

Tõenäosuste leidmine Bayesi võrkudes

Sven Laur
swen@math.ut.ee

Andmekaevandamise uurimisseminar MTAT.03.169.
Arvutiteaduse instituut, Tartu Ülikool
Detsember 2003, lk. 112–135

6. jaanuar 2004. a.

Kokkuvõte

Üheks kaasaegse matemaatika paeluvamaks haruks on Bayesi statistika ning selle analüüsimeetodid. Keskkel kohal nii teoorias kui ka praktikas on erinevate mudelite — tõenäosusjaotuste — konstrueerimine. Mudel peab adekvaatselt kajastama meie teadmisi uuritavast objektist. Üks suhteliselt lihtne, kuid samas väga efektiivne metodoloogia on Bayesi võrgud. Meetodi tuumaks on põhjuslike seoste kajastamine orienteeritud graafina ning selle hilisem analüüs.

Bayesi võrke on edukalt kasutatud nii ekspertsüsteemides kui masinõppes, sest struktureeritud lähenemine lubab lihtsalt kirjeldada eelteadmisi ning erinevaid tõenäosuslikke mudeleid. Kui ekspertsüsteemid kirjeldavad vaid apriorset tõenäosusjaotust, siis treenitavad Bayesi võrgud on võimelised näidetest õppima. Kahjuks peame oma ülevaates piirduma vaid tõenäosusjaotuse määramise ning arvutamise seotud probleemidega, jättes käsitlemata treenimise ning sellega seotud probleemid.

Ehkki ehituselt on Bayesi ja neurovõrgud sarnased, siis sellega sarnasus piirdub — Bayesi võrk kirjeldab alati tõenäosusjaotust, samas kui neurovõrgud kannavad endas suvalisi funktsioone.

Esmalt anname lühiülevaate Bayesi statistika alustest ning kirjeldame Bayesi võrgu semantikat. Seejärel keskendume praktikas üliolulisele keskmistamise probleemile, mille üheks lahenduseks on klikipuu ning klikkidele vastavate tõenäosusjaotuste leidmine.

1 Saateks

Ülevaade käsitleb eelkõige Bayesi võrku kui kogutud teadmuse mudelit. Sealjuures ignoreerime täielikult teadmiste kogumise probleemi. Teadmuse kogumiseks on kaks alternatiivset teed: ekspertide põhjalik intervjuerimine ning automaatselt õppimine näidetest. Eri lähenemised täiendavad teineteist. Juhtudel, kus ekspert ei suuda oma otsuseid ratsionaalselt põhjendada, on loomulik Bayesi võrgu treenimine. Samas nõuab treenimine ilma eelteadmisteta paljudel juhtudel liiga palju näited ning seetõttu on ekspertarvamus hädavajalik.

Lisaks teadmuse kirjeldamisele võib Bayesi võrku kasutada ka nähtuste modelleerimiseks, kuigi piir nende eesmärkide vahel on hägune. Teadmuste kirjeldus

on reeglina staatiline, kuid mudeleid kasutatakse eelkõige õppimiseks, sealjuures valitakse paljude erinevalt seadistatud Bayesi võrkude vahel. Nii staatilisel kui dünaamilisel juhul on tarvis efektiivselt arvutada erinevaid tõenäosusjao-tusi, seetõttu on ülevaate läbivaks probleemiks vastavate meetodite käsitlemine. Näiliselt lihtsate ekspertsüsteemide nagu *Microsoft Office Assistant* toimimise taga on keeruline teooria. Olgu siinjuures kurioosse kõrvalpõikena mainitud, et Bayesi võrkudel on tulevikku ka Bill Gatesi arvates [Hel96].

Materjali koostamisel lähtusime eelkõige küsimusest miks ja alles seejärel küsimusest kuidas. See mõneti ebaharilik lähenemine toob kaasa detailide rohkuse ning lugejal on kasulik hoida silme ees üldist ülesehitust. Töö jaguneb neljaks peamiseks osaks.

- **Tõenäosus kui ebakindluse mõõt.** Et mõista Bayesi võrgu semantikat on oluline käsitleda tõenäosust kui ebakindluse mõõtu. Seetõttu alustame lühikesest filosoofilisest ning praktilisest sissejuhatausest Bayesi statistikas.
- **Põhjuslikud seosed ning sõltumatus.** Kuigi Bayesi võrgu kirjeldus on ime-lihtne, siis selle peidetud struktuuri kasutamine nõuab teoreetilisi tulemusi. See sunnib vaatama lähemalt tinglikku sõltumatust ning selle abstraktsiooni — sõltumatuste graafi.
- **Bayesi võrgu teisendamine.** Lihtsate graafiteoreetiliste operatsioonidega saab avada Bayesi võrgu peidetud struktuuri, millel põhinevad edasised arvutused. Toome sisse Bayesi võrgu moraali ning klikipuu mõisted.
- **Ühisjaotuste arvutamine.** Esmalt kirjeldame tulemusi, mis lihtsustavad keerukaid ühisjaotusi. Seejärel toome ära konkreetse arvutuskeemi.

Ülevaade ei kata kaugeltki kõiki Bayesi võrgu täpse analüüsiga seotud küsimusi. Kindlasti tuleks eraldi käsitleda konkreetsete alamülesannete lahendamiseks kasutatavaid algoritme. Teine palju suurem ning kahtlemata olulisem valdkond on Bayesi võrkude treenimine, alates otsesest meetoditest lõpetades EM-algoritmi ja Monte-Carlo simulatsioonimeetoditega. Kuid iga selline teema nõuab omaette ülevaadet.

2 Tõenäosus kui ebakindluse mõõt

Üheks keskseks filosoofiliseks küsimuseks kaasaegses matemaatikas on tõenäosu-se mõiste interpreteerimine. On küllalt raske vastata küsimusele, miks on kulli ja kirja tõenäosus $1/2$ ning mida see ikkagi tähendab. Aksiomaatiline tõenäo-susteooria kujutab endast mehhaanilist arvutusreeglite kogu, kus tõenäosuste väärtused ning nende leidmine on teooriaväline. Tõenäosuse tõlgendamisel ja-gunevad arvamused kahte leeri. Objektivse tõenäosuse pooldajad (*frequentists*) eeldavad, et maailmas leidub igal huvitaval sündmusel üheselt määratud tõe-näosus ning see ilmneb katset lõpmatult korrates. Teise äärmuse järgi pole ob-jektiivset tõenäosust olemas, on olemas vaid üksikisiku veendumusi kajastav subjektiivne ebakindlus.

A priori pole selge, et ebakindlus (*uncertainty*) peaks rahuldama tõenäosu-sele esitatavaid aksioome. Loomulikult pole seda võimalik ka üldjuhul tõestada. Seetõttu tuuakse sisse ratsionaalne uskumus (*belief*), mis idealiseerib inim-mõis-tust. Erinevatest aksiomaatikatest lihtsamini mõistetav on 1937 aastal de Finetti

poolt pakutud *Dutch book* argument — ratsionaalselt mõtlev indiviid ei tohi aktsepteerida kihlvedusid, mis on talle kahjulikud iga lõpptulemuse korral.

Näiteks võib fikseerida järgmised reeglid. Alice ja Bob võivad omavahel müüa veksleid. Iga veksel on seotud sündmusega, mille toimumisel peab väljaandja maksma ühe krooni. Kui sündmust ei toimu, siis on veksel kehtetu. Veksleid võib müüa, osta ja vahetada.

Kui kõigi lõpptulemuste hulk on Ω , siis mõõduteoorias tuntud mõõtuvate sündmuste hulka \mathcal{F} saab käsitleda kui lubatud kihlveosündmuste hulka. Alice'i ebakindluse mõõduks p võime lugeda vekslite minimaalse müügi- või maksimaalse ostuhinna (on loomulik lugeda mõlemad hinnad võrdseks). Üsna lihtne on tõestada omadused:

(P0) sündmuse ebakindlus $p(A)$ ei sõltu selle esitamise viisist;

(P1) sündmuse ja vastandsündmuse ebakindluste summa on konstantne

$$p(A) + p(A^c) = \text{const};$$

(P2) teineteist välistavate sündmuste ebakindlus on aditiivne

$$A \cap B = \emptyset \quad \Rightarrow \quad p(A \cup B) = p(A) + p(B).$$

Ühte ja sama sündmust saab kirjeldada mitmel moel. Näiteks kulli saamine mündiviskel on samaväärne väitega, et katsel ei saadud kirja. Seetõttu kindlustab omadus (P0) ebakindluse ühesuse. Selle tõestamiseks vaatame olukorda, kus sündmused A_1 ja A_2 toimuvad alati samaaegselt. Kui $p(A_1) < p(A_2)$, siis saab Bob Alice'ile müüa vekslit $V_B(A_2)$ hinnaga $p(A_2)$ ning osta Alice'lt vekslit $V_A(A_1)$ hinnaga $p(A_1)$. Kuna mõlemad sündmused toimuvad alati korraga, siis pole veksleid vaja välja lunastada, samas teenib Bob alati $p(A_2) - p(A_1)$ krooni puhast kasu. Teiste kihlveoskeemide väljamõtlemise jätame lugejale iseseisvaks ülesandeks. Muuseas on oluline eeldada, et \mathcal{F} on σ -algebra.

Omadused (P0)-(P2) kindlustavad, et p on korrektselt defineeritud ja aditiivne mõõt, kus kindla sündmuse tõenäosus $P(\Omega) = \text{const}$. Et oluline on vaid erinevatele sündmuste ebakindluste vaherkord, siis võime ebakindluse normeerida 0 ja 1 vahele¹. Niisiis kui sündmuste ühendhulk Ω on lõplik, on ebakindlus p tõenäosusmõõt.

Kui sündmuste hulk Ω on lõpmatu, siis peab ebakindlus olema ülalt pidev:

(P3) iga sündmuste jada $A_1 \subseteq A_2 \subseteq \dots \subseteq A_n \subseteq \dots$ korral

$$\lim_{n \rightarrow \infty} p(A_i) = p(A), \quad A = A_1 \cup A_2 \cup \dots \cup A_n \cup \dots$$

Teisisõnu peab Alice oma otsustes olema järjepidev. Võrduse vasak pool kirjeldab olukorda, kus Alice täiendab vekslit lisaklauslitega ning parem pool vekslit, milles on kõik tingimused korraga üles loetud. On loomulik et mõlemad protsessid viivad sama lõpphinnani ehk pole vahet kui kaua võtab aega vekslite tingimuste täpsustamine.

Loomulikult sõltub ratsionaalse Alice ebakindlus tema eelnevatest teadmistest. Pole ju mõtet panustada kirjale teades, et mündil on kaks kulli. Ebakindlust sündmuse A suhtes, kui on toimunud (teada) sündmus B tähistatakse $p(A|B)$. Näiteks $p(\text{päike}|\text{vihm})$ tähistab päikesepaiste tõenäosust (ebakindlust), kui on

¹ Tegelikult peab meie defineeritud mudelis $p(\Omega) = 1$.

teada vaid see, et väljas sajab vihma. Aksiomaatilises tõenäosusteoorias vastab sellele tinglik tõenäosus, mis on defineeritud võrduse $p(B)p(A|B) = p(A, B)$ abil. Õnneks kehtib ebakindluse korral lisaks vormilisele ka sisuline sarnasus, mida saab tõestada kasutades näiteks kihlveoskeeme.

Teoreem 1 (Bayesi teoreem). Iga sündmuste paari A ja B korral kehtib

$$p(A)p(B|A) = p(A, B) = p(B)p(A|B).$$

Tulemus seab olulise kitsenduse ratsionaalsele uskumusele ning võimaldab formaliseerida ratsionaalse järeldamise protsessi. Reeglina ei õnnestu teadusliku eksperimendi käigus leida otsitava suuruse täpset väärtust. Olgu X otsitav suurus ning D eksperimendi käigus mõõdetud suurused, siis saame tulemuseks tõenäosusjaotuse

$$p(X = x|D) = \frac{p(D|X = x)p(X = x)}{p(D)} \propto p(D|X = x)p(X = x).$$

Valemi vasak pool kirjeldab Alice'i arvamust suuruse X väärtusest peale eksperimenti ning paremal olev liige $p(X = x)$ Alice'i veendumust, et otsitav suurus on x enne katset. Liiget $p(D|X = x)$ nimetatakse tõepäraks (*likelihood*). Võrduses esinev märk \propto tähistab seda, et vasak ning parem pool erinevad teineteisest konstandi võrra — on proportsionaalsed X varieerimisel.

Kuna erinevatel indiviididel on erinevad uskumused, siis järeldused katsest sõltuvad meie eelteadmistest! Vastolu objektiivsusega on näiline — teaduslik eksperiment peab olema laialdaselt aktsepteeritav. Ehk teisisõnu lõppjäreldused peavad olema sarnased sõltumata isiklikest eelteadmistest.

Klassikaline statistika ei tunnista aprioorsust tõenäosust $p(X = x)$ ning käsitleb vaid tõepära $p(D|X = x)$. Fikseerides $p(X = x) = \text{const}$ langeb Bayesi statistika kokku klassikalise statistikaga, seetõttu on selle rakendusvaldkond veidi laiem. Teiseks erinevuseks on analüüsi lõpptulemus. Klassikaline statistika annab reeglina tulemuseks punkt- või vahemikhinnanguid nagu keskmine ja veapiirid. Bayesi statistika ainus lubatav resultaat on tõenäosusjaotus ning igasugused järeldused väljuvad teooria piirest. Seetõttu on Bayesi statistika loomulik kaaslane otsustusteooria (*decision theory*).

Praktika seisukohalt olulisimaks erinevuseks on efektiivsus, reeglina viib järjekindel Bayesi printsiibi rakendamine arvutuslikult keerukate probleemideni, mille lahendamine on osutunud võimalikus vaid Monte-Carlo simulatsioonimeetoditega. Edasise huvi rahuldamiseks soovitame Bayesi statistikat käsitlevaid monograafiad [BT92, GCS⁺03].

3 Sõltumatud katsed ning keskmistamine

Kontseptuaalselt pole Bayesi statistika midagi muud kui järjekindel Bayesi teoreemi rakendamine. Tulemuseks on alati tõenäosusjaotus, kusjuures on oluline vaid selle kuju mitte jaotuse poolt kaetud pindala. Tihti loobutakse seetõttu jaotuse normaliseerimisest, kuna konstantsete kordajate hülgamine lihtsustab oluliselt avaldisi.

Järgnevalt vaatleme üht lihtsaimat katseskeemi — sõltumatute katsete seeriat. Vaatleme näite korras kulli ja kirja viskamist. Olgu katseseerias saadud

tulemused kodeeritud jadana (d_1, \dots, d_n) , kus $d_i \in \{0, 1\}$. Katse mudeliks loeme, et iga i korral $p(d_i = 1) = q$, kus q on meile tundmatu parameeter. Kuna katsed on sõltumatud, siis saame

$$p(d_1, \dots, d_n | q) = \overbrace{q \cdots q}^k \underbrace{(1-q) \cdots (1-q)}_{n-k} = q^k (1-q)^{n-k},$$

kus k on ühtede (kullide) arv katsetes. Ülesande täielikuks lahendamiseks peame määrama uskumist kajastava eeljaotuse². Näiteks võime postuleerida, et q on ühtlaselt võetud lõigust $[0, 1]$, siis saame lõpptulemuseks jaotuse

$$p(q | d_1, \dots, d_n) \propto q^k (1-q)^{n-k},$$

Samas kui meil on kindlaid andmeid $q \in [0.75, 1]$, saame lõppjaotuseks

$$p(q | d_1, \dots, d_n) \propto \begin{cases} q^k (1-q)^{n-k}, & \text{kui } q \in [0.75, 1], \\ 0, & \text{kui } q \in [0, 0.75). \end{cases}$$

Järgnev joonis 1 näitab vastavaid tõenäosusjaotusi, kui $n = 10$, $k = 5$ ja $n = 100$, $k = 50$. Jooniselt ilmneb eelduse $q \in [0.75, 1]$ ja katseseeria omavaheline vastolu — kui eeljaotuse väärtus on suur, siis on tõepära nullilähedane. Tekib loomulik küsimus, kas revideerida antud tulemuste valguses eeldust $q \in [0.75, 1]$ või mitte. Kui eeljaotus peegeldab meie “tõelisi” uskumusi, siis teooria garanteerib lõpptulemuse korrektsuse — tulemuse muutmine oleks irratsionaalne ning võimaldaks Bobil meilt raha välja petta. Tegelikuses pole küsimusel lihtsat vastust, sest väga raske kui mitte võimatu on leida meie uskumusi kajastavat eeljaotust. Teisalt võib kergekäeline eeljaotuse muutmine “ilustada” tulemusi ning anda võimaluse enesepettuseks. Seetõttu on adekvaatset eeljaotuse probleem üks tähtsamaid nii Bayesi teoorias kui ka praktikas.

Õnneks kehtib üldine kooskõla tulemus. Kui tegelikkusele vastava mudeli \mathcal{M}_0 ümbruses on eeltõenäosus nullist erinev, siis sõltumatute katsete arvu kasvades läheneb $p(\mathcal{M}_0) \rightarrow 1$, kus koondumine toimub tõenäosuse järgi. Maakeeli tähendab see, et kui me ei välista koheselt õiget tulemust, siis piisava arvu katsete järel koondub tõenäosusmass tegelikkusele vastava mudeli ümbrusesse, välja arvatud erandlikud juhud, mida juhtub üliharva. Näiteks ei saa kuidagi välistada juhtu, et katseseeria koosneb vaid kullidest, isegi kui q tegelik väärtus on 0.5.

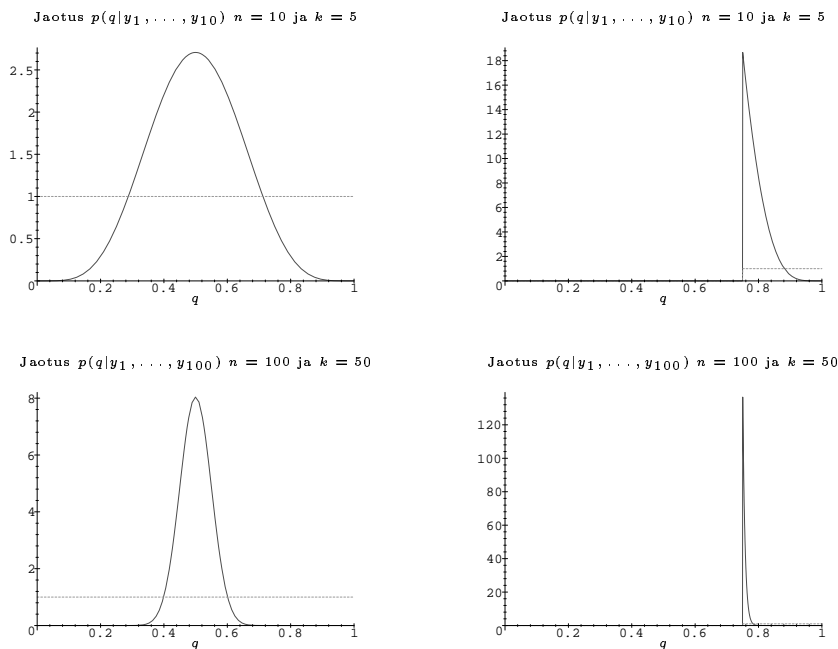
Tulles tagasi kulli ja kirja viskamise juurde võime loomulikult küsida, milline on kulli saamise tõenäosus järgmisel katsel. Sealjuures peame arvestama nii registreeritud katsed kui eeljaotust. Otseseselt ei oska me vastavat tõenäosust avaldada, küll aga suudame leida aposterioorse ühisjaotuse

$$p(y_{n+1} = 1, q | y_1, \dots, y_n) = p(y_{n+1} = 1 | q) p(q | y_1, \dots, y_n) = q \cdot q^k (1-q)^{n-k} p(q).$$

Kuna meid huvitab vaid y_{n+1} väärtus, siis peame keskmistama tulemuse üle kõigi q väärtuste

$$p(y_{n+1} = 1 | y_1, \dots, y_n) = \int_0^1 p(y_{n+1} = 1 | q) p(q | y_1, \dots, y_n) dq.$$

²Kuna matemaatiliselt korrektne nimetus aprioorne tõenäosusjaotus on kohmakas, siis siin ja edaspidi kasutame eeljaotust.



Joonis 1: Kulli saamise aposterioorne tõenäosus erinevate eeljaotuste ning katseandmete korral. Vasakpoolsed graafikud vastavad ühtlasele eeljaotusele $[0, 1]$ parempoolsed ühtlasele eeljaotusele $[0.75, 1]$. Katkendlik joon vastab eeljaotusele ning pidev aposterioorsele jaotusele.

Leides vastavad integraalid, saame katseandmete ning esimese mudeli korral kulli tõenäosuseks $p(y_{11} = 1|y_1, \dots, y_{10}) = p(y_{101} = 1|y_1, \dots, y_{100}) = 1/2$. Teise mudeli korral on see märgatavalt suurem $p(y_{11} = 1|y_1, \dots, y_{10}) \approx 0.792$ ja $p(y_{101} = 1|y_1, \dots, y_{100}) \approx 0.757$, sest meie aprioorne uskumus oli tugevalt kalutatud kulli poole.

Enamasti tuleb mudeli tõenäosusjaotust $p(\mathcal{M}|D)$ keskmistada üle üleliigsete parameetrite, mis viib kas integraalini

$$p(x|D) = \int_{-\infty}^{\infty} p(x, y|D) dy = \int_{-\infty}^{\infty} p(D|x, y)p(x, y) dy$$

või summani (kõdunud integraalini)

$$p(x|D) = \sum_{i=1}^m p(x, y_i|D) = \sum_{i=1}^m p(D|x, y_i)p(x, y_i),$$

kui üleliigne parameeter on diskreetne.

Keskmistamine ongi Bayesi statistika Pandora laekaks. Ühelt poolt lubab keskmistamine lahendada kõik ette tulevad probleemid, teisalt muutuvad arvutused keerukaks ning tulemused pole sageli analüütilised. Samas hoiab eeljaotuse kasutamine tihti ära üleõppimise (*overfitting*), sest kaalutud keskmise korral lähevad kõik mudelid tulemuse ennustamisel arvesse.

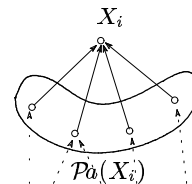
4 Põhjuslikud seosed

Bayesi statistika üheks raskemaks probleemiks on teadmiste formaliseerimine. Vähegi keerukamatel juhtudel on raske või koguni võimatu leida tõenäosusjaotust, mis kirjeldaks meie uskumisi adekvaatselt. Milline on näiteks teie subjektiivne uskumus, et lähima kümne aasta jooksul tabab Maad tuumasõda? Või kui suur on tõenäosus, et te võidate ümbermaailmareisi 123 käiguga?

Ümbermaailmareisi korral saab vastuse leidmiseks teha mängust mudeli — kirjeldada kõikvõimalikke nuppude paigutusi ning erinevate käikude tõenäosusi. Midagi analoogset saab teha subjektiivsete tõenäosuste korral. Reeglina suudavad inimesed küllaltki edukalt leida ning hinnata põhjuslikke seoseid ning nende tugevust. Tuumasõja puhul tuleb lihtsalt leida erinevad stsenaariumid ning neid käivitavate sündmuste tõenäosus. Peamiseks analüüsi vahendiks on keeruka süsteemi dekomponeerimine arusaadavateks põhjuslikeks seosteks. Loomulikult ei lähe selline protsess alati valutult. Erinevad otsustusteooria (*decision analysis*) koolkonnad on avaldanud mitmeid põhjalikke uurimusi, kuidas hinnata objektiivselt subjektiivset tõenäosust [SvH75, Ker91] ning milliseid psühholoogilisi eripärasid tuleb selle juures arvestada [Wan98, Wil94].

Üheks sündmuste omavahelist struktuuri graafiliselt illustreerivaks mudeliks on Bayesi võrk. Selle loomulik graafiline märgikeel võimaldab ka matemaatika-kaugel eksperdil kirjeldada oma intuiitviseid teadmisi. Selleks piisab põhjuslike seoste kirjeldamisest ning nende varustamisest tõenäosustega.

Definitsioon 1. Bayesi võrk (*belief network*) on tsükliteta orienteeritud graaf, mis määrab ära juhuslike suuruste $\mathcal{X} = (X_1, X_2, \dots, X_n)$ ühisjaotuse. Graafi tippudeks on juhuslikud suurused X_1, X_2, \dots, X_n ning kaared kirjeldavad põhjuslikke seoseid tippude vahel. Iga juhusliku suuruse X_i ja selle eellaste $\mathcal{Pa}(X_i)$ korral on määratud tinglik tõenäosus $p(X_i|\mathcal{Pa}(X_i))$.

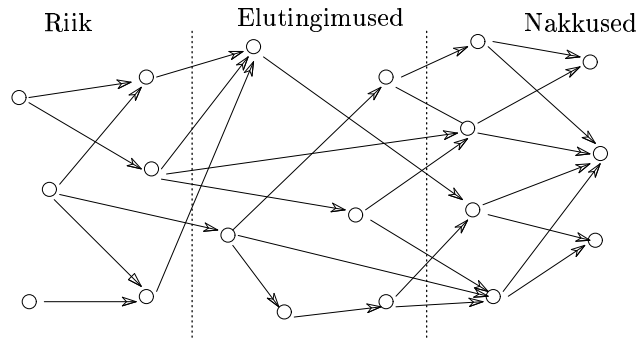


Klassikaline Bayesi võrkudel põhinevate ekspertüsteemide rakendusvaldkond on meditsiin, tuntumad neist on *pathfinder* [HHN92b, HHN92a] ja *intellipath*³. Teiseks näiteks toome programmi *Microsoft Office Assistant* [HBH⁺98], mille analüüsivõime põhineb Bayesi võrkudel. Enamasti ei piirduta vaid ekspertide teadmistega, vaid püütakse vajalike tõenäosusi täpsustada treeningandmete põhjal, kuid selle käsitlemine väljub antud materjali piiridest.

Juhuslikud suurused saab jagada kahte klassi, meid huvitavad parameetrid ning vähemtähtsad suurused — hüperparameetrid. Viimaste väärtus meid otseselt ei huvita, kuid nende olemasolu muudab mudeli kirjeldamise lihtsamaks. Näiteks pangalaenu tagasimaksmine sõltub inimese ametialasest edukusest lähima 5 – 10 aasta jooksul, kuid seda ei saa otselt mõõta. Küll aga saab seda ennustada teiste teadaolevate andmete põhjal. Kuna meile on oluline vaid laenu tagasimaksmine mitte ametialane edukus, siis peame keskmistama üle erinevate võimaluste. Keskmistamine üle hüperparameetrite muudabki arvutused Bayesi võrgus tehniliselt töömahukaks.

Paljudel juhtudel on erinevad tegurid grupeerunud ning seetõttu omandab mudel hierarhilise kuju. Näiteks nakkushaiguste levik sõltub väga paljudest teguritest, kuid laias laastus määrab selle arstiabi kvaliteet, elutingimused ning kultuurilised eripärad. Reeglina on tingimused ühe riigi piires sarnased. Mudeli

³*Intellipath* on *pathfinder*-i kommertsiaalne kloon.



Joonis 2: Hierarhilised kihid Bayesi võrgus

saab lühidalt kokku võtta hierarhiana

$$riik \rightarrow elutingimused \rightarrow nakkused.$$

Esimese lähendusena on see piisav, kuid loomulikult saab mudelit muuta palju detailsemaks. Lisades ning täpsustades erinevaid kvantitatiivseid tegureid, mis kirjeldavad arstiabi, elukvaliteeti ning väliskeskonna mõju, saame tulemusena keerulise Bayesi võrgu. Mõnedel juhtudel puudub osadel tippudel kindel tähendus, me ei suuda neid interpreteerida hoomatavate mõistetena nagu vanus, keskmine sissetulek jms.

Keerukamad mudelid on täpsemad, kuid ühes täpsusega kasvab ka hüperparameetrite arv. Ehkki suhteliselt lihtne on leida vastavat ühisjaotust, on tunduvalt keerulisem leida meid huvitavate parameetrite tõenäosusjaotust — analüütiline lahend võib üldse puududa. Teiseks muudab hüperparameetrite arv keskmistamise väga töömahukaks.

Diskreetsete jaotuste korral on lahend alati leitav, kuid parameetrite arvust tingitud arvutusmaht on siiski aukartustäratav. Näiteks juba üle 20 binaarse hüperparameetri korral on võimalike väärtustuste hulk üle miljoni ning seetõttu on oluline leida efektiivne keskmistamise viis. Küsimus taandub ühisjaotuse faktoriseerimisele — selle esitamisele mitme väiksema jaotise korrutisena.

5 Tinglik sõltumatus ja selle omadused

Bayesi võrgu koostamisel ning selle analüüsis on olulisel kohal tinglik sõltumatus, mis lubab analüüsida keerukaid seoseid erinevate juhuslike suuruste vahel. Kõik efektiivsed keskmistamise meetodid põhinevad sõltumatusel.

Definitsioon 2. Juhuslike suuruste hulgad A ja B on tinglikult sõltumatud juhuslike suuruste hulga C suhtes, kui teades juhuslike suuruste hulga C väärtusi, on suurused hulkades A ja B omavahel sõltumatud. Lühidalt tähistame tinglikku sõltumatust $I(A, B|C)$ ja tinglikku sõltuvust $D(A, B|C)$.

Näiteks palavik ja kõha on tugevas korrelatsioonis, olles külmetushaiguste sümptomiks. Kuid teades, kas patsiendil on kurguvalu või mitte, on mõlemad sümptomid teineteisest sõltumatud. Antud materjali piires tegeleme diskreetsete või absoluutselt pidevate juhuslike suurustega, millel tõenäosusjaotuse määrab

täielikult tihedusfuntsioon⁴. Seetõttu on juhuslike suuruste hulgad A ja B tinglikult sõltumatud parajasti siis kui iga juhuslike suuruste väärtustuse \mathcal{A} , \mathcal{B} ja \mathcal{C} korral lahutub tihedus

$$p(\mathcal{A}, \mathcal{B}|\mathcal{C}) = p(\mathcal{A}|\mathcal{C})p(\mathcal{B}|\mathcal{C}) \iff p(\mathcal{A}|\mathcal{B}, \mathcal{C}) = p(\mathcal{A}|\mathcal{C}).$$

Siinkohal on paslik meenutada, et diskreetsel juhul on tiheduseks elementaar-sündmuste tõenäosus $p(\mathcal{A}) = \Pr[A = \mathcal{A}]$.

Mõneti üllatavalt võib tinglik sõltumatus minna kaotsi keskmistamisel ja edasisel täpsustamisel. Selle illustreerimiseks vaatame kolme binaarset juhusliku suurust a , b ja c . Suurus c on võetud ühtlase jaotusega $p(c = 1) = 1/2$ ning a ja b ühisjaotus on esitatud tabelina. Esimeses näites on $I(a, b|c)$, kuid a ja b ühisjaotus ei ole sõltumatu. Tuletame meelde, et sõltumatuse korral on tabeli element rea- ja veerusumma korrutis.

$$\frac{1}{2} \cdot \begin{array}{c|cc} & c = 0 & \\ \hline a \backslash b & 0 & 1 \\ \hline 0 & 1 & 0 \\ 1 & 0 & 0 \end{array} + \frac{1}{2} \cdot \begin{array}{c|cc} & c = 1 & \\ \hline a \backslash b & 0 & 1 \\ \hline 0 & 1/4 & 1/4 \\ 1 & 1/4 & 1/4 \end{array} = \begin{array}{c|cc} & c = ? & \\ \hline a \backslash b & 0 & 1 \\ \hline 0 & 5/8 & 1/8 \\ 1 & 1/8 & 1/8 \end{array}$$

Teises näites keskmistub sõltuvus $D(a, b|c)$ välja ning a ja b on eraldi võttes sõltumatud.

$$\frac{1}{2} \cdot \begin{array}{c|cc} & c = 0 & \\ \hline a \backslash b & 0 & 1 \\ \hline 0 & 0 & 1/2 \\ 1 & 1/2 & 0 \end{array} + \frac{1}{2} \cdot \begin{array}{c|cc} & c = 1 & \\ \hline a \backslash b & 0 & 1 \\ \hline 0 & 1/2 & 0 \\ 1 & 0 & 1/2 \end{array} = \begin{array}{c|cc} & c = ? & \\ \hline a \backslash b & 0 & 1 \\ \hline 0 & 1/4 & 1/4 \\ 1 & 1/4 & 1/4 \end{array}$$

Intuiitiivselt tähendab sõltumatus $I(a, b|c)$ seda, et a väärtus ei anna uut lisainformatsiooni b kohta, kui teame c väärtust. Seda silmas pidades pole sõltumatuse kadumine keskmistamisel üllatav, lisainformatsiooni c puudumisel võib a teadmine paljastada b väärtuse. Teine näide illustreerib tuntud fakti, et erinevate andmete kombineerimisel võib nende tõestusväärtus suurenedada.

Bayesi võrku on kodeeritud kahte tüüpi informatsiooni: põhjuslikud seosed kaartena ning nende kvantitatiivne kirjeldus tinglike tõenäosustena. Kaared eri tippude vahel kannavad endas lokaalset sõltuvuste kirjeldust ning seetõttu saab graafi analüüsimiseks kasutada vaid tingliku sõltumatuse nelja põhiomadust:

(I1) sümmeetria (*symmetry*) — hulkade järjestus ei mõjuta sõltumatust

$$I(A, B|C) \iff I(B, A|C);$$

(I2) lahutatavus (*decomposition*) — alamhulgad pärivad sõltumatuse

$$I(A, B_1 \cup B_2|C) \implies I(A, B_1|C);$$

(I3) nõrk laiendatavus (*weak union*) — sõltumatus säilib liikmete fikseerimisel

$$I(A, B_1 \cup B_2|C) \implies I(A, B_1|C \cup B_2);$$

(I4) taanduvus (*contraction*) — sõltumatus kandub ahelas edasi

$$I(A, B|C_1 \cup C_2) \wedge I(A, C_2|C_1) \implies I(A, B \cup C_2|C_1).$$

⁴Tiheduse mõiste kasutamine ühtlustab antud kontekstis tähiseid ning on teoreetiliselt igati põhjendatud, vaata raamatut [Bil95, lk. 213, 419–425].

Esimene omadus tuleneb otseselt definitsiooni sümmeetriast. Intuitiivselt taandilt vaadates on ka lahutatavus ilmne, informatsiooni kasutamata jättes ei saa meie teadmiste hulk suureneda. Nõrk laiendatavus postuleerib fakti, et pole oluline kas me saame informatsiooni korruga $C \cup B_2$ või tükide C ja B_2 kaupa. Ning taanduvus ütleb, et kui suurused A ei anna mingit informatsiooni suuruste B kohta teades suurusi $C_1 \cup C_2$ ning suuruste C_2 kõik väärtused on võrdvõimalikud teades C_1 ja A , siis pole suurustest C_2 arutlustes kasu.

Teise omaduse formaalseks tõestamiseks oletame vastuväiteliselt, et leiduvad väärtustused \mathcal{A} , \mathcal{B}_1 ja \mathcal{C} nii, et $p(\mathcal{A}, \mathcal{B}_1 | \mathcal{C}) \neq p(\mathcal{A} | \mathcal{C})p(\mathcal{B}_1 | \mathcal{C})$. Võttes hulga \mathcal{B}_2 juhuslike suuruste B_2 kõikvõimalikud väärtustused on tulemuseks vastuolu tingliku sõltumatusega

$$p(\mathcal{A}, \mathcal{B}_1 \times \mathcal{B}_2 | \mathcal{C}) = p(\mathcal{A}, \mathcal{B}_1 | \mathcal{C}) \neq p(\mathcal{A} | \mathcal{C})p(\mathcal{B}_1 | \mathcal{C}) = p(\mathcal{A} | \mathcal{C})p(\mathcal{B}_1 \times \mathcal{B}_2 | \mathcal{C}).$$

Kolmanda omaduse tõestamiseks piisab väärtuste $p(\mathcal{B}_2 | \mathcal{C}) \neq 0$ vaatlemisest, sest vastasel korral pole tinglikud tõenäosused määratud. Siit saame

$$p(\mathcal{A}, \mathcal{B}_1 | \mathcal{C}, \mathcal{B}_2) = \frac{p(\mathcal{A}, \mathcal{B}_1, \mathcal{B}_2 | \mathcal{C})}{p(\mathcal{B}_2 | \mathcal{C})} = p(\mathcal{A} | \mathcal{C})p(\mathcal{B}_1 | \mathcal{B}_2, \mathcal{C}) = p(\mathcal{A} | \mathcal{C}, \mathcal{B}_2)p(\mathcal{B}_1 | \mathcal{C}, \mathcal{B}_2),$$

sest lahutatavuse tõttu $p(\mathcal{A} | \mathcal{C}) = p(\mathcal{A} | \mathcal{C}, \mathcal{B}_2)$. Neljanda omaduse tõestus on ilmne

$$p(\mathcal{A}, \mathcal{B}, \mathcal{C}_2 | \mathcal{C}_1) = p(\mathcal{B}, \mathcal{C}_2 | \mathcal{C}_1)p(\mathcal{A} | \mathcal{B}, \mathcal{C}_2, \mathcal{C}_1) = p(\mathcal{B}, \mathcal{C}_2 | \mathcal{C}_1)p(\mathcal{A} | \mathcal{C}_1).$$

Lisaks eelpool toodud tingimustele on olemas kolm olulist lisaomadust, mida täidavad vaid osad jaotused:

(I5) lõikemomadus (*intersection*) — nõrk laiendatavus on pööratav

$$I(\mathcal{A}, \mathcal{B}_1 | \mathcal{C} \cup \mathcal{B}_2) \wedge I(\mathcal{A}, \mathcal{B}_2 | \mathcal{C} \cup \mathcal{B}_1) \iff I(\mathcal{A}, \mathcal{B}_1 \cup \mathcal{B}_2 | \mathcal{C});$$

(I6) tugev laiendatavus (*strong union*) — täpsustamine säilitab sõltumatuse

$$I(\mathcal{A}, \mathcal{B} | \mathcal{C}) \implies I(\mathcal{A}, \mathcal{B} | \mathcal{C} \cup \mathcal{D});$$

(I7) tugev transitiivsus (*strong transitivity*) — tipp d ei sõltu mitmest hulgast

$$I(\mathcal{A}, \mathcal{B} | \mathcal{C}) \iff I(\mathcal{A}, \{d\} | \mathcal{C}) \vee I(\mathcal{B}, \{d\} | \mathcal{C}).$$

Tugev laiendatavus ja transitiivsus on väga spetsiifilised, võib julgelt öelda, et enamuse jaotusi pole tugevalt laiendatavad ega transitiivsed. Kuid lõiketingimuse kehtimiseks piisab, kui tihedusfunktsioon on igas punktis positiivne. Kui iga väärtustuse \mathcal{B}_1 ja \mathcal{B}_2 korral $p(\mathcal{B}_1, \mathcal{B}_2 | \mathcal{C}) \neq 0$, siis saame

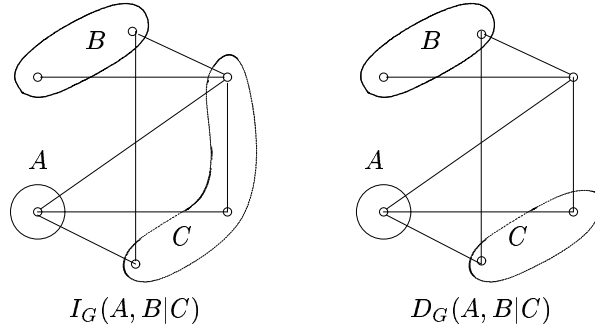
$$p(\mathcal{B}_1, \mathcal{B}_2 | \mathcal{C})p(\mathcal{A} | \mathcal{B}_1, \mathcal{C}) = p(\mathcal{A}, \mathcal{B}_1, \mathcal{B}_2 | \mathcal{C}) = p(\mathcal{B}_1, \mathcal{B}_2 | \mathcal{C})p(\mathcal{A} | \mathcal{B}_2, \mathcal{C}),$$

millest $p(\mathcal{A} | \mathcal{B}_1, \mathcal{C}) \equiv p(\mathcal{A} | \mathcal{B}_2, \mathcal{C})$. Teisisõnu ei sõltu avaldis väärtustustest \mathcal{B}_1 ja \mathcal{B}_2 ning seega

$$p(\mathcal{A} | \mathcal{B}_1, \mathcal{B}_2, \mathcal{C}) = p(\mathcal{A} | \mathcal{B}_1, \mathcal{C}) = p(\mathcal{A} | \mathcal{C}).$$

Formaalseks tõestuseks paneme tähele, et

$$p(\mathcal{A} | \mathcal{C}) = \int p(\mathcal{A}, \mathcal{B}_2 | \mathcal{C}) d\mathcal{B}_2 = \int p(\mathcal{B}_2 | \mathcal{C}) p(\mathcal{A} | \mathcal{B}_1, \mathcal{C}) d\mathcal{B}_2 = p(\mathcal{A} | \mathcal{B}_1, \mathcal{C}).$$



Joonis 3: Tipuhulkade sõltuvus ning sõltumatus graafis

6 Sõltumatuste kirjeldamine graafina

Kuna sõltumatus on sümmeetriline omadus, siis on loomulik kasutada selle kirjeldamiseks orienteerimata graafe. Samas põhineb Bayesi võrk orienteeritud graafil. Siit lähtuvalt on tarvis lahendada kaks ülesannet: anda tingliku sõltumatuste kirjeldus Bayesi võrgus ning leida sobiv lihtgraaf, mis kirjeldab sõltumatust. Alustame lihtsamast — sõltumatuse kirjeldamisest lihtgraafina. Järgnevas on kesksel kohal tipuhulkade kolmiku sõltumatus graafis.

Definitsioon 3. Olgu A , B ja C orienteerimata graafi G tipuhulgad. Tipuhulk C eraldab omavahel tipuhulgad A ja B , kui iga tee tippude $a \in A$ ja $b \in B$ vahel läbib mõnda tippu hulgast C . Sellist tipuhulkade kolmikut nimetakse sõltumatuks ja tähistakse $I_G(A, B|C)$. Vastasel korral nimetakse kolmikut sõltuvaks ja tähistakse $D_G(A, B|C)$.

Olemuselt on tinglik sõltumatus kvalitatiivne ning ei sõltu konkreetsetest tõenäosuste arväärtustest Bayesi võrgus, vaid nende omavahelisest vahekorrast. Seega on loomulik võtta selle mudeliks lihtgraaf tipuhulgaga \mathcal{X} .

Definitsioon 4. Graafi G nimetatakse juhuslike suurustele \mathcal{X} vastavaks täpseks sõltuvuste graafiks (*perfect-map*), kui tipukolmik A, B, C on sõltumatu parajasti siis, kui vastavad juhuslikud suurused on tinglikult sõltumatud, see tähendab

$$I_G(A, B|C) \iff I(A, B|C).$$

Kahjuks pole igal Bayesi võrgul (ühisjaotusel) täpset sõltumatuste graafi. Joonisel 4 toodud kolm binaarset juhuslikku suurust ei oma täpset sõltuvuste graafi. Sest ühelt poolt peaks tippude a ja b vahel olema serv, kuna $D(a, b|c)$, kuid serva lisamine läheb vastuollu a ja b sõltumatusega. Täpseid sõltuvuste graafe omavatel jaotustel on olemas lihtne kirjeldus.

Teoreem 2. *Juhuslike suuruste \mathcal{X} ühisjaotusel on olemas täpne sõltuvuste graaf parajasti siis, kui jaotus rahuldab eelpool sõnastatud tingimusi (I1)-(I7).*

TÕESTUS. Tõestuse võib leida Gutierrez'e ja Hadi õpikust [CGH97]. □

Kuna graafis kehtib alati tugev laiendatavus, siis on omadus (I6) vältimatu. Tugev transitiivsus (I7) hoiab ära joonisel 4 kujutatud vastuolud. Sisuliselt välistab tugev lõiketingimus kõik Bayesi võrgud, milles mitu kaart suubuvad ühte

tippu. Et enamus jaotusi ei rahulda nii rangeid nõudeid, siis tuleb defineerida ohutud aproksimatsioonid.

Definitsioon 5. Graaf G on juhuslike suurustele \mathcal{X} vastav sõltumatuste graaf (I -map), kui igale sõltumatule kolmikule $I_G(A, B|C)$ vastavad juhuslikud suurused on tinglikult sõltumatud, see tähendab

$$I_G(A, B|C) \implies I(A, B|C).$$

Definitsioon 6. Graaf G on juhuslike suurustele \mathcal{X} vastav sõltuvuste graaf (D -map), kui igale sõltuvale kolmikule $D_G(A, B|C)$ vastavad juhuslikud suurused pole tinglikult sõltumatud, see tähendab

$$D_G(A, B|C) \implies D(A, B|C).$$

Kuna täisgraafis pole ühtegi mittetriviaalset sõltumatut kolmikut ja nullgraafis pole sõltuvusi, siis leiduvad igal ühisjaotusel alati ohutud aproksimatsioonid. Ülimaks eesmärgiks on leida minimaalne sõltumatuste graaf ning maksimaalne sõltuvuste graaf. Seda realiseerivat algoritmi vaatame järgmises peatükis koos tingliku sõltuvuse kirjeldusega.

7 Bayesi võrgu moraal

Bayesi võrgu moraaliks võetakse võimalikult täpne sõltumatuste graaf. Esimese sammuna leiame minimaalse sõltumatuste graafi kirjelduse. Selleks vaatame kolme erineva rangusega Markovi tingimust.

Definitsioon 7. Graaf tippudega \mathcal{X} rahuldab ühisjaotuse $p(\mathcal{X})$ suhtes:

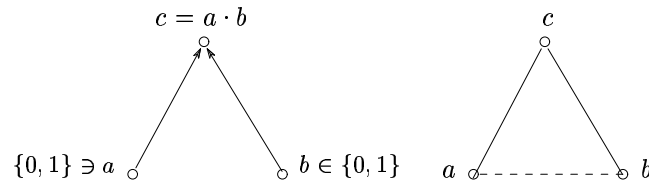
(PM) Markovi paaritingimust (*pairwise markov property*), kui puuduv serv tippude X_i ja X_j vahel viitab X_i ja X_j tinglikule sõltumatusele $\mathcal{X} \setminus \{X_i, X_j\}$ suhtes ehk $I_G(X_i, X_j | \mathcal{X} \setminus \{X_i, X_j\}) \implies I(X_i, X_j | \mathcal{X} \setminus \{X_i, X_j\})$;

(LM) lokaalset Markovi tingimust (*local markov property*), kui X_i ja $\mathcal{X} \setminus \{X_i\}$ on tinglikult sõltumatud tipu X_i naabritehulga $Bd(X_i)$ suhtes ehk

$$I_G(X_i, \mathcal{X} \setminus \{X_i\} | Bd(X_i)) \implies I(X_i, \mathcal{X} \setminus \{X_i\} | Bd(X_i));$$

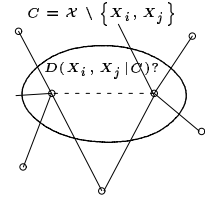
(GM) globaalset Markovi tingimust (*global markov property*) kui graaf on sõltuvuste graaf ehk $I_G(A, B|C) \implies I(A, B|C)$.

Siinjuures tuletame lugejale meelde, et X_i naabriteks on tipud, millel on tipuga X_i ühine serv. Neist omadustest on kergesti kontrollitav vaid Markovi



Joonis 4: Näide Bayesi võrgust, millel pole täpset sõltumatuste graafi

paaritingimus, teiste kontrollimine on suuremate graafide korral raske, kui mitte võimatu. Ilmselgelt kehtib implikatsioonide jada $(GM) \implies (LM) \implies (PM)$, kuid vastupidised implikatsioonid üldjuhul ei kehti. Seetõttu tuleb sõltumatus-
graafi konstrueerimist alustada Markovi paaritingimuse kontrollimisest ning seejärel lisada omaduste (LM) ja (GM) täitmiseks vajalikud servad. Ilmneb, et lõiketingimust rahuldavate jaotuste korral piisab paaritingimuse rahuldamisest.



Teoreem 3. *Kui ühisjaotus rahuldab eelpool sõnastatud tingimusi (I1)-(I5), siis on Markovi paaritingimus samaväärne globaalse Markovi tingimusega.*

TÕESTUS. Tõestus on ära toodud lisas 1. □

Kaared Bayesi võrgus paljastavad tipupaare, mis ei rahulda paaritingimust. Nii viitab kaar $X_i \rightarrow X_j$ enamasti sõltuvusele $D(X_i, X_j | \mathcal{X} \setminus \{X_i, X_j\})$. Kuid leidub erandeid, näiteks kui leidub tipp $X_k = X_j$, siis on tipud X_i ja X_j tinglikult sõltumatud. Ka samasse tippu suubuvad kaared $X_i \rightarrow X_k$ ja $X_j \rightarrow X_k$ viitavad enamasti sõltuvusele $D(X_i, X_j | \mathcal{X} \setminus \{X_i, X_j\})$, mida illustreerib joonis 4. Erandiks on siin näiteks juhuslikud suurused a, b ja c , kus c on võetud ühtlaselt hulgast $\{0, 1, 2\}$ ning binaarsed juhuslikud suurused a ja b on määratud tabeliga.

c	0	1	2
$p(a = 1 c)$	1/10	1/10	4/10
$p(b = 1 c)$	1/10	3/10	2/10

Nüüd tinglikust sõltumatusest c suhtes järeldub a ja b sõltumatus, kuid suuruse c jaotus sõltub siiski mõlemast tipust ehk $D(a, \{b, c\} | \emptyset)$ ja $D(\{a, c\}, b | \emptyset)$. Seega jaotuse Bayesi võrk vastab joonisele 4, kuid tippude a ja b vahele pole serva vaja!

Ülaltoodud kontranäidete tõttu pole minimaalse sõltumatuste graafi konstrueerimine Bayesi võrgu topoloogiast lähtudes võimalik isegi siis, kui kehtib lõiketingimus. Samas saame topoloogiast lähtudes üldjuhul parima aproksimatsiooni.

Definitsioon 8. Bayesi võrgu G moraaliks nimetatakse lihtgraafi G^m , kus iga kahe tippu X_i ja X_j vahel on serv parajasti siis, kui X_i ja X_j vahel on kaar või leidub tipp X_k nii, et tippudest X_i ja X_j lähtuvad kaared tippu X_k .

Teoreem 4. *Bayesi võrgu moraal on sõltumatuste graaf.*

TÕESTUS. Teoreemi tõestus on toodud lisas 1. □

Tinglikult võib seoseid Bayesi võrgus mõista teineteist mõjutavate tegurite jadana, kus tippudel pole oma eellastele mingit mõju. Seda silmas pidades saame küllaltki objektiivse sõltumatuse definitsiooni.

Definitsioon 9. Me ütleme, et tipuhulgad A ja B on d -eraldatud (d -separated) hulga C poolt, kui A ja B on eraldatud hulga C poolt tipuhulga $A \cup B \cup C$ eellashulgale vastava võrgu moraaliks $(\mathcal{A}_n(A \cup B \cup C))^m$, kus

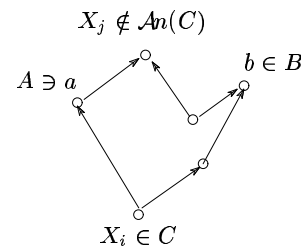
$$\mathcal{A}_n(A \cup B \cup C) = \{x : \text{leidub suunatud tee } x \rightsquigarrow A \cup B \cup C\}.$$

Kuna tipuhulkade A, B ja C eellashulgale vastav alamgraaf on ka Bayesi võrk, siis järeldub tipuhulkade d -eraldatusest ka nende sõltumatus, kuid üldjuhul vastupidine implikatsioon ei kehti. Sellest hoolimata annab d -sõltumatus lihtsa viisi tingliku sõltumatuse tuvastamiseks.

Teoreem 5 (d-eraldatuse kirjeldus). Bayesi võrgu tipuhulgad A ja B on d -eraldatud tipuhulga C suhtes, kui iga tee korral alusgraafis hulgast A hulka B on täidetud üks kahest:

(D1) teel leidub tipp $X_i \in C$ nii, et mõlemad kaared pole vastakuti tipus X_i ;

(D2) teel leidub vastakuti olevate kaartega tipp X_j , mis pole C eellane.



TÕESTUS. Tingimus (D1) välistab sama tee eellasgraafi graafi moraalis. Tingimus (D2) kindlustab, et tippu X_j ei vaadelda sõltumatuse kontrollil. Kui graafis leidub tee, mis ei rahulda tingimusi (D1) ja (D2), siis esiteks kuulub see tee eellasgraafi. Teiseks avab tingimus (D1) moraalis möödapääsud hulga C tippudest antud teel. Ja eellasgraafi moraalis täpselt neid tippe läbiv tee, mis on vastuolus sõltumatusega. \square

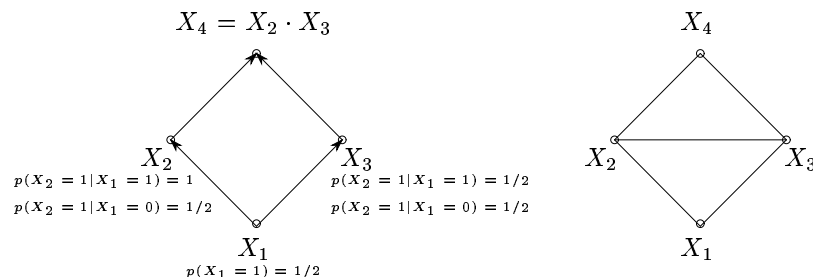
8 Triangulatsioon ning klikipuu

Eespool saadud tulemused võimaldavad, meil keskenduda ühisjaotuse faktoriseerimisele. Keskset rolli hakkab mängima Bayesi võrgu moraal ning sellest saadud klikipuu. Kuid esmalt fikseerime lihtsa, kuid olulise tulemuse.

Teoreem 6. Sõltumatuste graafi servade lisamisel jääb graaf ikkagi sõltumatuste graafiks.

TÕESTUS. Kuna servade lisamine ei kaota ühtegi teed tipuhulkade A ja B vahel, siis tipuhulkade sõltuvus säilib lisamisel. Teisisõnu uusi sõltumatuid tipukolmikuid juurde ei tule. \square

Järgneva teoreetilise käsitluse illustreerimiseks vaatame lihtsat mängu, kus eesmärgiks on saada kaks järjestikust kulli maksimaalselt kolmel viskel. Kui esimesel katsel tuleb kull, siis peab mängija võitmiseks saama ka teisel katsel kulli. Kui esimesel katsel tuleb kiri, siis peavad järgmised katsed andma kulli. Kodeerides kulli ühena ning kirja nullina, saame mängu Bayesi võrgu esitada joonisena 5. Tipud X_2 ja X_3 ei vasta otseselt katsetele: X_2 vastab ainuvõimaliku



Joonis 5: Kahe kulli mängu Bayesi võrk ja selle moraal

kullide paari esimesele ja X_3 teisele komponendile. Bayesi võrgu definitsioonist lähtuv lahendus oleks keskmistada üle X_1 , X_2 ja X_3 väärtuste

$$p(X_4 = 1) = \sum_{X_1} \sum_{X_2} \sum_{X_3} p(X_1, X_2, X_3, X_4 = 1).$$

Kuid see on üldiselt liiga töömahukas — summa liikmete arv on eksponentsiaalne, seetõttu üritatakse alati ühisjaotust faktoriseerida.

Ühisjaotuse faktoriseerimisel on eesmärgiks leida võimalikult väiksed indeksite hulgad $\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_r$ ja potentsiaalifunktsioonid f_j nii, et

$$p(X_1, X_2, \dots, X_n) = \prod_{j=1}^r f_j(X_{\mathcal{I}_j}), \quad X_{\mathcal{I}_j} = \{X_i | i \in \mathcal{I}_j\}.$$

Paljud võimalikud faktoriseeringud on triviaalsed, näiteks valem

$$p(X_1, X_2, \dots, X_n) = p(X_1)p(X_2|X_1)p(X_3|X_2, X_1) \cdots p(X_n|X_1, X_2, \dots, X_{n-1})$$

ei anna mingisugust eelist keskmistamisel. Näidet üldistades loeme faktoriseeringu triviaalseks kui indeksite hulgad saab järjestada nii, et $\mathcal{I}_1 \subseteq \mathcal{I}_2 \subseteq \dots \subseteq \mathcal{I}_r$. Loomulikult leidub jaotusi, millel pole mittetriviaalseid faktoriseeringuid. Kerge on taibata, et selliste Bayesi võrkude moraaliks on täisgraaf. Kui täisgraaf esineb uuritavas graafis alamgraafina, siis nimetatakse selle tippe klikiks. Teisisõnu kliki iga kahe erineva tipu vahel on serv. Üldiselt annavad sõltumatuste graafi klikid märku juhuslike suuruste komplektidest, millel enamasti pole mittetriviaalseid faktoriseeringuid. Veelgi enam, kui jaotuse tihedusfunktsioon on igas punktis positiivne, saab ühisjaotust alati esitada klikkide väärtustuste vastavate potentsiaalide korrutisena.

Teoreem 7 (Grimmett'i esitus). *Kui ühisjaotuse tihedusfunktsioon on rangelt positiivne, siis leiduvad potentsiaalifunktsioonid $\phi_C : X_C \rightarrow \mathbb{R}$, $C \in \mathcal{C}$ nii, et*

$$p(X_1, X_2, \dots, X_n) = \prod_{C \in \mathcal{C}} \phi_C(X_C),$$

kus \mathcal{C} on kõigi maksimaalsete klikkide hulk.

TÕESTUS. Täielik tõestus on toodud ära näiteks Paget'i doktoritöös [Pag99]. \square

Meie uuritavas näites on klikkideks \emptyset , $\{X_1\}$, $\{X_2\}$, $\{X_3\}$, $\{X_4\}$, $\{X_1, X_2\}$, $\{X_1, X_3\}$, $\{X_2, X_4\}$, $\{X_3, X_4\}$, $\{X_1, X_2, X_3\}$ ja $\{X_2, X_3, X_4\}$. Neist maksimaalsed on vaid viimased kaks. Seega kui Grimmeti teoreemi eeldused oleks täidetud⁵, siis saaks ühisjaotuse esitada kujul

$$p(X_1, X_2, X_3, X_4) = \phi_1(X_1, X_2, X_3) \cdot \phi_2(X_2, X_3, X_4).$$

Grimmett'i esitust tuntakse ka Gibbs'i jaotuse ja Markovi väljade ekvivalentsiteoreemina. Tõestuses konstrueeritud potentsiaalidel ϕ_C pole kergesti hoomatavat tähendust. Meie lähimaks eesmärgiks on leida selline sõltumatuste graaf, mille korral on potentsiaalidel selge tähendus — klikile vastav tõenäosusjaotus. Kuid ennem peame tegema graafiteoreetilise kõrvalepõike ning defineerima klikigraafid.

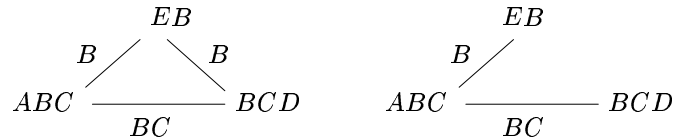
Definitsioon 10. Graafile G vastava klikigraafi \mathcal{G} tippudeks on maksimaalsed klikid C_1, C_2, \dots, C_m ning klikkide C_i ja C_j vahelisteks servaks (eraldajahulgaks) loeme nende mittetühja lõiget $S_{ij} = C_i \cap C_j$.

⁵Selle, miks Grimmeti teoreemi eeldused pole täidetud, jätame lugejale iseseisvaks põhendamiseks.

Meie näite korral on klikigraafis kaks tippu $\{X_1, X_2, X_3\}$ ja $\{X_2, X_3, X_4\}$, mille vahel on serv $\{X_2, X_3\}$. Järgnevas hakkavad meid huvitama graafid, mille klikigraafideks on puud või “peaaegu” puud. Selleks defineerime erikujulise graafide klassi, mille korral lubame endale teatud servade eemaldamist.

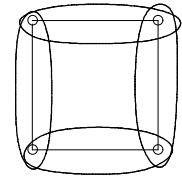
Definitsioon 11. Graafi maksimaalsete klikkide jada C_1, C_2, \dots, C_m rahuldab jagamatu lõike tingimust, kui iga klikile C_k vastav ühisosa $C_k \cap (C_1 \cup \dots \cup C_{k-1})$ sisaldub ühes eelnevas klikis C_j , $j < k$. Graafi klikipuu koosneb maksimaalsetest klikkidest, kuid iga kliki C_k ja sellele eelnevate klikkide vahele jätame alles vaid suurima mittetühja lõike $C_k \cap C_j$, $j < k$.

Kui graafi klikigraafiks on puu, siis langeb see kokku klikipuuga. Kuid klikipuu definitsioon võimaldab osad korduvad klikigraafi servad eemaldada. Igal



Joonis 6: Jagamatu lõike tingimusega klikigraaf ning vastav klikipuu

graafil pole klikipuud, väikseimaks näiteks on neljane tsükkel. Probleemi lahendamaks serv, mis ühendaks vastastippe. Seda tüüpi servi, mis ühendavad tsükli kahte mittejärjestikust tippu, nimetatakse kõõludeks. Ilmneb, et vähemalt nelja tipuga tsükklite lõhkumine kõõludega on tarvilik ja piisav klikipuu olemasoluks.



Definitsioon 12. Trianguleeritud graafi iga tsükkel, milles on vähemalt neli tippu, omab kõõlu.

Teoreem 8. Sidusal graafil on klikipuu parajasti siis, kui ta on trianguleeritud.

TÕESTUS. Olgu graafis on ilma kõõluta tsükkel, milles on vähemalt neli tippu. Lihtne on mõista, et tsükkel ei saa täielikult kuuluda ühtegi klikki. Teisalt moodustavad tsükli teele jäävad klikid meta-tsükli ning jagamatu lõike tingimus on rikutud.

Kui ei leidu klikipuud, siis peab leiduma vähemalt neljane metatsükkel, mille klikke ei saa sobivalt järjestada (kolmene tsükkel vastab klikile ja vastuolu ei teki!). Vaatame lühimat metatsükli ning selle servi. Kuna klikkide lõiked on mittetühjad, siis saame vastava pikkusega tsükli ka esialgses graafis. Et graaf on trianguleeritud, siis leidub kõõl ning see on vastuolus metatsükli minimaalsusega. \square

Trianguleerides Bayesi võrgu moraali, saame klikipuuga graafi. Loomulikult vastab minimaalse servade arvuga trianguleeritud graaf parimale ohutule aproksimatsioonile, millel on klikipuu. Kahjuks on minimaalse trianguleeritud graafi leidmine \mathcal{NP} -raske ülesanne ning sellepärast kasutatakse erinevaid ahneid algoritme nagu Kjærulff'i heuristika [Kær90]. Õnneks on meie näitegraaf on juba trianguleeritud kujul. Samas on trianguleeritud graafides klikkide otsimine lihtsam, selleks saab kasutada erikujulisi algoritme [Gol80].

Iga eraldajahulk indutseerib klikipuu täieliku lahutuse $\mathcal{T} = \mathcal{T}_1 \dot{\cup} \mathcal{T}_2$, kus \mathcal{T}_1 ja \mathcal{T}_2 on servaga eraldatud alampuud. Järgnev tulemus annab võtme ühisjaotuse faktoriseerimiseks.

Lemma 1. *Tipuhulgad A_1 ja A_2 , mis vastavad klikipuu täielikule lahutusele, on tinglikult sõltumatud alampuude vahelise eraldajahulga suhtes.*

TÕESTUS. Olgu alampuude vaheline eraldajahulk S . Lihtne induktsioon üle lisatud klikkide annab, et erinevatest alampuudest võetud klikkide lõige kuulub eraldajahulka S .

Suvaline tee tipuhulgast A_1 tipuhulka A_2 peab sisaldama serva, mille üks ots kuulub tipuhulka A_1 ning teine A_2 . Kahtlemata peab antud serv kuuluma mõnda klikki ja seega serva tipud samasse hulka. Kuna $A_1 \cap A_2 = S$, siis üks serva otspunkt on hulgas S . \square

See lihtne lemma sillutab teed teooria sõlmtulemusele, millele on rajatud efektiivne keskmistamine üle erinevate juhuslike suuruste.

Teoreem 9 (Klikipuu potentsiaalide esitus). *Kui sõltumatuste graafide vastab klikipuu, siis faktoriseerub ühisjaotus vastavalt klikkide ja eraldajahulkade tõenäosusjaotustele*

$$p(X_1, X_2, \dots, X_n) = \prod_{k=1}^m \frac{p(C_k)}{p(S_k)} = \prod_{k=1}^m p(R_k|S_k),$$

kus eraldajahulk S_k on klikki C_k ning sellele eelnenud klikkide vaheline serv ning jääkhulk $R_k = C_k \setminus S_k$.

TÕESTUS. Üldsust kitsendamata vaatame sidusaid sõltumatuste graafe. Tõestuseks piisab, kui näidata võrdust $p(R_k|R_{k-1}, \dots, R_1) = p(R_k|S_k)$. Lemmast 1 saame, et C_k ja $C_1 \cup \dots \cup C_{k-1}$ on tinglikult sõltumatud S_k suhtes, mistõttu

$$p(R_k|C_1 \cup \dots \cup C_{k-1}) = p(R_k|S_k).$$

\square

Näiliselt süütu teoreem annab tohutut arvutuslikku kokkuhoidu. Kui meid huvitavad parameetrid kuuluvad ühte klikki, siis võime keskmistada vaid selle. Kui muutujad on mitmes klikis, siis on ühisjaotuse leidmine veidi keerulisem. Kuid ikkagi saab mitteoluliste klikkide potentsiaali ignoreerida, valides vastavalt klikipuu potentsiaalide esitusele minimaalse arvu ühisjaotuse taastamiseks vajalikke klikke. Meie näiteülesande korral saame nüüd esituse

$$p(X_1, X_2, X_3, X_4) = \frac{p(X_1, X_2, X_3)p(X_2, X_3, X_4)}{p(X_2, X_3)}.$$

Ning meid huvitava tõenäosuse arvutamiseks peame leidma $p(X_2, X_3, X_4)$ ning keskmistama ainult üle X_2 ja X_3 .

9 Klikkidele vastavate ühisjaotuste leidmine

Eelnevad tulemused kindlustavad klikipuu olemasolu. Sealjuures on klikipuu konstrueerimisel mõned nipid, mis vähendavad edasist arvutusmahtu. Vastavad algoritmid on toodud ülevaateartiklites [HD96, JJ94], kuid nendega tutvumise jätame lugejale iseseisvaks ülesandeks.

Klikkidele vastavad potentsiaalid rahuldavad kooskõlatingimusi, kusjuures need on piisavad ning tarvilikud.

Teoreem 10. Iga klikipuu potentsiaalid rahuldavad kooskõlatingimusi:

(LK) lokaalne kooskõla (local consistency) — iga klikki C ja selles sisalduva eraldajahulga S korral

$$\int \phi_C(X_C) dX_R = \phi_S(X_S), \quad R = C \setminus S;$$

(GK) globaalne kooskõla (global consistency) — potentsiaalid vastavad ühisjaotusele

$$p(X_1, X_2, \dots, X_n) = \prod_{k=1}^m \frac{p(C_k)}{p(S_k)}. \quad (1)$$

Kooskõla tingimused on tarvilikud ja piisavad potentsiaalide fikseerimiseks.

TÕESTUS. Tarvilikkuseks paneme tähele, et klikki potentsiaali $\phi_C(X) = p(X_C)$ keskmistamine annab

$$\int p(X_C) dX_R = p(X_S), \quad R = C \setminus S.$$

Teisalt on globaalne kooskõla ilmne, kuna lokaalselt kooskõlalised potentsiaalid defineerivad ühisjaotuse vastavalt valemile (1). Globaalse kooskõla tõttu on see otsitav ühisjaotus. \square

Kui muutujad \mathcal{X} on diskreetsed, siis kõdub integraal summaks ning ühisjaotused saab salvestada tabelitesse. Lokaalse ja globaalse kooskõla saavutamiseks kasutatakse kolme erinevat võtet. Põhimõtteliselt pole algoritm piiratud diskreetsete juhuslike suurustega, kuid jaotusi peab olema võimalik salvestada ning keskmistada. Üheks selliseks sobivaks pidevaks jaotuseks on normaaljaotus.

Esimeseks sammuks on potentsiaalide initsialiseerimine. Selleks võetakse klikkide potentsiaalid $\phi_C \equiv 1$, sama tehakse eralajahulkadega $\phi_S \equiv 1$. Iga tipu X_i korral otsitakse klikk C , mis sisaldab X_i ja eellasi $\mathcal{P}a(X_i)$, ning parandatakse klikki potentsiaali

$$\phi_C(X) = \phi_C(X)p(X_i|\mathcal{P}a(X_i)).$$

Klikk C leidub tänu moraali konstruktsioonile! Kuna iga tipp toob valemisse (1) sisse parandusliikme $p(X_i|\mathcal{P}a(X_i))$, siis kehtib globaalse kooskõla tingimus, kuid lokaalne kooskõla on reeglina rikutud.

Lokaalse kooskõla saavutatakse naaberklikkide mõjutamisel läbi eraldajahulkade. Ühte sellist vastasmõju nimetatakse teateks (*message*) ning teate edastamine klikkide C ja D vahel koosneb kahest etapist.

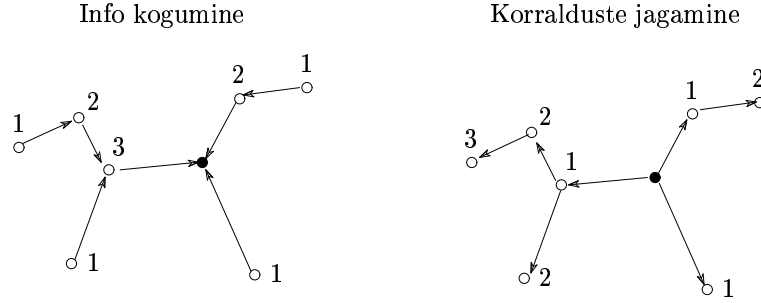
1. **Keskmistamine.** Salvestakse eraldajahulga potentsiaal ϕ_S^{old} ning leitakse uus

$$\int \phi_C(X_C) dX_R = \phi_S(X_S), \quad R = C \setminus S.$$

2. **Neelamine.** Klikki D potentsiaaliks võetakse

$$\phi_D = \phi_D \frac{\phi_S}{\phi_S^{old}},$$

sealjuures loetakse $0/0 = 0$.



Joonis 7: Kooskõla saavutamise kampaaniad

Et teate edastamisel muutub vaid eraldajahulga ning kliki D potentsiaal, siis võrdus

$$\frac{\phi_D^{old}}{\phi_S^{old}} = \frac{\phi_D^{old}}{\phi_S^{old}} \frac{\phi_S}{\phi_S} \frac{\phi_S^{old}}{\phi_S} = \frac{\phi_D}{\phi_S}$$

kindlustab globaalse kooskõla.

Lokaalne kooskõla saavutatakse kahe kampaaniaga. Esmalt kogutakse tõestusmaterjal juurtippu ning seejärel saadetakse korraldused tagasi lehtedesse. Puu juurtipuks on kasulik võtta klikipuu tsenter, sest siis on teed puu lehtedeni lühimad ning arvutusvigade mõju on väiksem. Tõestusmaterjali kogumiseks saadavad esmalt lehed teateid vahetutele ülemustele. Iga vahetipp edastab teate oma ülemusele alles siis, kui kõik teated alluvatele on neelatud.

Tõestusmaterjali jagamisel saadab juurtipp teated oma alluvatele, need omakorda oma alluvatele. Jällegi ei saada ükski tipp teadet enne oma teate neelamist.

Teoreem 11. *Pärast tõestusmaterjali kogumist ning selle laialisaatmist on rahuldatud nii globaalne kui lokaalne kooskõla tingimus.*

TÕESTUS. Teoreem tõestatakse induksiooniga üle klikipuu tippude. Vaatame vaid tõestuse olulisemat fragmenti — sõnumite vahetust kahe tipu vahel. Olgu

$$\phi_S = \int \phi_C dX_R, \quad \phi'_S = \int \phi_D dX'_R, \quad R = C \setminus S, \quad R' = D \setminus S,$$

siis pärast teist sõnumit on $\phi_S^{new} = \phi'_S \phi_S / \phi_S^{old}$ ja kooskõla kehtib D ja S vahel. Teisest küljest

$$\int \phi_C^{new} dX_R = \int \phi_C \frac{\phi_S \phi'_S}{\phi_S \phi_S^{old}} dX_R = \phi'_S \phi_S / \phi_S^{old}.$$

Sama arutelu kehtib ka mitme alluva korral ning sealjuures ei pea tipuhulgas olema klikid.

Täielikuks tõestuseks tuleb eraldada lehe ülemus ning vaadata selle sõnumivahetust. Kuid täpsemad detailid jätame lugejale. \square

Kuigi kampaaniad ise võivad olla kulukad, on arvutuslik kokkuhoid märkimisväärne — iga keskmistamisel hüljatud parameeter vähendab keerukust kordades. Erinevate tõenäosuste leidmiseks piisab ühest kampaaniast. Teisest küljest

lubab arvutusmahu kahanemine kasutada keerukamaid Bayesi võrke — täpsemaid statistilisi mudeleid. Sealjuures tuleb tunda piire: igast valdkonnas saab luua ekspertsüsteemi, kuid kaugeltki mitte kõik neist pole vajalikud.

Tulles tagasi meie näite juurde on pärast initsialiseerimist saadud jaotuste tabelid ilma nullideta järgnevad.

Kliik { X_1, X_2, X_3 }			
X_1	X_2	X_3	p
0	0	★	1/8
0	1	★	1/8
1	1	0	1/4
1	1	1	1/4

Kliik { X_2, X_3, X_4 }			
X_2	X_3	X_4	p
0	0	0	1
0	1	0	1
1	0	0	1
1	1	1	1

Lokaalne koosõla saabub kohe pärast esimest teadet, sest X_2 ja X_3 määravad üheselt X_4 .

Kliik { X_1, X_2, X_3 }			
X_1	X_2	X_3	p
0	0	★	1/8
0	1	★	1/8
1	1	0	1/4
1	1	1	1/4

Kliik { X_2, X_3, X_4 }			
X_2	X_3	X_4	p
0	0	0	1 · 1/8
0	1	0	1 · 1/8
1	0	0	1 · 3/8
1	1	1	1 · 3/8

Ning meid huvitava kahe järjestikuse kulli tõenäosus $p(X_4 = 1) = 3/8$.

10 Täiendava tõestusmaterjali kasutamine

Bayesi võrk kajastab meie aprioorsest uskumust, mille keskmistamisel saame meid huvitavate parameetrite tõenäosusjaotuse. Lisaks aprioorsele uskumusele võib aja jooksul kerkida esile uut ja huvitavat informatsiooni — täiendavat tõestusmaterjali. Seega on igati õigustatud küsimus, kuidas muudab täiendav informatsioon meie uskumust. Ka sellele küsimusele annab vastuse Bayesi valem. Olgu aprioorne jaotus $p(\theta, \psi)$, kus meie tõestusmaterjal D käib parameetrite ψ kohta. Meid huvitab seega aposterioorne tõenäosus

$$p(\theta, \psi|D) = \frac{p(D|\psi)p(\theta, \psi)}{p(D)} \propto p(D|\psi)p(\theta, \psi),$$

kus $p(D|\psi)$ kvantifitseerib tõestusmaterjali väärtust ning $p(D)$ on tõestusmaterjali täistõepära (*complete data likelihood*). Ning kontseptuaalselt on probleem lahendatud.

Praktiliseks lahendamiseks piisab, kui täiendame Bayesi võrku nii, et see kannaks normeerimata tõenäosusjaotust $p(D|\psi)p(\theta, \psi)$. Teisisõnu tuleb muuta klikipuu initsialiseerimisprotseduuri.

Vaatame esmalt lihtsaimat juhtu, kus tõestusmaterjal määrab täielikult parameetrite ψ väärtused x_i^{obs} , $i \in \psi$. Klikid initsialiseeritakse standardisel viisil, seejärel leitakse igale parameetrile $X_i \in \psi$ vastav klikk ning tehakse parandus

$$\phi_C(X) = \begin{cases} 0, & \text{kui } X_i \neq x_i^{obs}, \\ \phi_C(X), & \text{kui } X_i = x_i^{obs}. \end{cases}$$

Klikkide potentsiaalid on globaalses kooskõlas jaotusega $p(\theta, \psi = \psi^{obs})$ ning seega annab teadete edastamise kampaania vajalikud marginaaljaotused konstandi $p(\psi = \psi^{obs})$ täpsusega. Tõenäosuse $p(\psi = \psi^{obs})$ saame lihtsalt esialgse jaotuse $p(\theta, \psi = \psi^{obs})$ keskmistamisel üle θ . Meie näiteülesande korral nullib teadmine $X_1 = 1$ paljud tabeli read ning kampaaniad annavad tulemuseks

Klikk $\{X_1, X_2, X_3\}$			
X_1	X_2	X_3	p
1	1	0	1/4
1	1	1	1/4

Klikk $\{X_2, X_3, X_4\}$			
X_2	X_3	X_4	p
1	0	0	1 · 1/4
1	1	1	1 · 1/4

ja oodatud vastus on $p(X_4 = 1) = (1/4)/(1/4 + 1/4) = 1/2$.

Nõrk sõltumatu tõestusmaterjal koosneb sõltumatute sündmuste tõenäosustest $p(D_i|X_i)$, $i = 1, \dots, k$. Nüüd on igale tipule X_i vastav parandus

$$\phi_C(X) = \phi_C(X)p(D_i|X_i)$$

ning tulemuseks on normeerimata tõenäosusjaotus $p(D|\psi)p(\theta, \psi)$. Meie näite puhul võiks nõrgaks tõestusmaterjaliks olla mängija naeratus peale esimest viset. Oletame, et meie mängija naerataks tõenäosusega 1/2 peale kirja saamist ning tõenäosusega 3/4 peale kulli saamist. Nüüd võttes arvesse nõrka tõestusmaterjali saame uued algtabelid

Klikk $\{X_1, X_2, X_3\}$			
X_1	X_2	X_3	p
0	0	★	1/8 · 1/2
0	1	★	1/8 · 1/2
1	1	0	1/4 · 3/4
1	1	1	1/4 · 3/4

Klikk $\{X_2, X_3, X_4\}$			
X_2	X_3	X_4	p
0	0	0	1
0	1	0	1
1	0	0	1
1	1	1	1

ning pärast lokaalset kooskõla saavutamist on tulemuseks

Klikk $\{X_1, X_2, X_3\}$			
X_1	X_2	X_3	p
0	0	★	1/16
0	1	★	1/16
1	1	0	3/16
1	1	1	3/16

Klikk $\{X_2, X_3, X_4\}$			
X_2	X_3	X_4	p
0	0	0	1/16
0	1	0	1/16
1	0	0	1/4
1	1	1	1/4

Seega peaks kahe kulli saamise tõenäosus olema 2/5, mis suurem kui lisateadmisteta saadud tõenäosus 3/8 ja väiksem, kui teadmises $X_1 = 1$, leitud tõenäosus 1/2.

Kui $p(D|\psi)$ sõltub mitmest parameetrist, siis on mõistlik Bayesi võrku lisada tipp D seda määravate eellastega ning tõenäosusjaotusega $p(D|\psi)$. Täiendatud Bayesi võrgus vastab tõestusmaterjal parameetrite fikseerimisele.

Tõestusmaterjali saab ka iteratiivselt lisada, siis on võimalik osad vahetulemused säilitada. Lisaks sellele saab ära kasutada mitmeid võimalusi arvutuste optimeerimiseks, mõningaid vihjeid ning edasisi viiteid pakub Huang ja Darwiche ülevaateartikkel [HD96].

11 Kokkuvõtte asemel

Loodan, et ülevaade pakkus lugejale esmase mulje Bayesi võrkudest. Loomulikult ei kata see kaugeltki kõiki Bayesi võrkudega seotud probleeme, tegelikult on see vaid kontseptsioonide esmatutvustus, mille aluseks on Duke'i ülikooli statistika kursus [Peo99]. Nii saab igast analüüsi osast: võrgu konstrueerimisest, moraali trianguleerimisest, klikkide leidmisest ning potentsiaalifunktsioonidest kirjutada eraldi ülevaate.

Palju huvitavamaid rakendusi pakuvad treenitavad Bayesi võrgud, näiteks rämpsposti filtrid. Lihtsaimal juhul vastab võrgu treenimisele kaarte tõenäosuste $p(X_i | \mathcal{Pa}(X_i))$ automaatne määramine. See toob endaga kaasa suurima aposterioorse tõepära printsiibi või üle mudelite keskmistamise. Pahatihti kasutatakse arvutusmahu piiramiseks ligikaudseid Monte-Carlo meetodeid nagu Gibbs'i ja Metropolis'e algoritme või hoopiski EM-algoritmi.

Olemasoleva teooria ja rakenduste hulka vaadates tuleb Bill Gates'iga nõustuda — Bayesi võrkudel on tulevikku.

Viited

- [Bi195] Patrick Billingsley. *Probability and Measure*. Probability and Mathematical Statistics. Wiley and Sons, third edition, 1995.
- [BT92] George E. P. Box and George C. Tiao. *Bayesian Inference in Statistical Analysis*. Wiley and Sons, 1992.
- [CGH97] Enrique Castillo, Jose Manuel Gutierrez, and Ali S. Hadi. *Expert Systems and Probabilistic Network Models*. Springer, 1997.
- [GCS⁺03] Andrew Gelman, John B. Carlin, Hal S. Stern, Donald B. Rubin, and A. Gelman. *Bayesian Data Analysis*. CRC Press, 2003.
- [Gol80] Martin Charles Golumbic. *Algorithmic Graph Theory and Perfect Graphs*. Academic Press, New York, 1980.
- [HBH⁺98] E. Horvitz, J. Breese, D. Heckerman, D. Hovel, and K. Rommelse. The lumiere project: Bayesian user modeling for inferring the goals and needs of software users. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, pages 256–265, 1998. Rohkem informatsiooni *Lumiere* projekti kohta saab <http://research.microsoft.com/~horvitz/lum.htm>.
- [HD96] Cecil Huang and Adnan Darwiche. Inference in belief networks: A procedural guide. *International Journal of Approximate Reasoning*, 15:225–263, 1996. Kättesaadav veebist.
- [Hel96] Leslie Helm. Improbable inspiration. Los Angeles Times, October 28 1996. Lõbus artikkel *Microsofti* arendustööst.
- [HHN92a] D. Heckerman, E. J. Horvitz, and B. Nathwani. Toward normative expert systems I: The PATHFINDER project. *Methods of Information in Medicine*, 31(2):90–105, June 1992. Kahjuks kättesaamatu.

- [HHN92b] D. Heckerman, E. J. Horvitz, and B. Nathwani. Toward normative expert systems: Part I the pathfinder project. KSL 92-66, Stanford University, 1992. Kättesaadav veebist.
- [JJ94] Finn V. Jensen and Frank Jensen. Optimal junction trees. In *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, pages 360–366, 1994. Kättesaadav veebist.
- [Kær90] Uffe Kærulff. Triangulation of graphs — algorithms giving small total state space. Technical Report 90-09, Aalborg University, 1990. Kättesaadav veebist.
- [Ker91] G. Keren. Calibration and probability judgments: Conceptual and methodological issues. *Acta Psychologica*, 77(3):217–273, 1991. Kahjuks kättesaamatu.
- [Pag99] Rupert D. Paget. *Nonparametric Markov Random Field Models for Natural Texture Images*. PhD thesis, University of Queensland, Feb 1999. Kättesaadav veebist.
- [Peo99] Mark Peot. Sta 294: Special topics belief networks. Loengukonstpektid, 1999. Vaata lähemalt <http://www.stat.duke.edu/courses/Spring99/sta294/>.
- [SvH75] C. Spetzler and C. Stael von Holstein. Probability encoding in decision analysis. *Management Science*, 22(340–358), 1975. Kahjuks kättesaamatu.
- [Wan98] Hongbin Wang. *Order effects in human belief revision*. PhD thesis, Ohio State University, 1998. Annab ülevaate ning sisaldab palju viiteid antud ainevallale.
- [Wil94] Alyson G. Wilson. Cognitive factors affecting subjective probability assessment. Discussion Paper 94-02, Duke University, 1994.

12 Lisa 1: Tõestused

Teoreem 12. *Kui ühisjaotus rahuldab tingimusi (I1)-(I5), siis on Markovi paaritingimus samaväärne globaalse Markovi tingimusega.*

TÕESTUS.

(PM) \implies (LM)

Lõikeomadus lubab seostest $I(X_i, B_1 | \mathcal{X} \setminus (X_i \cup B_1))$ ja $I(X_i, B_2 | \mathcal{X} \setminus (X_i \cup B_2))$ järeldada $I(X_i, B_1 \cup B_2 | \mathcal{X} \setminus (X_i \cup B_1 \cup B_2))$. Võttes arvesse kõik X_i seotud Markovi paarid, saamegi $I(X_i, \mathcal{X} \setminus \{X_i\} | \mathcal{Bd}(X_i))$.

(LM) \implies (GM)

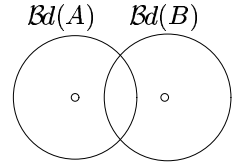
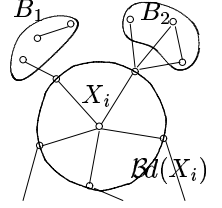
Toome sisse tähistuse $A^c = \mathcal{X} \setminus (A \cup \mathcal{Bd}(A))$ ning näitame et iga tipuhulga A korral kehtib $I(A, A^c | \mathcal{Bd}(A))$. Esmalt iga tipuhulga A ja B korral

$$I(A, A^c | \mathcal{Bd}(A)) \wedge I(B, B^c | \mathcal{Bd}(B)) \implies I(A \cup B, (A \cup B)^c | \mathcal{Bd}(A) \cup \mathcal{Bd}(B)).$$

See tuleneb otseselt nõrgast laiendatavusest ning taandatavusest. Teisalt jällegi $\mathcal{Bd}(A) \cup \mathcal{Bd}(B) = \mathcal{Bd}(A \cup B) \cup D$, kus $D \subseteq A \cup B$. Kuna kehtib $I(C, D | \mathcal{Bd}(A \cup B) \cup (A \cup B))$, siis lõikeomadusest saame

$$I(A \cup B, (A \cup B)^c | \mathcal{Bd}(A \cup B)).$$

Kuna lokaalne Markovi omadus kindlustab $I(\{a\}, \{a\}^c | \mathcal{Bd}(a))$, saame kõiki selliseid seoseid arvestades näidata, et $I(A, A^c | \mathcal{Bd}(A))$. Vaatleme nüüd graafis olevat sõltumatut kolmikut A, B ja C . Kuna C on eraldaja, siis leidub hulk $A \subseteq A_0$ ja mille piiiriks on C . Saab näiteks hulga A laiendamisel kuni C elementideni. Lihtne on taibata $B \subseteq A_0^c$ ning eelnevast $I(A_0, A_0^c | C) \implies I(A, B | C)$. \square



Teoreem 13. *Bayesi võrgu moraal on sõltumatuste graaf.*

TÕESTUS. Iga tipukolmiku A, B ja C korral Bayesi võrgu moraal, piisab, kui vaatame vaid tippe, millest leidub suunatud tee hulkadesse A, B või C . Ülejäänud tipud ei tule A, B, C ühisjaotuse avaldamisel arvesse.

Tõestame väite esmalt tipukolmiku A, B ja C korral, kus $C \subseteq \mathcal{Nd}(A \cup B)$ (mittejärglaste hulk). Tähistame sümboliga A^+ kõigi A tippude eellashulka, mis on saadud hulga A laiendamisel kuni tipuhulgani C . Paneme tähele, et $A^+ \cap B^+ = \emptyset$ ja iga tipu $a \in A^+$ vanemad kuuluvad kas hulka A^+ või C . Nüüd on lihtne induktsiooniga üle topoloogilise järjestuse näidata $p(A^+ | C, B^+) = p(A^+ | C)$, millest saamegi $I(A, B | C)$.

Üldjuhul jaguneb $C = C_0 \cup C_1 \cup C_2$, kus $C_0 \subseteq \mathcal{Nd}(A \cup B)$ ja $C_1, C_2 \subseteq \mathcal{D}_e(A \cup B)$ (järglastehulk) ning kõik suunatud teed hulgast A või B hulka C_2 läbivad C_1 . Tähistame nüüd A^+ tippude hulka, millest leidub suunatud tee hulka A või leidub hulgast A suunatud tee, kusjuures teed ei läbi hulka C . On selge, et $A^+ \cap B^+ = \emptyset$ ja A^+ vanemad kuuluvad kas hulka A^+ või C . Eelnev tulemus kindlustab $I(A^+, B^+ | C_0)$.

Vastavalt C_2 konstruktsioonile on C_2 eellaste hulga piiiriks C_1 , seega saab induktsiooniga üle topoloogilise järjestuse näidata $p(C_2 | A^+, B^+, C_0, C_1) = p(C_2 | C_1)$ ehk $I(C_2, A^+ \cup B^+ \cup C_0 | C_1)$. Nõrgast laiendatavusest $I(A^+, C_2 | C_0 \cup C_1 \cup B^+)$.

Kui C_1 on minimaalne võimalik, siis on iga tipp C_1 kas A^+ või B^+ vahetu naaber. Samas ei saa ükski tipp olla korraga A^+ ja B^+ naaber, sest A ja B on moraalil sõltumatud C suhtes. Olgu vastav lahutus $C_1 = C_1^A \cup C_1^B$, siis on lihtne näidata $I(A^+, B^+ \cup C_1^B | C_0 \cup C_1^A)$. See omakorda tähendab $I(A^+, B^+ | C_0 \cup C_1)$ ning taanduvusest saame $I(A^+, B^+ | C)$. \square

