k^i paraméterezéséhez tartozó *a priori* bizonytalan elvárás egy szorzatként fejezhető ki:

$$p(\theta_k^i, M_k^i, \mathcal{M}^i) = p(\mathcal{M}^i)p(M_k^i|\mathcal{M}^i)p(\theta_k^i|M_k^i). \tag{1.3}$$

A modellek eloszlásainak specifikációját a megfigyelhető mennyiségekre vonatkozó $p(x|\theta, \phi)$

(avagy $p(x|\theta_k^i, M_k^i)$ feltételes eloszlás egészíti ki a Bayes-statisztikai megközelítéshez tartozó teljes együttes eloszlássá.

1.2.2. A prediktív következtetés

A Bayes-statisztikai keret felhasználására és a fogalmak bevezetése végett foglaljuk össze a főbb következtetéstípusokat (az induktív következtés részletesen az Intelligens adatelemzés című jegyzetben található). Az a priori elvárások specifikálása lehetővé teszi az a priori prediktív következtetéseket az x megfigyelhető mennyiségek felett:

$$p(x) = \sum_k p(M_k) \int p(x|\theta_k) p(\theta_k|M_k)\, \mathrm{d}\theta_k. \tag{1.4}$$

Az integrálás és/vagy összegzés a modellek felett, akár modellstruktúrák és paraméterezésük felett, a valószínűségszámításbeli vetítésnek (marginalizációnak) felel meg, és a bayesi kontextusban *bayesi modellátlagolásnak* nevezik [35, 37, 36, 25]. Az a posteriori valószínűségi eloszlások formális bevezetését megelőlegezve, egy D megfigyelési adat esetében a *a posteriori prediktív eloszlás* a D adathalmazon vett feltétellel így írható fel:

$$p(x|D) = \sum_k p(M_k|D) \int p(x|\theta_k) p(\theta_k|D, M_k)\, \mathrm{d}\theta_k. \tag{1.5}$$

Ezek az egyenletek jól illusztrálják a Bayes-statisztikai keret megkülönböztető jegyeit a frekventista valószínűségi értelmezéshez kapcsolódó frekventista statisztikai kerethez képest: a bayesi megközelítés egyetlen fixnek tekintett adathalmaz esetében modellek sokasága felett átlagol (azaz nincs sem modellkiválasztás, sem hipotetikus adathalmazok feletti vizsgálat). A fenti bayesi modellátlagolás normatív származtatása ellenére mint technika is széles körben használt, ilyen például regressziós és klasszifikációs modelleknél a „committee” megközelítésés [5]; Bayes-hálóknál [37]. Mivel a modellátlagolás általában analitikusan nem oldható meg, Monte Carlo módszerek használatosak (általános áttekintését lásd [25]).

Fontos hangsúlyozni, hogy a Bayes-statisztikai megközelítésben a predikció és a prediktív eloszlások jelentik a célt, a modellek csupán eszközkén jelennek meg (bár mint láttuk szükségszerű eszközként). Következésképpen az ideális Bayes-statisztikai eredmény a prediktív eloszlás, amely természetesen nem önmagában jelenik meg (jelentődik), hanem a bayesi döntéselméleti keretben kerül felhasználásra és vezet optimális döntésekhez, például a jelentett mennyiségek révén. Az a priori bizonytalanságok $p(\theta)$ és a $p(x|\theta)$ modell felhasználásával a következtetések másik típusát is bemutatjuk.

1.2.3. A parametrikus következtetés és a Bayes-szabály

Az együttes eloszlás, amelyben a megfigyelt mennyiségeknek és a modellparamétereknek egyforma státuszuk van, a paraméterekre való következtetést is lehetővé teszi. A híres *Ba-*

yes szabály felhasználásával a megfigyelt mennyiségekkel vett feltétel szerinti parametrikus következtetés válik lehetővé:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{\int p(x|\theta)p(\theta)\,\mathrm{d}\theta} \propto p(x|\theta)p(\theta). \tag{1.6}$$

Az 1.6 egyenletben $p(\theta)$ az *a priori eloszlás* vagy prior, $p(x|\theta)$ a *mintavételi eloszlás* amely a *likelihood*-ot és az $L(\theta;x)$ a *likelihood függvény*-t is definiálja. A $p(x)$ az *adat marginális likelihood*-ja, amely csupán egy normalizációs konstanst definiál és $p(\theta|x)$ az *a posteriori eloszlás* a paraméterek felett, vagy egyszerűen poszterior. Az az 1.6 egyenlet azt is mutatja, hogy a poszterior egyensúlyi állapotot jelent a prior és likelihood között, és a prior konstans volta miatt a megfigyelések növekvő száma esetén a poszteriort a likelihood fogja dominálni, míg a prior hatása elhanyagolhatóvá válik.

A paraméterek a poszteriori eloszlása már a prediktív poszterior eloszlás kapcsán is megjelent az 1.5 egyenletben mint $p(\theta|D)$, $P(M_k|D)$ és $p(\theta_k|D,M_k)$ a D adat megfigyelése után (lásd [4]). Diszkrét modellek esetében a $p(M_k|D)$ poszterior felírható úgy, hogy

$$p(M_k|D) = \frac{p(D|M_k)p(M_k)}{p(D)} \tag{1.7}$$

ahol a *marginális modell likelihood* avagy M_k evidenciája

$$p(D|M_k) = \int p(D|\theta_k, M_k)p(\theta_k|M_k)\,\mathrm{d}\theta_k \tag{1.8}$$

és a *marginális adat likelihood* pedig

$$p(D) = \sum_k p(D|M_k)p(M_k). \tag{1.9}$$

A bayesi megközelítés elsődlegesen elméleti volta ellenére a bayesi elemzés célja gyakran a modellre vagy legalábbis a modell tulajdonságaira történő következtetés, ami a szubjektív értelmezés szerint a prior elvárások megfigyelések szerinti normatív „frissítését” jelenti.

1.3. Valószínűségi eloszlások függetlenségeinek rendszere

A Bayes-statisztikai keret nyitva hagyja a felhasználandó valószínűségi modellosztály kérdését, így akár a teljes tárgyterületi vagy akár egyetlen változó függésének a modellezése is lehetséges. A továbbiakban a teljes tárgyerületi modellezésnél legelterjedtebb valószínűségi gráfos modelleket, azon belül a Markov-hálókat, s különösen az oksági modellezésben is központi szerepet játszó Bayes-hálókat fogjuk áttekinteni. A feltételes modellek alkalmazhatóságát az Intelligens adatelemzés jegyzet PGM-ek tanulása című fejezetében foglaljuk össze.

A nagyszámú változó feletti együttes eloszlás hatékony reprezentálásánal kulcskérdés az eloszlásban megfigyelhető függetlenségek kihasználása, amihez pedig az explicit reprezentálásuk szükséges. A reprezentációs lehetőségek megértéséhez áttekintjük a függetlenségek rendszerét, s annak szabályszerűségeit.

1.3.1. A függetlenség és feltételes függetlenség fogalmai

A feltételes függetlenség fogalma központi szerepet játszik a valószínűségszámításban, s mint látni fogjuk a relevancia-irrelevencia kérdéskörének tisztázásában a logika, a mesterséges intelligencia, és a gépi tanulás terén is (ezt a valószínűségek szubjektív értelmezése is jelzi). Követve a Dawid [11] által bevezetett jelölést, diszkrét véletlen változók esetében a feltételes függetlenség a következőképpen definiálható.

1.3.1. Definíció. []Legyen $p(\underline{V})$ együttes eloszlás esetén $\underline{X}, \underline{Y}, \underline{Z} \subseteq \underline{V}$ diszjunkt részhalmazok. Jelölje $\underline{X}$ és $\underline{Y}$ $\underline{Z}$ feltétel melletti függetlenségét $I_p(\underline{X}|\underline{Z}|\underline{Y})$, azaz

$$(X \perp\!\!\!\perp Y|Z)_p \text{ \textit{iff} } (\forall \underline{x}, \underline{y}, \underline{z}\ p(\underline{x}, \underline{y}|\underline{z}) = p(\underline{x}|\underline{z})p(\underline{y}|\underline{z}) \text{ \textit{ha} } p(\underline{z}) > 0). \tag{1.10}$$

Egyéb ekvivalens definíciók még a következők is (feltéve, hogy a szükséges mennyiségek jól definiáltak) [31].

$$\begin{aligned}
(X \perp\!\!\!\perp Y|Z)_p &\Leftrightarrow p(\underline{x}|\underline{y}, \underline{z}) = p(\underline{x}|\underline{z}) \\
(X \perp\!\!\!\perp Y|Z)_p &\Leftrightarrow p(\underline{x}, \underline{y}, \underline{z}) = p(\underline{x}, \underline{z})p(\underline{y}, \underline{z})/p(\underline{z}) \\
(X \perp\!\!\!\perp Y|Z)_p &\Leftrightarrow p(\underline{x}, \underline{y}, \underline{z}) = h(\underline{x}, \underline{z})k(\underline{y}, \underline{z}) \textit{valamely h,k-ra} \\
(X \perp\!\!\!\perp Y|Z)_p &\Leftrightarrow p(\underline{x}, \underline{z}|\underline{y}) = p(\underline{x}|\underline{z})p(\underline{z}|\underline{y})
\end{aligned} \tag{1.11}$$

Az $(X \perp\!\!\!\perp Y|Z)_p$ feltételes függetlenségre egy másik jelölés az $I_p(X|Z|Y)$ és az $I_p(X;Y|Z)$. Egyértelműség esetén az alsóindexet és a feltételt elhagyjuk. A függetlenség hiányát, azaz a függést $(X \not\perp\!\!\!\perp Y|Z)_p$ jelöli. Érdemes észrevenni, hogy a feltételes függetlenség $\underline{Z}$ minden releváns értékére megkövetelt. A feltételes függetlenség gyengébb formája a kontextuális függetlenség, amely esetében a feltételes függetlenség csak adott $\underline{c}$ értékek esetén áll fenn egy diszjunkt $\underline{C}$ halmaznál. A *kontextuális feltételes függetlenséget* $\underline{X}$ és $\underline{Y}$ között $\underline{Z}$ feltétel mellett a c kontextusban $I_p(\underline{X}|\underline{Z}, \underline{c}|\underline{Y})$ jelöli, azaz amikor

$$I_p(\underline{X}|\underline{Z}, \underline{c}|\underline{Y}) \text{ \textit{iff} } (\forall \underline{x}, \underline{y}, \underline{z}\ p(\underline{y}|\underline{z}, \underline{c}, \underline{x}) = p(\underline{y}|\underline{z}, \underline{c}) \text{ \textit{ha} } p(\underline{z}, \underline{c}, \underline{x}) > 0). \tag{1.12}$$

Az X és Y közötti függés erősségének kvantitatív jelzésére nagyon sok asszociás mérték ismert. Egy igen általános standard mérték a (feltételes) *kölcsönös információ*

$$MI_p(X;Y|Z) = \mathrm{KL}(p(X,Y|Z)|p(X|Z)p(Y|Z)). \tag{1.13}$$

1.3.2. Egyéb valószínűségszámítási alapfogalmak

A függetlenség fogalmán túl a valószínűségszámítás elemi eszközkészletét fogjuk csak használni, amelynek összefoglalásaként ajánlható az MI Almanach [47]. Tételesen a következőkre lesz szükség.

1. Eseménytér. Elemi és összetett esemény, additivitás.

2. Együttes eloszlás. Diszkrét eseménytér esetén eloszlás táblázatmodell.
3. Vetítés/bővítés. Események feletti „ki-átlagolás/szummázás/integrálás”.
4. Véletlen változó. Várható érték, variancia, medián, módusz.
5. Feltételes valószínűség. Kétváltozós és általános eset.
6. Bayes szabály. Kétváltozós és általános eset, prior, posterior fogalma, "$\propto$" jelölés.
7. Láncszabály. Tetszőleges sorrend melletti alkalmazhatóság.
8. Naiv következtetés. Tetszőleges feltételes eloszlás származtatása az együttes eloszlásból, levezetés.
9. Egyenlőtlenségek. Markov, Csebisev, Cauchy-Schwarz, Jensen

1.3.3. A Markov-takaró, Markov-határ és közvetlen függés fogalmai

A feltételes függetlenség lehetővé teszi egy adott változó szempontjából irreleváns változók definiálását is, méghozzá többváltozós módon, illetve elégséges és szükséges szempontokból is.

1.1. definíció. *Egy változóhalmazt $MB_P(X_i)$-t X_i Markov-takaró-jónak nevezünk $P(X_1,\ldots,X_n)$ eloszlásban, ha $(X_i \perp\!\!\!\perp V \setminus MB(X_i)|MB(X_i))_P$ (egyértelműség esetén P nem jelölt). A minimális Markov-takarót Markov-határnak nevezzük és $MBo(X_i)_P$ jelöli.*

Ha a Markov-takaró egyértelműen létezik, akkor bevezethető egy szimmetrikus páronkénti reláció a Markov-határbeliségre [15]: $MBM(X_i,X_j)_P$ fennáll X_i és X_j között P-ben, ha

$$MBM(X_i,X_j)_P \leftrightarrow X_j \in MBo(X_i)_P \tag{1.14}$$

A Markov-határbeliségen belül definiálható egy szigorúbb kategória is, amelyet *közvetlen függésnek* nevezünk, ha minden diszjunkt $Z \subseteq V$ halmazra $(X \not\perp\!\!\!\perp Y|Z)$ fennáll (ebben az esetben a függés két változó között is létezik, amikor $Z =$, ami nem feltétlenül igaz a Markov-határbeli változópároknál).

1.3.4. A grafoid axiómák

Egy adott valószínűségi eloszlás esetén a feltételes függetlenségek eleget tesznek a következő tulajdonságoknak, amelyek a valószínűség szubjektív értelmezése esetén irrelevancia tulajdonságokként is olvashatók [11, 41].

a Szimmetria: Az irrelevancia szimmetrikus.

$$I_p(\underline{X};\underline{Y}|\underline{Z})\ iff\ I_p(\underline{Y};\underline{X}|\underline{Z})$$

b Dekompozíció: Irreleváns információ része is irreleváns.

$$I_p(\underline{X};\underline{Y}\cup\underline{W}|\underline{Z}) \Rightarrow\ I_p(\underline{X};\underline{Y}|\underline{Z})\ and\ I_p(\underline{X};\underline{W}|\underline{Z})$$

c Gyenge unió: Irreleváns információ irreleváns marad más irreleváns információ megismerése után is.

$$I_p(\underline{X};\underline{Y}\cup\underline{W}|\underline{Z}) \Rightarrow\ I_p(\underline{X};\underline{Y}|\underline{Z}\cup\underline{W})$$

d Összevonás: Irreleváns információ irreleváns marad más irreleváns információ elfelejtése után is.

$$I_p(\underline{X};\underline{Y}|\underline{Z})\ and\ I_p(\underline{X};\underline{W}|\underline{Z}\cup\underline{Y}) \Rightarrow\ I_p(\underline{X};\underline{Y}\cup\underline{W}|\underline{Z})$$

e Metszet : Szimmetrikus irrelevancia együttes irrelevanciát jelent, ha nincs más függés.

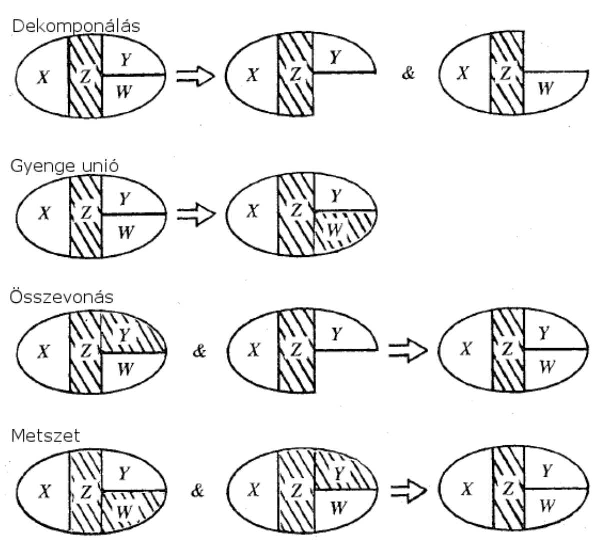
$$I_p(\underline{X};\underline{Y}|\underline{Z}\cup\underline{W})\ and\ I_p(\underline{X};\underline{W}|\underline{Z}\cup\underline{Y}) \Rightarrow\ I_p(\underline{X};\underline{Y}\cup\underline{W}|\underline{Z})$$

Ezek a tulajdonságok más területeken, az adatbázisok elméletében, illetve a gráfelméletben is megjelennek, ami indokolta az alábbi fogalmaj bevezetését:

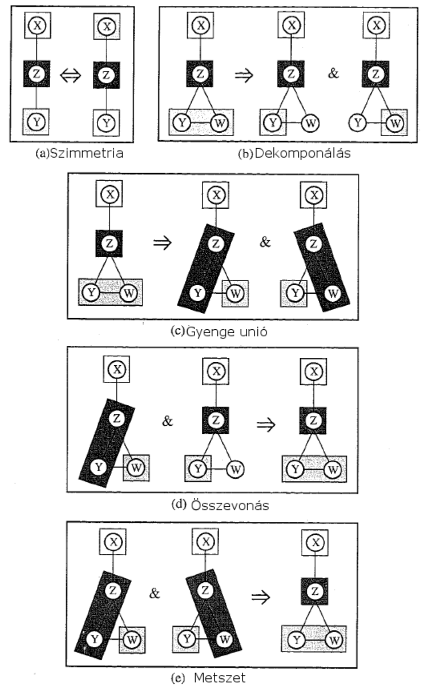
1. Szemi-grafoid (SG) axiómáknak nevezzük a szimmetria, dekompozíció, gyenge unió, összevonás tulajdonságokat, amelyek minden eloszlásban teljesülnek.

2. Grafoid axiómáknak nevezzük a szemi-grafoid axiómákat és a metszet tulajdonságot, amely csak szigorúan pozitív eloszlásokban áll fenn.

A tulajdonságokat az 1.1 ábra és az 1.2 ábra illusztrálják.



1.1. ábra. A szemi-grafoid axiómák vizualizációja.



1.2. ábra. A grafoid axiómák vizualizációja.

A tulajdonságok bevezetése felvetette, hogy esetleg lehetséges egy helyes és teljes logikai kalkulust létrehozni a függetlenségek feletti következtetésre. Elsőként vezessük be a függetlenségi modell fogalmát.

1.2. definíció. *Egy $P(X_1, \ldots, X_n)$ eloszlás M_P függetlenségi modellje pontosan a P-ben érvényes $I_P(X, Y|Y)$ függetlenségi állításokat tartalmazza.*

Sajnos azonban a Pearl&Paz által1985-ben megfogalmazott teljességi feltevéssel ellentétben a teljesség nem elérhető, mivel vannak ezen kívül is érvényes tulajdonságok, amelyeknek nincsen véges karakterizciójuk [52].

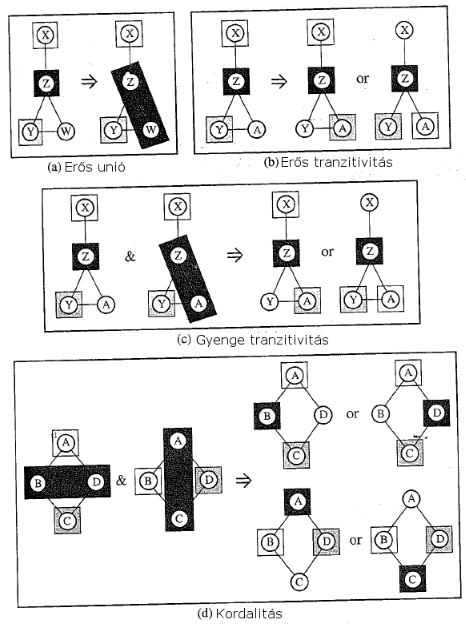
Függetlenségi térképek

Az eloszlások strukturális, sok esetben szinte egyedül lényeges tulajdonságainak, azaz a függetlenségi modelljének a reprezentálásához több megközelítés kínálkozik.

1.3. definíció. *Egy három argumentumú bináris függvényt $I(X, Y|Y) : V \times V \times V \to 0, 1$ $X, Y, Z \subseteq V$ P eloszlás*

1. *függetlenségi térképének nevezünk, ha $I \Rightarrow I_p$,*
2. *függési térképének nevezünk, ha $I \Leftarrow I_p$,*
3. *perfekt térképének nevezünk, ha $I \Leftrightarrow I_p$.*

A grafoid axiómák gráfokban való fennállása miatt a gráfok természetes jelöltek ilyen térkép szerepre, bár korlátaikat látni fogjuk (az 1.3 ábra nem-grafoid axiómákat illusztrál).



1.3. ábra. Nem-grafoid axiómák vizualizációja.

1.4. Valószínűségi gráfos modellek

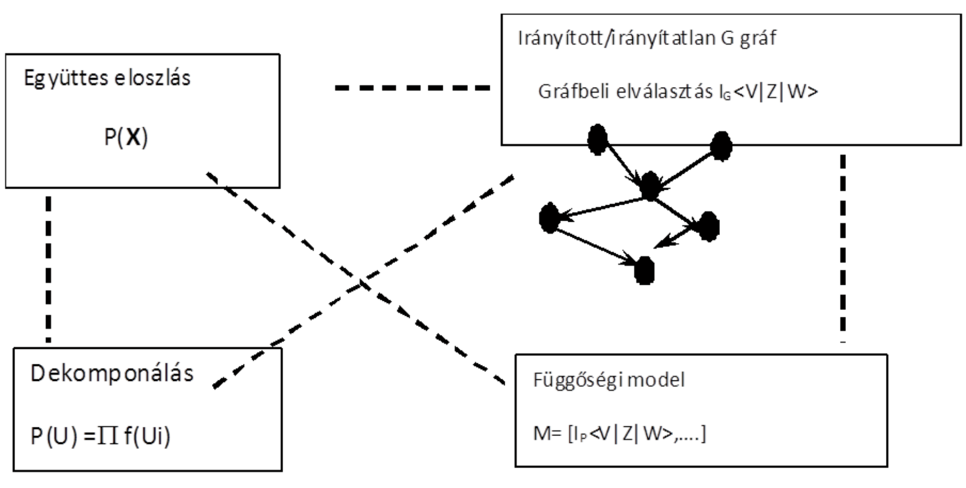
A gráfelmélet és valószínűségszámítás a véletlen gráfok elméletétől a mai hálózatkutatásig sok területen kapcsolódik egymáshoz. Mindkét kutatási irány foglalkozik a valószínűségi eloszlások reprezentálásával, s ez megnehezíti a terminológia választást. Az angol probabilistic graphical models (PGMs) kifejezés jelentése gráfos vagy gráf-alapú valószínűségi-modellek†, amit a meghonosodni látszó valószínűségi gráfos modellek fordítás remélhetőleg jól kifejez. Mint látni fogjuk, a gráfstruktúrák feletti eloszlások megadásánál a véletlen gráfok elmélete is szerephez jut, de az alapvető kapcsolatot a gráfok felhasználása, a valószínűségi modellek strukturális és parametrikus reprezentálása jelenti. A függetlenségi modell reprezentálásán túl további kérdés az eloszlás kvantitatív reprezentálása, a következtetésben való felhasználása és az oksági modellezésben való felhasználás kérdése. A jelen fejezetben az első két kérdést tárgyaljuk, a következtetés kérdését és az oksági modellek aspektusait a jegyzet két másik fejezete tárgyalja. A jelen fejezetben az oksági kutatásban betöltött szerepe miatt a Bayes-hálós modellosztály kap nagyobb hangsúlyt, de ebben a fejezetben a formális tárgyalás során a valószínűségi (akauzális) értelmezés keretein belül maradunk.

†Nem pedig véletlen gráfokon alapuló modellek jelentés.

1.4.1. Bayes-hálók kutatásának áttekintése

A Bayes-hálók (BN) a valószínűségi gráfos modellek egy alosztálya, amelyben irányított, körmentes gráfokat (DAG) használunk a sokváltozós eloszlás függetlenségeinek és kvantitatív jellemzőinek a reprezentálására, illetve opcionálisan az eloszlást generáló oksági mechanizmusok reprezentálására is. Egy intuitív értelmezés szerint a csomópontok a véletlen változókat, az élek pedig közvetlen oki ráhatást jelentenek, így definálva a modell struktúráját, amely egy irányított körmentes gráf. A struktúrát definiáló DAG lokális valószínűségi modellekkel van kiegészítve, annotálva, nevezetesen minden csomóponthoz tartozik egy lokális modell, amely megadja azon csomópont által reprezentált valószínűségi változó valószínűségi függését a gráfban szülőként jelen lévő valószínűségi változóktól. Ezen lokális modellek paraméterei a modell paraméterei.

A Bayes-háló modellosztályt több tudományterületen is sokoldalúan felhasználják. A teljesség igénye nélkül ide tartozik a tudásmérnökség, a gépi tanulás, az adat- és tudásfúzió, a biomarker kutatások vagy az oksági kutatások. A Bayes-hálók sokoldalúsága abból a tényből következik, hogy három autonóm kutatási szintet kapcsol egybe: az oksági modellt, a valószínűségi modell függetlenségi struktúráját és a kvantitatív eloszlást. Egy másik dimenzió mentén szintén három szerepet tölt be: a tudásreprezentálás, az adatokból történő tanulás és az intelligens döntéstámogatás problémaköreit fedi le. Ezen két dimenzió mentén létrejövő kombinációk kimeríthetetlen tárházat jelentenek, például az a priori ismeretek és megfigyelések, beavatkozások fúziójánál vagy az optimális szekvenciális beavatkozások megtervezésénél. A valószínűségi gráfos modellek és eloszlások akauzális relációinak sokféleségét az 1.4 ábra illusztrálja.



1.4. ábra. A valószínűségi gráfos modellek és eloszlások akauzális reláció.

A gráfos modellek kutatása valószínűségi és oksági modellezésben visszavezethető az 1920-as évekig, Wright útvonal-diagrammokat vizsgáló munkájáig [55]. Az első (orvosi) alkalmazása a Bayes-hálóknak, mint valószínűségi szakértői rendszereknek 1970-ben jelent meg, amely mind tudásmérnöki és gépi tanulási jegyeket is felmutatott [13]. Sokváltozós, nagyobb szakterületet lefedő alkalmazások az 1980-as évek végétől láttak napvilágot. Valószínűségi eloszlások függetlenségeinek szisztematikus vizsgálata 1979-ben publikálták [11], amit J.Pearl mérföldkőnek számító könyve követett a DAG-ok felhasználhatóságáról [41]. Az eloszlások dekomponálásának lehetősége annotált DAG-okkal 1982-ben merült fel először (a gráf alapú dekomponálás részletes tárgyalását lásd [31]).

A Bayes-hálók oksági felhasználása a kezdetektől jelen van [41, 53, 51], bár eleinte csupán emberi segédeszköznek tekintették és az okság valószínűségi alapú kutatása fő kérdésének a passzív megfigyelésből való tanulás határainak a tisztázását tartották (például a hatáserősség identifikálhatóságának kérdését [41, 39, 42]). Ezt később egészítette ki a függetlenségi modell mögötti oksági modell kutatása és a kontrafaktuálisok szemantikájának modell alapú definiálása [17, 43]. Hatékony következtetési eljárásokról szóló publikáció Bayes-hálók egy speciális osztályára, a polifákra 1983-ban jelent meg, amelyet 1988-ban követett egy általános esetben is használható, az egzakt következtetésben dominánssá váló megoldás, a klikkek fájában való következtetés [48]. A paraméterek bayesi kezelése rögzített struktúra esetén Dirichlet priorok felhasználásával 1990-ben jelent meg [50], a kapcsolódó „'prekvenciális' (predictive sequential,prequential) keret pedig 1993-ból származik [49]. A paraméterek átfogó, strukturális és kauzális aspektusait is figyelembe vevő megoldása 1995-ben született meg [24].

A struktúrák feletti bayesi megközelítés 1991-ig vezethető vissza, amely még feltette a változók oksági sorrendjének az ismeretét [6]. Az általános elméleti és gyakorlati keret 1992-ből származik [9]. Egy teljes bayesi megközelítést strukturális modelltulajdonságok következtetésére 1995-ben közöltek [37], amelyet 2000-ben adaptáltak nagyszámú változóra [15, 16].

Bayes-hálók dekomponált reprezentálása már az 1990-es évektől jelen van a szakirodalomban [19], bár kezdetben a kontextuális függetlenségek vezérelték ezt az irányt. A propozicionális reprezentációtól való elmozdulás a relációs és az általánosabb elsőrendű logika felé egy jelenleg is aktívan kutatott irány [23, 27, 29, 30].

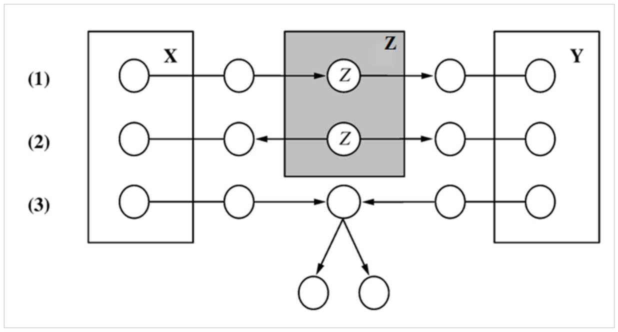
1.4.2. Irányított elválasztás, és egyéb gráfelméleti fogalmak

A függetlenségek tulajdonságainak gráfszerűsége és a függetlenségi térképre vonatkozóan megfogalmazott kívánalmak az 1.3 definícióban természetes módon vetik fel a gráfok felhasználását a feltételes függetlenségek reprezentálására. A lefogás/elválasztás reláció lehet az intuitív jelölt erre. Irányított gráfokban ennek a következő definícióját fogjuk használni.

1.4.1. Definíció. [] Egy G irányított, körmentes gráfban az $X, Y, Z \subseteq V$ diszjunkt csomópont halmazok esetében jelölje $I_G(X|Z|Y)$, illetve $I_G(X;Y|Z)$, ha X és Y *d-elválasztottak* Z által, azaz ha minden p út X és Y között blokkolt Z által a következőképpen

1,2 a p út tartalmaz egy Z-beli n csomópontot nem összetartó élekkel (azaz így $\rightarrow n \rightarrow$ vagy így $\leftarrow n \rightarrow$), 3 a p út tartalmaz egy nem Z-beli n csomópontot összetartó élekkel (azaz így $\rightarrow n \leftarrow$), amelynek nincs leszármazottja Z-ben.

Az 1.4.9 definíciót az 1.5 ábra illusztrálja.



1.5. ábra. Az irányított elválasztás (d-szeparáció) vizualizációja. Az X és Y csomóponthalmazok akkor d-szeparáltak, ha köztük minden út blokkolt vagy áthaladó (1) és széttartó (2) módon egy Z-beli csomóponttal, vagy konvergáló módon (3) Z-n kívüli elemmekkel (leszármazottak sem Z-beliek).

Az elválasztáson/lefogáson kívül szükséges gráfelméleti fogalmak még a következőek (lásd MI Almanach [47]):

- Irányítatlan, irányított, részlegesen irányított gráf fogalma
- Gráfreprezentációk, adjacencia mátrix
- Szülő, gyermek, ős, leszármazott, út, hurok, klikk
- Fa, polifa, többszörösön összekötött gráfok
- Topológiai (ősiségi) sorrendezés
- Kordális/háromszögesített gráfok [perfekt gráfok].

Az irányított Markov-feltételek

Egy Bayes-háló struktúrája és a reprezentálni kívánt eloszlás közti kapcsolatot az alábbi négy feltételre alapozhatjuk [41, 31, 10, 43].

1.4.2. Definíció. [] A $p(X_1,\ldots,X_n)$ eloszlás faktorizálható a G DAG szerint, ha

$$p(X_1,\ldots,X_n)=\prod_{i=1}^{n} p(X_i|\,\mathrm{Pa}(X_i)), \tag{1.15}$$

ahol $\mathrm{Pa}(X_i)$ az X_i csomópont szülői halmaza G-ben.

1.4.3. Definíció. [] A $p(X_1, \ldots, X_n)$ eloszlásra teljesül a sorrendi Markov-feltétel G szerint, ha

$$\forall\, i = 1, \ldots, n : (X_{\prec(i)} \perp\!\!\!\perp \{\{X_{\prec(1)}, \ldots X_{\prec(i-1)}\} \setminus \mathrm{Pa}(X_{\prec(i)})\} | \, \mathrm{Pa}(X_{\prec(i)}))_p, \qquad (1.16)$$

ahol $\prec$ egy topologikus sorrend G esetén (azaz az élek kompatibilisek G-vel) és $\{X_{\prec(1)}, \ldots X_{\prec(i-1)}\} \setminus \mathrm{Pa}(X_{\prec(i)})$ $X_{\prec(i)}$ összes ősét jelöli kivéve a szüleit.

1.4.4. Definíció. [] A $p(X_1, \ldots, X_n)$ eloszlásra teljesül a lokális (szülői) Markov-feltétel G szerint, ha bármely változó független a nem-leszármazottaitól feltéve a szüleit

$$\forall\, i = 1, \ldots, n : (X_i \perp\!\!\!\perp \mathrm{Nondescendants}(X_i) | \, \mathrm{Pa}(X_i))_p, \qquad (1.17)$$

ahol Nondescendants(X_i) jelöli X_i nem-leszármazottait G-ben (azaz akikhez nem vezet út G-ben X_i-től).

1.4.5. Definíció. [] A $p(X_1, \ldots, X_n)$ eloszlásra teljesül a globális Markov-feltétel G szerint, ha

$$\forall\, X, Y, Z \subseteq V : \; I_G(X; Y|Z)_G \Rightarrow (X \perp\!\!\!\perp Y|Z)_p. \qquad (1.18)$$

Ezen feltételek segítségével megfogalmazható egy alapvető kapcsolat eloszlások és DAG reprezentációjuk között [31].

1.4.1. Tétel. [[31]] Egy $p(V)$ eloszlás és G DAG esetén az 1.4.2, 1.4.3, 1.4.4 és 1.4.5 feltételek ekvivalensek:

(F) p Markovi G-hez avagy p faktorizálódik G szerint,

(O) p eleget tesz a sorrendi Markov-feltételnek G szerint,

(L) p eleget tesz a lokális Markov-feltételnek G szerint,

(G) p eleget tesz a globális Markov-feltételnek G szerint.

A feltételek ekvivalenciája miatt ezekre a feltételekre együttesen is mint irányított Markov-feltételekre hivatkozhatunk (p, G) pár viszonylatában.

1.4.3. Bayes-háló definíciók

A Markov-feltételek felhasználásával az alábbi meghatározás adható.

1.4.6. Definíció. [] A G irányított körmentes gráf a $P(V)$ eloszlás Bayes-hálója, ha minden változót a gráf egy csomópontja reprezentál, a gráfra teljesül valamelyik (és így az összes) Markov-feltétel, és a gráf minimális (azaz bármely él elhagyásával a Markov-feltétel már nem teljesül).

Míg ez a definíció egyértelműen a valószínűségi függetlenségek rendszerének reprezentációjaként tekint a Bayes-hálóra, addig a mérnöki gyakorlatban közkedvelt az alábbi, praktikus meghatározás.

1.4.7. Definíció. [] A V valószínűségi változók Bayes-hálója a (G, θ) páros,ha G egy irányított körmentes gráf, amelyben a csomópontok jelképezik V elemeit, θ pedig a csomópontokhoz tartozó $P(X_i|Pa(X_i))$ feltételes eloszlásokat leíró numerikus paraméterek összessége.

A Markov-feltétel teljesülése biztosítja, hogy minden gráfból kiolvasott függetlenség teljesüljön az eloszlásban, azonban a másik irányhoz, ahhoz tehát, hogy minden függetlenség kiolvasható is legyen a gráfból, annak stabilnak is kell lennie.

1.4.8. Definíció. [] Egy P(U) eloszlás stabil, ha létezik olyan G DAG, hogy P(U)-ban pontosan a G-ből d-szeparációval kiolvasható függések és függetlenségek teljesülnek benne (azaz G perfekt térkép).

A DAG-reprezentáció korlátját alapvetően az jelenti, hogy numerikusan a struktúra szerint nem szükségszerű függetlenségek is lekódolhatóak. A triviális redundanciákon túl ezek rejtett formákban is megjelenhetnek, például nem tranzitív függések képében vagy alacsonyabbrendű függetlenségek képében (például egy Markov-láncban megfelelő paraméterezés mellett előfordulhat, hogy a függések nem tranzitívak).

A d-szeparáció szükséges és elégséges voltát a következő tétel mutatja, amely szerint egy adott G DAG-gal kompatibilis összes eloszlásban érvényes függetlenségeknek a G-beli d-szeparáció egzakt reprezentációja [41].

1.4.2. Tétel. [[41]]

$$\forall\, X, Y, Z \subseteq V :\ (X \perp\!\!\!\perp Y|Z)_G \Leftrightarrow ((X \perp\!\!\!\perp Y|Z)_p \text{ in all } p \text{ Markov relative to } G.$$

A tétel élesítése, hogy általános bayesi megközelítésben azon eloszlások mértéke 0, amelyeknek G nem perfekt térképe [39].

Azonban bizonyos típusú függetlenségekhez, például a négy csomópontos gyémánt struktúrához, a Bayes-hálós megközelítés nem alkalmas, s ez a reprezentációs korlát felveti más gráftípusok használatát is.

1.4.4. Markov-hálók

Markov-hálók esetében egy G=(P,E) irányítatlan gráf alapján definiálunk egy függetlenségi térképet a G-beli (irányítatlan) elválasztásra/lefogásra alapozva (P a (csomó)pontok halmaza és E a P-be tartozó pontpárok, élek halmaza).

1.4.9. Definíció. [] Egy G irányítatlan gráfban az $X, Y, Z \subseteq V$ diszjunkt csomópont halmazok esetében jelölje $I_G(X|Z|Y)$, illetve $I_G(X;Y|Z)$, ha X és Y *elválasztottak* Z által, ha minden p út X és Y között tartalmaz egy Z-beli elemet.

Az irányított gráfokkal analóg módon irányítatlan gráfoknál is megfogalmazhatóak Markov-feltételek.

1.4.5. Markov-feltételek irányítatlan gráfokban

F Klikk faktorizálás: Ha $P(X_{1:n})$ G klikkjein definiált szorzatként felírható.

P Páronkénti markoviság: Bármely két nem szomszédos változó független egymástól az összes többi változóval vett feltétellel.

L Lokális markoviság: Egy változó független minden más változótól a szomszédaival vett feltétellel.

G Globálil markoviság: Bármely két változóhalmaz független egymástól egy őket elválasztó halmazzal vett feltétellel.

Az irányított gráfoktól eltérően ezek a feltételek általános esetben nem ekvivalensek, bár $G \Rightarrow L \Rightarrow P$. Kordális gráfokban szorosabb a kapcsolatuk, és pozitív eloszlásokban viszont ekvivalensek.

1.1. tétel (Hammersley-Clifford, 1971, unpublished)**.** *Pozitív eloszlásokban az irányítatlan Markov feltételek ekvivalensek: $F \Leftrightarrow P \Leftrightarrow L \Leftrightarrow G$.*

1.2. tétel. *Ha egy G kordális gráf lokálisan (vagy globálisan) markovi a p eloszláshoz, akkor az F tulajdonság is teljesül.*

Ennek következményeként is a Markov-háló definíciója így adható meg.

1.4. definíció. *G irányítatlan gráf a p eloszlás Markov-hálója (azaz G függetlenségi térképe p függetlenségi modelljének), ha G globálisan‡ markovi ($I_G(X|Z|Y) \Rightarrow I_P(X|Z|Y)$) és G minimális (azaz egy ;l törlésével ezt a tulajdonságát elvesztené).*

Pozitív eloszlásokban a Markov-hálót/Markov-(véletlen) mezőt Gibbs-(véletlen) mezőnek is nevezik, mivel a Hammersley–Clifford elmélet szerint a következő faktorizálható lehetséges.

1.5. definíció. *Egy eloszlás P_Φ Gibbs mérték, ha faktorokkal (potenciálokkal) definiálható $\Phi = \{\Phi_1(V_1), \ldots, \Phi_K(V_K)\}$:*

$$P_\Phi(V) = \frac{1}{Z} \prod_{i=1}^{K} \Phi_i(V_i), \tag{1.19}$$

ahol $V_1, \ldots, V_K \subset V$ és Z egy normalizációs konstans. Gibbs-(véletlen) mezőkhöz tartozó G esetében $V_1, \ldots, V_K$ pontosan G klikkjei.

Mivel a global markoviságból következik a lokális markoviság (p, G) esetében, X szomszédos csomópontjai G gráfban $bd(X, G)$ mindig X Markov-takarója $(MB_p(X, bd(X, G)))$. Továbbá pozitív eloszlásokban ez egy egyértelmű Markov-takaró.

‡Nem ekvivalens definíciókban a Lokális feltétel is használt.

1.3. tétel (Pearl,Paz, 1985). *Ha G egy Markov-hálója egy p pozitív eloszlásnak, akkor X szomszédos csomópontjai G gráfban egy egyértelmű Markov-határt formálnak.*

A tulajdonságok a tudásmérnökség és gépi tanulás számára is sokrétűen felhasználhatóak, például:

F Klikk faktorizálás: következtetésben.

P Páronkénti markoviság: élenkénti konstrukció esetében.

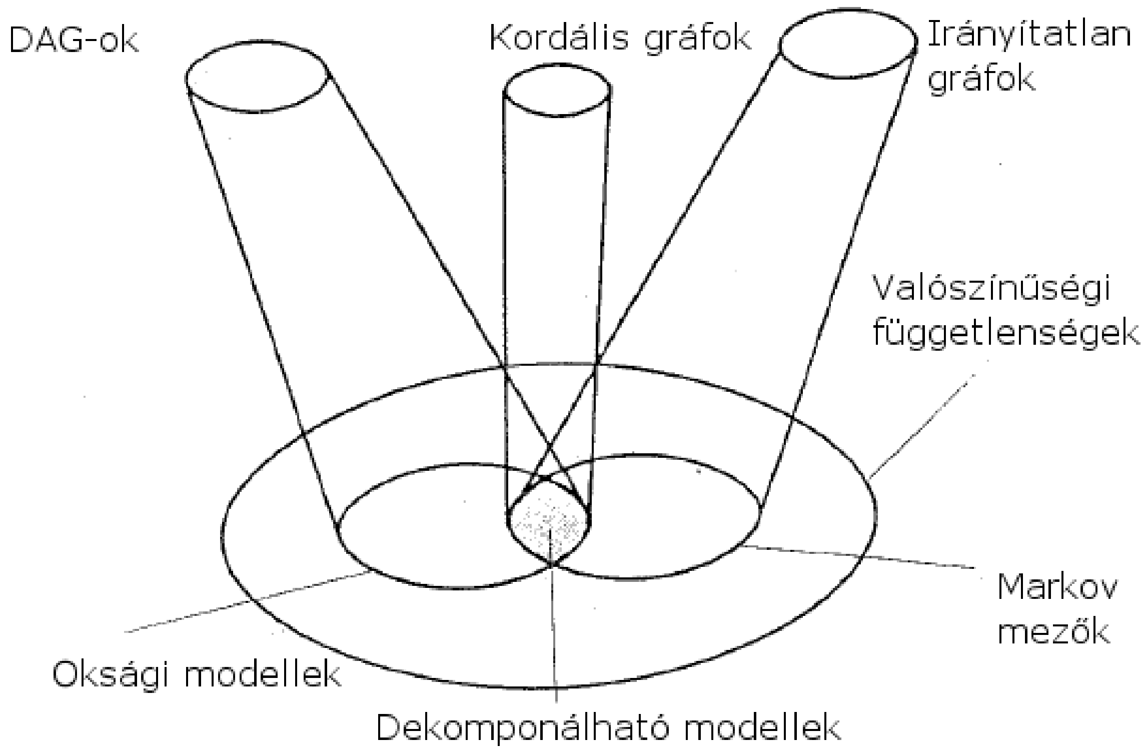
L Lokális markoviság: Markov-határ alapú konstrukcióban.

G Globális markoviság: globális relevencia viszonyok gyors kikövetkeztetésében (elválasztásra hatékony algoritmusok léteznek).

1.4.6. Bayes-hálók és Markov-hálók reprezentációs képessége

A Markov-hálók, a Bayes-hálókhoz hasonlóan helyes, de nem teljes reprezentációi a függetlenségeknek. Reprezentációs képességeik (avagy hiányosságaik) azonban bizonyos mértékig komplementerek. Például a gyémántstruktúrájú Markov-hálókkal egzakt módon reprezentált eloszlások Bayes-hálókkal egzakt módon nem reprezentálhatóak, viszont a Bayes-hálókban egzakt módon reprezentálható v-struktúra egzakt módon nem reprezentálható Markov-hálókkal. A két modellosztály reprezentációs képességeit az 1.6 ábra illusztrálja.

1.6. definíció (Intranzitív hármas/v-struktúra). *X, Y, Z véletlen változók egy Intranzitív hármast/v-struktúrát alkotnak p eloszlásbn, ha fennáll $D_p(X;Y)$, $D_p(Y;Z)$ és $I_p(X;Z)$.*



1.6. ábra. A különböző típusú gráfok reprezentációs képességei.

1.5. Egyszerű Bayes-hálók

A valószínűségi gráfos modellek strukturális képességeinek áttekintése után a modellek kvantitatív specifikálásának kérdéseit vizsgáljuk. Ennek bevezetéseként három egyszerű, de széleskörben használt modellosztály paraméterezését és felhasználását mutatjuk be.

1.5.1. Naiv Bayes-hálók

Tételezzük fel, hogy a változók halmaza tartalmaz egy hipotézisváltozót (Y) és megfigyeléseket ($X_1, \ldots, X_n$). (Elterjedt más elnevezések ugyanezen fogalmakra: ok, modell, diagnózis, illetve okozat, bizonyíték, megfigyelés, tünet, szimptóma.) A megfigyelések csoportját tekinthetjük különböző típusú megfigyeléseknek (például tüneteknek egy orvosi problémában) vagy azonos típusú, de eltérő idejű megfigyelések szekvenciáinak. A naiv Bayes-hálók definiáló tulajdonsága, hogy a $P(Y, X_1, \ldots, X_n)$ eloszlásban X_i-k feltételesen függetlenek Y feltétellel, azaz $I_P(X_j; X'|Y) = P(X_j|Y)$ bármely diszjunkt X' részhalmazra. A modell parametrikus megadását Y $p(Y)$ eloszlásának megadása és a $P(X_1|Y)$,...,$P(X_n|Y)$ feltételes eloszlások megadása jelenti. Ekkor

$$P(Y|X_{i_1}, \ldots, X_{i_m}) \propto P(X_{i_1}|Y) \ldots P(X_{i_m}|Y)P(Y). \tag{1.20}$$

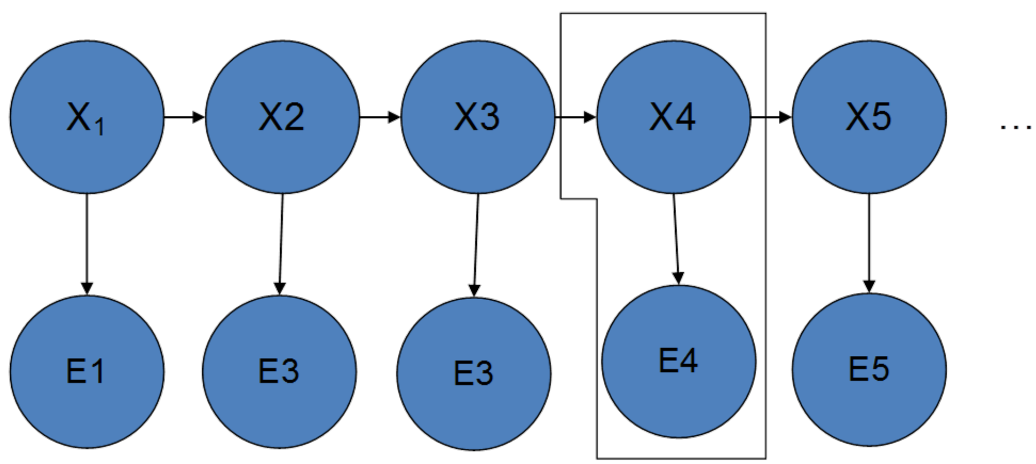
Ha Y bináris, akkor az esély így írható fel:

$$\frac{P(Y=1|X_{i_1},\ldots,X_{i_m})}{P(Y=0|X_{i_1},\ldots,X_{i_m})} = \frac{P(Y=1)}{P(Y=0)} \prod_i \frac{P(X_{i_1}|Y=1)}{P(X_{i_1}|Y=0)}. \tag{1.21}$$

Az eredmény jól tükrözi a valószínűségi gráfos modellek, nevezetesen itt a Bayes-hálók azon képességét, hogy az általános diszkrét, véges esetben exponenciális számosságú paramétert és következtetést lineáris számosságú paraméterre és lineáris számítási idejű következtetésre lehetett redukálni a függetlenségek kihasználásával.

1.5.2. Markov-láncok és rejtett Markov modellek

Elsőrendű Markov-láncok esetében feltételezzük, hogy az $X_1,\ldots,X_n$ valószínűségi változók körében minden i-re teljesül, hogy X_i feltételesen független az $X_1,\ldots,X_{i-2}$ változóktól az X_{i-1} ismeretében. Homogén lánc esetében továbbá az úgynevezett átmeneti valószínűségeket egy indexfüggetlen $P(X_i|X_{i-1})$ feltételes eloszlás definiálja. Ez a modell a rejtett Markov-modelleknél kiegészül a megfigyelhető evidenciák $E_1,\ldots,E_n$ valószínűségi változó halmazzal, amit – egy statikus érzékelő modellt feltételezve – egy szintén indexfüggetlen $P(E_i|X_i)$ köt a közvetlenül már nem megfigyelhetőnek vélelmezett „rejtett állapot” X_i változókhoz. A Markov-láncot és rejtett Markov modelleket az 1.7 ábra illusztrálja.



1.7. ábra. Markov-lánc és rejtett Markov modell illusztrálása.

A rejtett Markov-modellek esetében szintén hatékony, a változók számában lineáris és a változók értékkészletének méretében négyzetes futási idejű eljárások léteznek a következő tipikus feladatokra:

1. Szűrés (filtering) vagy ellenőrző megfigyelés (monitoring): ez a bizonyossági állapot (belief state) kiszámításának a feladata, ami a jelenlegi állapot feletti a posteriori eloszlás, az adott időpontig vett összes bizonyíték ismeretében; vagyis szeretnénk

kiszámítani a $P(X_t|e_{1:t})$ mennyiséget, feltéve, hogy a bizonyítékok folyamatos sorozatban érkeznek a $t = 1$ időponttól kezdve.

2. Előrejelzés (prediction): ez egy jövőbeli állapot feletti a posteriori eloszlás kiszámításának a feladata az adott időpontig vett összes bizonyíték ismeretében; azaz, szeretnénk kiszámítani a $P(X_{t+k}|e_{1:t})$ mennyiséget valamely $0 < k$ esetén.

3. Simítás (smoothing) vagy visszatekintés (hindsight): ez egy múltbeli állapot feletti a posteriori eloszlás kiszámításának a feladata a jelen időpontig vett összes bizonyíték ismeretében; azaz, szeretnénk kiszámítani a $P(X_k|e_{1:t})$ mennyiséget valamely $0 < k < t$ esetén.

4. Legvalószínűbb magyarázat (most likely explanation): a megfigyelések egy sorozatának ismeretében szeretnénk megtalálni azt az állapotsorozatot, amely a legvalószínűbben generálta az adott megfigyeléseket; vagyis szeretnénk kiszámítani az $argmax_{x_{1:t}} P(x_{1:t}|e_{1:t})$ értékét.

5. Likelihood paraméterezés: megfigyelések egy vagy több sorozatának ismeretében (x) szeretnénk megtalálni azt a paraméterezést (θ), amely az adott megfigyeléseket legnagyobb valószínűséggel generálta; azaz szeretnénk kiszámítani az $argmax_\theta P(x|\theta)$ értékét.

1.6. Parametrizáció, priorok definiálása és tudásmérnöki kérdések

A teljes bayesi megközelítésben a modelleket nem csupán egy pontparametrizációval kell ellátni, hanem a priori kényszereknek megfelelő struktúrák és paramétertér feletti eloszlások megadásával is. Ezek tárgyalása viszont az oksági aspektusok figyelembevételét is igényli, így tárgyalásuk a tudásmérnöki és kiterjesztett PGM reprezentációkhoz hasonlóan az Oksági modellek című fejezetben kapott helyet.

Irodalomjegyzék

[1] C. Andrieu, A. Doucet, and C. P. Robert. Computational advances for and from Bayesian analysis. *Statistical Science*, 19(1):118–127, 2004.

[2] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, 1980.

[3] J. O. Berger. Bayesian analysis: A look at today and thoughts of tomorrow. *Journal of the American Statistical Association*, 95(452):1269–1276, 2000.

[4] J. M. Bernardo. *Bayesian Theory.* Wiley & Sons, Chichester, 1995.

[5] C. M. Bishop. *Neural Networks for Pattern Recognition.* Clarendon Press, Oxford, 1995.

[6] W. L. Buntine. Theory refinement of Bayesian networks. In *Proc. of the 7th Conf. on Uncertainty in Artificial Intelligence (UAI-1991)*, pages 52–60. Morgan Kaufmann, 1991.

[7] P. Cheeseman. In defense of probability. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence (IJCAI-85)*, pages 1002–1009. Morgan Kaufmann, 1985.

[8] M. Chen, Q. Shao, and J. G. Ibrahim. *Monte Carlo Methods in Bayesian Computation.* Springer-Verlag, New York, 2000.

[9] G. F. Cooper and E. Herskovits. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9:309–347, 1992.

[10] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter. *Probabilistic networks and expert systems.* Springer-Verlag, New York, 1999.

[11] A. P. Dawid. Conditional independence in statistitical theory. *J. of the Royal Statistical Soc. Ser.B*, 41:1–31, 1979.

[12] A. P. Dawid. Probability, causality and the empirical world: A bayes-de finetti-popper-borel synthesis. *Statistical Science*, 19(1):44–57, 2004.

[13] F. T. de Dombal, D. J. Leaper, J. C. Horrocks, and J. R. Staniland. Human and computer-aided diagnosis of abdominal pain. *British Medical Journal*, 1:376–380, 1974.

[14] B. Efron. Bayes theorem in the 21st century. *Proc. Natl. Acad. Sci.*, 340(7):1177–78, 2013.

[15] N. Friedman and D. Koller. Being Bayesian about network structure. In *Proc. of the 16th Conf. on Uncertainty in Artificial Intelligence(UAI-2000)*, pages 201–211. Morgan Kaufmann, 2000.

[16] N. Friedman and D. Koller. Being Bayesian about network structure. *Machine Learning*, 50:95–125, 2003.

[17] D. Galles and J. Pearl. Axioms of causal relevance. *Artificial Intelligence*, 97(1-2):9–43, 1997.

[18] D. Gamerman. *Markov Chain Monte Carlo*. Chapman & Hall, London, 1997.

[19] D. Geiger and D. Heckerman. Knowledge representation and inference in similarity networks and Bayesian multinets. *Artificial Intelligence*, 82:45–74, 1996.

[20] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman & Hall, London, 1995.

[21] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London, 1996.

[22] L. Győrfi. *T omegkiszolgálás informatikai rendszerekben*. Műegyetemi Kiadó, 1996. (in Hungarian).

[23] J. Y. Halpern. An analysis of first-order logics of probability. *Artificial Intelligence*, 46:311–350, 1990.

[24] D. Heckerman, D. Geiger, and D. Chickering. Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20:197–243, 1995.

[25] J. A Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky. Bayesian model averaging: A tutorial. *Statistical Science*, 14(4):382–417, 1999.

[26] L. Hunyadi. Bayes gondolkodás a statisztikában. *Statisztikai szemle*, 89(10-11):1150–1171, 2011.

[27] Manfred Jaeger. Relational Bayesian networks. *Proc. of the 13th Conference on Uncertainty in Artificial Intelligence (UAI-1997)*, pages 266–273, 1997.

[28] W. H. Jefferys and J. O. Berger. Sharpening ockham's razor on a Bayesian strop, 1991.

[29] D. Koller and A. Pfeffer. Object-oriented Bayesian networks. In Dan Geiger and Prakash P. Shenoy, editors, *Proc. of the 13th Conf. on Uncertainty in Artificial Intelligence (UAI-1997)*, pages 302–313. Morgan Kaufmann, 1997.

[30] D. Koller and A. Pfeffer. Probabilistic frame-based systems. In *Proc. of the 15th National Conference on Artificial Intelligence (AAAI), Madison, Wisconsin*, pages 580–587, 1998.

[31] S. L. Lauritzen. *Graphical Models*. Oxford, UK, Clarendon, 1996.

[32] J. S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer-Verlag, 2004.

[33] D. J. C. MacKay. Probable networks and plausible predictions - a review of practical Bayesian methods for supervised neural networks. *Neural Computation*, pages 469–505, 1996.

[34] D. J. C. Mackay. *Learning in graphical models*, chapter Introduction to Monte Carlo Methods. MIT Press, Cambridge, MA, 1999.

[35] D. Madigan and R. Almond. Test selection strategies for belief networks. StatSci Research Report 20., 1993.

[36] D. Madigan, S. A. Andersson, M. Perlman, and C. T. Volinsky. Bayesian model averaging and model selection for Markov equivalence classes of acyclic digraphs. *Comm.Statist. Theory Methods*, 25:2493–2520, 1996.

[37] D. Madigan and J.York. Bayesian graphical models for discrete data. *Internat. Statist. Rev.*, 63:215–232, 1995.

[38] D. Malakoff. Bayes offers a 'new' way to make sense of numbers. *Science*, 286:1460–1464, 1999.

[39] C. Meek. Causal inference and causal explanation with background knowledge. In *Proc. of the 11th Conf. on Uncertainty in Artificial Intelligence (UAI-1995)*, pages 403–410. Morgan Kaufmann, 1995.

[40] R. M. Neal. *Bayesian Learning for Neural Networks*. Springer, Berlin, 1996.

[41] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Francisco, CA, 1988.

[42] J. Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–710, 1995.

[43] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2000.

[44] K. R. Popper. *The Logic of Scientific Discovery*. Hutchinson, London, 1959.

[45] A. Rényi. *Probability Theory.* Akadémiai Kiadó, Budapest, 1970.

[46] C. P. Robert. *The Bayesian Choice.* Springer-Verlag, New York, 2001.

[47] S. Russel and P. Norvig. *Artificial Intelligence.* Prentice Hall, 2001.

[48] D. J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, Series B*, 50(2):157–224, 1988.

[49] D. J. Spiegelhalter, A. Dawid, S. Lauritzen, and R. Cowell. Bayesian analysis in expert systems. *Statistical Science*, 8(3):219–283, 1993.

[50] D. J. Spiegelhalter and S. L. Lauritzen. Sequential updating of conditional probabilities on directed acyclic graphical structures. *Networks*, 20(.):579–605, 1990.

[51] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search.* MIT Press, 2001.

[52] M. Studeny. Semigraphoids and structures of probabilistic conditional independence. *Annals of Mathematics and Artificial Intelligence*, 21(1):71–98, 1997.

[53] T. Verma and J. Pearl. *Causal Networks: Semantics and Expressiveness*, volume 4, pages 69–76. Elsevier, 1988.

[54] G. Vita'nyi and K. Li. *An introduction to Kolmogorov complexity and it applicationes.* Springer-Verlag, New York, 1990.

[55] S. Wright. Correlation and causation. *J. of Agricultural Research*, 20:557–585, 1921.

