Bayesi võrgud: seitse sammu täiuseni

Sven Laur

24. oktoober 2003. a.

Kokkuvõte

Üheks kaasaegse matemaatika paeluvamaks haruks on Bayesi statistika ning selle analüüsimeetodid. Kesksel kohal Bayesi teoorias kui 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 keerukate põhjuslike seoste kajastamine orienteeritud graafina ning selle hilisem analüüs.

Bayesi võrke on edukalt kasutatud ekspertsüsteemides, sest lihtne modulaarne lähenemine lubab kirjeldada keerukaid teadmisi. Teine oluline Bayesi võrkude rakendusala on ?masinõpe?, kus võrku treenitakse andmeid klassifitseerima. Ehki ehituselt on Bayesi ja neurovõrgud sarnased, siis sellega see piirdub—Bayesi võrk kirjeldab alati tõenäosusjaotust, samas kui neurovõrgud kannavad endas suvalisi funktsioone.

Esmalt anname lühiülevaate Bayesi statistika alustest ning seejärel kirjeldame Bayesi võrgu semantikat. Seejärel keskendume praktikas üliolulisele keskmistamise probleemile, mille üheks lahenduseks on Bayesi võrgu analüüs.

Sisukord

1 Sissejuhatus

* Probleemipüstitus * Lühikokkuvõte

2 Tõenäosus kui ebakindluse mõõt

üheks keskseks filosoofiliseks küsimuseks kaasaegses matemaatikas on tõenäosuse mõiste interpreteerimine. Klassikaline näide tõenäosusest on kulli ja kirja viskamine. Reeglina võetakse kulli saamise tõenaäosuseks automaatselt $1/2$, aga mida see ikkagi tähendab pole kaugeltki selge. Aksiomaatiline tõenäosusteooria käsitleb tõenäosust kui teooriväliselt antud suurust ning ei anna ühtegi meetodit selle määramiseks või interpreteerimiseks.

Objektiivse tõenäosuse pooldajad (*frequentists*) eeldavad, et maailmas leidub igal huvitaval sündmusel üheselt määratud tõenäosus, mida saab kindlaks teha katset lõpmatult korrates. Teise äärmuse järgi pole objektiivset tõenäosust olemas, on olemas vaid üksikisiku veendumusi kajastav subjektiivne ebakindlus. Ebakindlust kasjastav arvuline väärtus ongi tõenäosus.

A priori pole selge, et ebakindlus peaks rahuldama tõenäosusele esitatavaid aksioome. Saab näidata, et ratsionaalse üksikisiku ebakindluse mõõt peab alati rahuldama tõenäosusele seatud kitsendusi. Vaatleme hüpoteetilist olukorda, kus Alice kirglik kihlvedude sõlmija ning salakaval Bob otsustab Alice'lt maksku mis maksab raha välja petta. Selleks saab Bob sõlmida hulga kihlvedusid omapoolse panusega 1. Bob on edukas, kui sõlmitud kihlvedude skeem toob talle iga võimaliku lõpptulemuse korral tulu. Kui Alice käitub ratsionaalselt, siis Bobil ei õnnestu teda nii tüssata. Alice'i ebakindluse mõõduks võime võtta panustatava summa.

ülesanne 1 Olgu Ω kõikvõimalike sündmuste ühendhulk, $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ seadusega lubatud kihlveosündmuste hulk ning $p(A)$ sündmusele A vastav Alice'i panus. Näidata, et Alice käitub on ratsionaalselt, siis

1. sündmuse ebakindlus $p(A)$ ei sõltu selle esitamise viisist;
2. iga lubatud sündmuse ja vastandsündmuse ebakindluste summa on kostantne
$$p(A) + p(A^c) =;$$
3. teineteist välistavate sündmuste ebakindlus on aditiivne
$$A \cap B = \emptyset \qquad \Rightarrow \qquad p(A \cup B) = p(A) + p(B).$$

Milline peab olema lubatud sündmuste hulk $\mathcal{F}$? Kõrvutada seda aksiomaatilises tõenäosusteoorias kasutatava σ-algebraga.

Ühesõnaga on p korrektselt defineeritud ja aditiivne. Kuna kindla sündmuse ebakindlus on konstantne, siis saame ebkindluse normeerida 0 ja 1 vahele. Kui sündmuste ühendhulk Ω on lõplik, siis ebakindlus p on tõenäosusmõõt.

Ülesanne 2 Milliseid tingimusi peab rahuldama lubatud sündmuste hulk $\mathcal{F}$?

Kui sündmuste ühendhulk Ω on lõpmatu, tuleb lisaks eelnevale eeldada Alice'i järjekindlust. Olgu $A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n \subseteq \ldots$ kasvavte sündmuste jada, näiteks lubab Bob Alice'l täpsustada kihlveo tingimusi. Teisalt võib ettenägelik Alice kohe Bobiga kihla vedada sündmuse $A = A_1 \cup A_2 \cup \cdots A_n \cdots$ üle. Alice on järjekindel kui iga sellise sündmuste jada korral $\lim p(A_i) = p(A)$ ehk pole vahet kui kaua võtab aega kihlveo tingimuste täpsustamine.

Ülesanne 3* Näidata, et kui Alice on ratsionaalne, järjekindel ning lubatud on vaid lõplikud panused, siis ebakindlus p on normeerimata tõenäosusmõõt.

Loomulikult sõltub ratsionaalse Alice ebakindlus tema eelnevatest teadmistest. Pole ju mõtet panustada kirjale, kui sa tead, et mündil on kaks kulli. Ebakindlust sündmuse A suhtes, kui on toimunud sündmus B tähistatakse $p(A \mid B)$. Aksiomaatilises tõenäosusteoorias vastab sellele tinglik tõenäosus, mis on defineeritud tautoloogiliselt $p(A \mid B) = p(A,B)/p(B)$, kui $p(B) \neq 0$. õnneks kehtib lisaks vormilisele ka sisuline sarnasus.

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

$$p(A)p(B \mid A) = p(A,B) = p(B)p(A \mid B).$$

Erinevalt klassikalisest lähenenisest vajab väide tõestust, sest $p(A \mid B)$ pole defineeritud läbi tautoloogilise võrduse.

Ülesanne 4* Tõestada Bayesi teoreem, leides vastavad kihlveoskeemid.

Bayesi teoreem formaliseerib ratsionaalse järeldamise protsessi. Reeglina ei õnnestu teadusliku eksperimendi käigus leida otsitava suuruse täpset väärtust. Olgu X otsitav suurus ning D ekperimendi käigus mõõdetud suurused, siis saame tulemuseks tõenäosusjaotuse

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

Esimene liige $p(D \mid X)$ iseloomustab andmete D nägemise tõenäosust kui otsitava suuruse väärtus x. Teine liige iseloomustab meie veendumusi enne katset, st. milline on meie panus x saamiseks. Erinevatel indiviididel võib $p(X = x)$ olla kardinaalselt erinev, seega *järeldused katsest sõltuvad meie eelteadmistest!*

Vastolu objektiivsusega on näiline—teaduslik eksperiment peab olema laialdaselt aksepteeritav. Ehk teissisõnu lõppjäreldused peavad olema sarnased sõltumata isiklikest eelteadmistest.

Klassikaline statistika on oma olemuselt ortogonaalne Bayesi statistikale. Kui Bayesi statistika korral on tulemuseks tundmatu suuruse tõenäosusjaotus, siis klassikalise statistika annab vastuseks arvu või intervalli. Loomulikult läheb punkt- ja intervallhinnangute saamisel kaduma osa võimalikust informatsioonist.

Teine oluline erinevus on tulemuste interpreteerimisel—usaldusintervallid pole otseselt seotud tõenäosusega. Teisteks miinusteks on rakused puuduvate andmete töötlemisel ja eelteadmiste ignoreerimine. Oluliseks eeliseks on efektiivsus, reeglina viib järjekindel Bayesi printsiibi rakendamine arvutuslikult keerukate probleemideni. Teine eelis on eeltõenäosuse (*prior*) puudumine—paljudel juhtudel on raske kui mitte võimatu kodeerida eelnevaid teadmisi tõenäosusena.

Edasist lugemist Selleks, et ehitada üles ratsionaalset ebakindluse mõõtu on mitmeid viise. Valisime kontseptuaalselt lihtsaima *Dutch Book* argumendi [**?**], samas on olemas mitmeid alternatiivseid aksiomaatikaid nagu [**?**, **?**]. Teiseks on subjektiivne tõenäosus otseselt seotud otsustusteooria (*decision theory*) [**?**] ja mänguteooriaga[**?**]. Edasise huvi rahuldamiseks soovitame monografiaid [**?**, **?**].

3 Erinevad lahendusmeetodid

Detailidesse laskumata võib öelda, et Bayesi statistika pole midagi muud kui Bayesi teoreemi rakendamine. Olenemata analüüsi tüübist saame alati tulemuseks tõenäosusjaotuse. Sealjuures on oluline vaid jaotuse kuju mitte selle alla jääv pindala. Sageli loobutakse jaotuse normaliseerimisest, sest kontstantsete kordajate hülgamine lihtsustab oluliselt avaldisi.

Järgnevalt vaatleme üht lihtsaimat katseskeemi—sõltumatute katsete seeriat. Vaatleme näite korras Mendeli eksperimenti hernestega [**?**]. Teatavasti määrab herneste värvi üks geenide paar. Kui üks geenidest on dominantne, siis on herned kollased, vastasel korral rohelised. Olgu kahe taime ristanditele vastavad värvid $d_1, d_2, \dots, d_n$, millist tüüpi geenidega on ristatud taimed, kui mõlemad neist on kollaste hernestega.

Olgu dominantne geen A ja retsesiivne geen a, siis on lihtne arvutada erinevate ristatavatele vastavaid tõenäosusi

$$\begin{array}{lclll} p(r|AA,AA)=0 & \multirow{2}{*}{$\dots$} & p(r|Aa,Aa)=1/4 & p(r|Aa,aa)=1/2 & p(r|aa,aa)=1 \\ p(k|AA,AA)=1 & & p(k|Aa,Aa)=3/4 & p(k|Aa,aa)=1/2 & p(k|aa,aa)=0 \end{array}$$

Kuna katsed y_i on sõltumatud, siis on kerge leida

$$p(d_1,\dots d_n \mid \mathcal{M}) = p(d_1 \mid \mathcal{M})\cdots p(d_n \mid \mathcal{M}),$$

kus $\mathcal{M}$ on võimalik ristatavate kombinatsioon. Ülesande täielikuks lahendamiseks peame määrama eelneva tõenäosusjaotuse $p(\mathcal{M})$. Kui loeme kõik paarid $\mathcal{M}$ võrdtõenäolisteks, siis on vastuseks tõenäosusjaotus

$$p(\mathcal{M} \mid d_1,\dots d_n) = p(d_1 \mid \mathcal{M})\cdots p(d_n \mid \mathcal{M}),$$

muidu tuleb arvestada lisaliikmega ning

$$p(\mathcal{M} \mid d_1,\dots d_n) \propto p(d_1 \mid \mathcal{M})\cdots p(d_n \mid \mathcal{M})p(\mathcal{M}).$$

Ülesanne 5 Leida kõikidele paaridele AA, AA, AA, Aa, Aa, AA, Aa, Aa, Aa, aa, aa, Aa ja aa, aa vastavad lõpptõenäosused, kui ristanditele vastavad värvid on r, k, r, r, k, k ja kõik paarid on võrdtõenäolised.

Kuigi ühtlane jaotus on üks levinumaid eeljaotusi, siis pole see alati sobilik. Näiteks kui me teame, et ristatavad taimed on saadud Aa tüüpi taimede omavahelisel ristamisel on Aa sagedasem kui AA.

Ülesanne 6 Leida vastav eeljaotus ning sellele vastavad lõpptõenäosused.

Kuigi lõpptulemus sõltub eeljaotusest, kehtib üldine kooskõla tulemus. Kui tegelikkusele vastava mudeli $\mathcal{M}_0$ eeltõenäosus on nullist erinev, siis sõltumatute katsete arvu kasvades kasvab $p(\mathcal{M}_0) \to 1$. Tulemust saab üldistada juhule, kus mudelite arv on lõpmatu.

kontrollida sõnastust

Enamasti on mudelis $\mathcal{M}$ ka selliseid parameetreid, mis meid otseselt ei huvita. Seetõttu tuleb tõenäosusjaotust $p(\mathcal{M} \mid D)$ keskmistatakse üle üleliigsete parameetrite. Näiteks kui mudelil on parameetrid x ja y ning meid huvitab vaid parameeter x, siis leiame esmalt $p(D \mid x, y)$ ja $p(x, y)$ ning seejärel integreerime üle y väärtuste

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

Kui parameeter y saab omada vaid diskreetsedid väärtusi $y_1, y_2, \ldots, y_m$, siis kõdub integraal summaks ning vastav avaldis on

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

Ühelt poolt lubab keskmistamine lahendada kõik ette tulevad probleemid, teisalt muudab see arvutused keerukaks—tekkivad integraalid ja summad pole lihtsalt leitavad. See on peamine Bayesi statistika nõrkus.

Teiseks kummastavaks tehnikaks on keskmistamine üle mudelite. Enamasti vajatakse mudelit selleks, et ennustada uusi mõõtmistulemusi. Kuna analüüsi tulemuseks on tõenäosusjaotus üle mudelite mitte optimaalne mudel, siis ennustamiseks tuleb jällegi kesmistada. Olgu ennustatav suurus d, siis

$$p(d \mid D) = \int p(d, \mathcal{M} \mid D) d\mathcal{M} = \int p(d \mid \mathcal{M}) p(\mathcal{M} \mid D) d\mathcal{M}.$$

Ehk teisisõnu kõik mudelid võetakse ennustamisel arvesse kuid iga mudeli kaaluks on selle tõenäosus.

Ülesanne 7 Leida tõenäosused, eelnevalt vaadeldud ristandid on geenipaariga AA, Aa ja aa. Milline on edasisel ristamisel tekkivate geenipaaride tõenäosusjaotus?

4 Põhjuslikud seosed

Bayesi statistika üheks raskemaks probleemiks on teadmiste formaliseerine. Vähegi keerukamatel juhtudel on raske raske või koguni võimatu leida tõenäosusjaotus, 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[1] 123 käiguga?

Neile küsimustele on raske kohest vastust anda, samas suudab iga üks meist hinnata erinevate sündmuste vahelisi põhjuslikke seoseid ning nende tugevust. Nii on lihtne leida järgneva seisu tõenäosus, kui on teada eelnev nuppude paigutus mängulaual. Sama moodi saab analüüsida ka tuumasõja puhkemise erinevaid stsenaariume ning neid käivitavate sündmuste tõenäosust. Peamiseks analüüsi vahendiks on keeruka süsteemi dekomponeerimine arusaadavateks põhjuslikeks seosteks.

Bayesi võrk ongi graafiline mudel, mis kannab endas sündmuste vahelist struktuuri. Loomulik graafiline märgikeel ja erinevad subjektiivse tõenäosuse hindamise metoodikad [**?**] võimaldavad ka matemaatikakaugel eksperdil kirjeldada oma intuitiivseid teadmisi. Selleks piisab põhjuslike seoste kirjeldamisest ning nende varustamisest tõenäosustega. Selliselt loodud ekspertsüsteemid nagu **??? Vaata slaide ???** on sisuliselt deitailsed ?eeljaotuste? kirjeldused. Kui sellele lisada veel ?andmemudel?, siis saab esialgset jaotust täpsustada, see tähendab treenida Bayesi võrku uute andmetega.

Definition 1. Bayesi võrk (*belief network*) on tsükliteta orienteeritud graaf, määrab ära juhuslike suuruste $\mathcal{X} = (X_1, X_2, \ldots, X_n)$?ühendjaotuse?. Graafi tippudeks on juhulikud suurused $X_1, X_2, \ldots, X_n$ ning kaared kirjeldavad põhjuslikke seoseid tippude vahel. Iga juhusliku suuruse X_i ja selle eelaste korral on määratud tinglik tõenäosus $p(X_i|\mathcal{P}a(X_i))$.

Kõik muutujad pole võrdses seisus, sageli vastab Bayesi võrk hierarhilisele mudelile. Hierarhilise mudeli korral jagatakse muutujad mitmesse klassi. Tipud millel pole järglasi on meid otsitavad parameetrid ning ülejäänud tipud moodustavad hüperparameetrite hulga. Analüüsi lõpptulemusena soovime näha otsitavate parameetrite ühisjaotust ning seetõttu peame keskmistama üle hüperparameetrite.

Klassikaliseks näiteks hierarhilisest mudelist on ravimi edukuse hindamine. Lisaks ravile mõjutavad haigete seisundit paljud muud tegurid. Lihtsaim seda

[1] Hetkel ei tule paremat objektiivse tõenäosusega seotud näidet meelde.

arvestav mudel käsitleb eri haiglates toimuvat eraldi. Kuna haigla mõju ravile on *a priori* raske hinnata, siis selle asemel täpsustatakse erinevate haiglate vaheline sarnasus. Tulemuseks on kaks tõenäosusjaotust: $p(y|\Theta)$ ja $p(\Theta|\Psi)$. Neist esimene kirjeldab ravi edukust konkreetsete hüperparameetrite väärtuse korral. Teine kirjeldab kui tõenäoline on olukord, et haiglas valitsevad tingimused Θ. Mudeli graafiline kuju Bayesi võrguna on imelihtne

$$haigla \rightarrow ravimeetod \rightarrow edukus$$

Loomulikult saame me mudelit alati täpsustada näiteks võime arvesse võtta väliskeskonna mõju ravile.

Lihtne on leida mudelile vastavat ühisjaotust, kuid reeglina on keeruline leida sellest edukuse tõenäosust. Peamisteks raskusteks on analüütilise lahenduse puudumine ning suur hüperparameetrite arv. Diskreetesete jaotuste korral langeb esime küsimus ära, 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 algoritm. Küsimus taandub ühisjaotuse faktoriseerimisele—selle esitamisele mitme väiksema jaotise korrutisena.

5 Tingimuslik sõltumatus ja selle omadused

Bayesi võrgu koostamisel ning selle analüüsis on olulisel kohal tingimuslik sõtumatus, mis lubab analüüsida keerukaid seoseid erinevate juhuslike suuruste vahel. See omakorda annab võtme keskmistamise protseduuri optimiseerimiseks.

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

Antud materjali piires tegeleme diskreetsete või absoluutselt pidevate juhuslike suurustega, millel tõenaäosuse määrab täielikult tihedus. 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

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

Mõneti üllatav on tingliku sõltumatuse omadus minna kaotsi keskmistamisel ja edasisel täpsustamisel. Selle omaduse illustreerimiseks vaatame kolme binaarset juhuslikku suurust X_1, X_2 ja X_3. Näites võtame X_3 uhtlase jaotusega, see tähendab $p(X_3 = 1) = 1/2$ ning esitame X_1 ja X_2 ühisjaotuse tabelina. Vasakpoolne tabel vastab juhule kui on teada $X_3 = 0$ ja parem $X_3 = 1$. Esimeses näites on $I(X_1, X_2|X_3)$,

$$1/2\cdot\begin{array}{|c|cc|}\hline & 0 & 1 \\ \hline 0 & 1 & 0 \\ 1 & 0 & 0 \\ \hline \end{array} \quad + \quad 1/2\cdot\begin{array}{|c|cc|}\hline & 0 & 1 \\ \hline 0 & 1/4 & 1/4 \\ 1 & 1/4 & 1/4 \\ \hline \end{array} \qquad \Rightarrow \qquad \begin{array}{|c|cc|}\hline & 0 & 1 \\ \hline 0 & 5/8 & 1/8 \\ 1 & 1/8 & 1/8 \\ \hline \end{array}$$

kuid x_1 ja X_2 liitjaotus pole enam sõltumatu. Teises näites on $D(X_1, X_2|X_3)$,

$$1/2\cdot\begin{array}{|c|cc|}\hline & 0 & 1 \\ \hline 0 & 0 & 1/2 \\ 1 & 1/2 & 0 \\ \hline \end{array} \quad + \quad 1/2\cdot\begin{array}{|c|cc|}\hline & 0 & 1 \\ \hline 0 & 1/2 & 0 \\ 1 & 0 & 1/2 \\ \hline \end{array} \qquad \Rightarrow \qquad \begin{array}{|c|cc|}\hline & 0 & 1 \\ \hline 0 & 1/4 & 1/4 \\ 1 & 1/4 & 1/4 \\ \hline \end{array}$$

kuid X_1 ja X_2 ühendjaotus on sõltumatu.

Juhuslike suuruste omavaheliste seoste analüüsis on oluline teada tingliku sõltumatuse nelja põhiomadust:

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

$$I(A, B|C) \quad \Longleftrightarrow \quad I(B, A|C)$$

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

$$I(A, B_1 \cup B_2|C) \quad \Longrightarrow \quad I(A, B_1|C)$$

(I3) nõrk laiendatavus (*weak union*)—liikmete fikseerimine säilitab sõltumatuse

$$I(A, B_1 \cup B_2|C) \quad \Longrightarrow \quad I(A, B_1|C \cup B_2)$$

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

$$I(A, B|C_1 \cup C_2) \wedge I(A, C_2|C_1) \quad \Longrightarrow \quad I(A, B \cup C_2|C_1)$$

Esimene omadus tuleneb otseselt definitsiooni sümmetriast. Teise omaduse tõestamiseks oletame vastuväiteliselt, et leiduvad väärtustustused $\mathcal{A}$, $\mathcal{B}_1$ 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 hulgaks $\mathcal{B}_2$ juhuslike suuruste B_2 kõikvõimalikud väärtustused on tulemuseks vastuolu tingimusliku 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 tingimuslikud 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

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

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

(I5) lõikemomadus (*intersection*)—järelduvusnool nõrgas laiendatavuses on pööratav

$$I(A, B_1|C \cup B_2) \wedge I(A, B_2|C \cup B_1) \qquad \Longleftrightarrow \qquad I(A, B_1 \cup B_2|C).$$

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

$$I(A, B|C) \quad \Longrightarrow \quad I(A, B|C \cup D)$$

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

$$I(A, B|C) \qquad \Longleftrightarrow \qquad I(A, \{d\}|C) \vee I(B, \{d\}|C).$$

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

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

saame $p(\mathcal{A}|, \mathcal{B}_1, \mathcal{C}) \equiv p(\mathcal{A}|, \mathcal{B}_2, \mathcal{C})$. Teisisõnu ei sõltu avaldis $\mathcal{B}_1$ 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

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

Tugev laiendatavus ning ja transitiivsus on väga spetsiifilised omadused. Võib julgelt öelda, et enamus joutusi ei täida neid omadusi.

6 Sõltumatuste kirjeldamine graafina

Kuna sõltumatus on sümmetriline omadus, siis on loogiline kasutada selle modellerimiseks orienteerimata graafe. Teisest küljest on Bayesi võrk orienteeritud kirjeldus juhuslike suuruste omavahelisest vahekorrast. Siit lähtuvalt on meil tarvis lahendada kaks ülesannet: anda tingliku sõltumatuste kirjeldus Bayesi võrgus ning leida sobiv lihtgraaf, mis kirjeldab sõltumatust.

Tingimusliku sõltumatuse kirjeldamiseks Bayesi võrgus on esmalt otstarbekas defineerida orienteerimata sõltumatuste graaf. Järgnevas on kesksel kohal tipuhulkade kolmiku sõltumatus graafis.

Definition 3. Olgu A, B ja C orienteerimata graafi G tippuhulgad. 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 kolmikut nimetakse sõltumatuks ja tähistakse $I_G(A, B|C)$. Vastasel korral nimetakse kolmikut sõltuvaks ja tähistakse $D_G(A, B|C)$.

Olemuselt on tingimuslik sõltumatus kvalitatiivne ning ei sõltu konkreetsetest tõenäosuse arvväärtustest Bayesi võrgus, vaid nende omavahelisest vahekorrast. Seega on loomulik seda mudeldada lihtgraafina, mille tipuhulgaks on $\mathcal{X}$.

Definition 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 tingimuslikult sõltumatud, see tähendab

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

Kahjuks ilmneb, et igal Bayesi võrgul (ühisjaotusel) puudub täpne sõltumatuste graaf. Näiteks järgmisel joonisel toodud kolm binaarset juhuslikku suurust ei rahulda $I(X_1, X_2|X_3)$ kuid samas on X_1 ja X_2 sõltumatud.

Seega ühelt poolt peaks tippude X_1 ja X_2 vahel olema serv, kuid see läheb vastuollu X_1 ja X_2 sõltumatusega. Täpseid sõltuvuste graafe omavatel jaotustel on olemas lihtne kirjeldus.

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

Korrektuur. Tõestuse võib leida Guiterreze ja Hardi õpikust [**?**]. ☐

Intuitiivsel tasemel põhjendades võimaldab lõikeomadus (I5) leida sõltumatuse graafi. Kuna igas graafis kehtib tugev laiendatavus (I6), siis on see vältimatu. Tugev transitiivsus (I7) väldib olukordi, kus üks tipp sõltub mitmest tipust mis on tingimuslikult sõltumatud.

Enamus jaotusi ei rahulda nii rangeid nõudeid ning seetõttu tuleb defineerida ohutud aproksimatsioonid.

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

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

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

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

Kuna täielikus graafis pole ühtegi mittetriviaalset sõltumatut kolmikut ja nullgraafis pole sõltuvusi, siis leidub igal ühisjaotusel alati ohutud aprksimatsioonid. Ülim eesmärk on leida minimaalne sõltumatuste graaf ning maksimaalne sõltuvuste graaf. Seda realiseerivat algoritmi vaatame järgmises peatükis koos tingimusliku sõltuvuse kirjeldusega.

7 Bayesi võrgu moraal

Bayesi võrgu moraaliks võetakse võimalikult täpne sõltumatuste graaf. Esimeseks sammuks sellel teel on leida minimaalse sõltumatuste graafi kirjeldus. Selleks defineerime kolm erineva rangusega tingimust.

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

(PM) Markovi paaritingimust (*pairwise markov property*), kui puuduv serv tipude X_i ja X_j vahel tähendab, et X_i ja X_j on tinglikult sõltumatud $\mathcal{X} \setminus \{X_i, X_j\}$ suhtes ehk $I_G(X_i, X_j|\mathcal{X} \setminus \{X_i, X_j\}) \Longrightarrow I(X_i, X_j|\mathcal{X} \setminus \{X_i, X_j\})$;

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

$$I_G(X_i, \mathcal{X} \setminus \{X_i\} \,|\mathcal{B}d(X_i)) \Longrightarrow I(X_i, \mathcal{X} \setminus \{X_i\} \,|\mathcal{B}d(X_i));$$

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

Neist omadustest on suhteliselt kergesti kontrollitav vaid Markovi paaritingimus, teiste kontrollimine on suuremate graafide korral raske, kui mitte võimatu. Enne kui uurime, kuidas kontrollida tingimust (PM), vaatame tingimuste omavahelist vahekorda. Imselgelt kehtib implikatsioonde jada (GM) $\Longrightarrow$ (LM) $\Longrightarrow$ (PM). Kuid vasupidised implikatsioonid üldjuhul ei kehti.

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

Korrektuur.
(PM) $\Longrightarrow$ (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 $I(X_i, \mathcal{X} \setminus \{X_i\} \,|\mathcal{B}d(X_i))$.
(LM) $\Longrightarrow$ (GM)
Toome sisse tähistuse $A^c = \mathcal{X} \setminus (A \cup \mathcal{B}d(A))$ ning näitame et iga tipuhulga A korral kehtib $I(A, A^c|\mathcal{B}d(A))$. Esmalt iga tipuhulga A ja B korral

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

See järeldub otseselt nõrgast laiendatavusest ning aheldatavuse omadusest. Teisalt $\mathcal{B}d(A) \cup \mathcal{B}d(B) = \mathcal{B}d(A \cup B) \cup D$, kus $D \in A \cup B$. Et triviaalselt kehtib $I(C, D|\mathcal{B}d(A \cup B) \cup (A \cup B))$, siis lõikeomadusest saame järeldada

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

Kuna lokaalne Markovi omadus kindlustab $I(\{a\}, \{a\}^c | \mathcal{B}d(a))$, saame kõiki selliseid seoseid arvestades $I(A, A^c|\mathcal{B}d(A))$. Vaatleme nüüd graafis olevat sõltumatut kolmikut A, B ja C. Kuna C on eraldaja, siis leidub hulk $A \subseteq A_0$ ja mille piiriks on C. See koosneb kõigist tippudest hulgast A lähtuvatest teedel, mis ei läbi C. Lihtne on taibata $B \subseteq A_0^c$. Seega eelnevast $I(A_0, A_0^c|C) \Longrightarrow I(A, B|C)$. □

Sisuliselt annab teoreem võtme minimaalse sõltumatuste graafi konstrueerimiseks servhaaval. Vaadates Bayesi võrku tundub, et kaar tippude a ja b vahel viitab sõltuvusele $D(a, b|\mathcal{X}\setminus\{a, b\})$. Kuid see pole alati nii, näiteks kui leidub tipp $c = b$, siis $I(a, b|\mathcal{X} \setminus \{a, b\})$. Enamasti viitab tipu c jaotuse sõltumine tipupaarist a ja b, sellele $D(a, b|\mathcal{X} \setminus \{a, b\})$. Kuid seegi pole kindel reegel. Näiteks kui c on võetud ühtlaselt hulgast $\{0, 1, 2\}$ ning binaarsed juhuslikud suurused a ja b on määratud tabeliga.

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

Siis tinglikust sõltumatusest c suhtes järeldub a ja b on sõltumatus , kuid c jaotus siiski sõltub mõlemast tipust ($D(a, \{b, c\})$ ja $D(\{a, c\}, b)$).

Antud kontranäited viitavad sellele, et minimaalse sõltumatuste graafi konstrueerimine Bayesi võrgust topoloogiast lähtudes on võimatu isegi siis kui kehtib lõiketingimus. Seega tuleb meil valida parim võimalik aproksimatsioon.

Definition 8. Bayesi võrgu G moraaliks nimetatakse lihtgraafi G^m, kus iga kahe tipu X_i ja X_j vahel on serv parajasti siis, kui on täidetud üks kahest tingimusest:
(M1) X_i ja X_j vahel on kaar;

(M2) leidub tipp X_k nii, et X_i ja X_j lähtuvad kaared X_k.

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

Korrektuur. On kerge taibata, et tipukolmiku A, B ja C korral Bayesi võrgu moraalis, piisab kui vaatame, vaid tippe millest leidub suunatut 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 $C \subseteq \mathcal{N}d(A \cup B)$. Eelnevat märkust arvestades lõpeb iga suunatud tee hulgast C hulgas A või B. Tähistame A^+ kõigi tippude hulka teedel A ja B^+ tippude hulka teedel B, kus ükski tee ei läbi 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 $C_0 \subseteq \mathcal{N}d(A \cup B)$ ja $C_1, C_2 \subseteq \mathcal{D}e(A \cup B)$ 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, kusjuutes 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 piiriks 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 saame $I(A^+, C_2|C_0 \cup C_1 \cup B^+)$. Võttes C_1 minimaalse võimaliku, peab iga tipp C_1 olema kas A^+ või B^+ vahetu naaber, vastasel korral poleks moraalis A ja B 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 aheldatuse omaduse tõttu saamegi $I(A^+, B^+|C)$. $\square$

Kui me heidame kõrvale lisainformatsiooni, mida saab leida tipuhulkade A ja B järglaste fikseerimisest, siis saame me küllaltki objektiivse sõltumatuse definitsiooni.

Definition 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 tippuhulga $A \cup B \cup C$ eelashulgale vastava võrgu moraalis $(\mathcal{A}n(A \cup B \cup C))^m$.

Uus d-erladatuse mõiste annab loomuliku üldistuse Markovi tingimustele.

Definition 10. Orienteeritud atsükliline graaf tippudega $\mathcal{X}$ rahuldab:

(DP) suunatud Markovi paaritingimust (*directed pairwise Markov property*), kui iga tipupaari X_i ja X_j korral, mille korral puudub suunatud tee X_i ja X_j vahel, on $I(X_i, X_j|\mathcal{N}d(X) \setminus \{X_i, X_j\})$;

(DL) suunatud lokaalset Markovi tingimust (*directed local Markov property*), kui iga tipu X_i ja temast mittepõlvnevad tipud $\mathcal{N}d(X_i)$ on tingimuslikult sõltumatud X_i vahetute eellaste suhtes $\mathcal{P}a(X_i)$ ehk $I(X_i, \mathcal{N}d(X_i)|\mathcal{P}a(X_i))$;

(DG) suunatud globaalset Markovi tingimust (*directed global Markov property*), kui tipuhulkade A ja B d-eraldatus hulga C suhtes viitab $I(A, B|C)$.

Erinevalt orienteerimata graafist saab tõestada rohkem samaväärsusi. Jällegi on ahel (DG) $\Longrightarrow$ (DL) $\Longrightarrow$ (DP) triviaalne ning üldjuhul implikatsioon (DP) $\Longrightarrow$ (DL) ei kehti.

Theorem 5. *Suunatud Markovi tingimused rahuldavad (DP) $\Longleftarrow$ (DL) $\Longleftrightarrow$ (DG).*

Korrektuur.
(DL) $\Longrightarrow$ (DG)
Suunatud lokaalsest Markovi tingimusest on lihtne järeldada, et iga juhusliku suuruse X_i korral $p(X_i|\mathcal{N}d(X_i)) = p(X_i|\mathcal{P}a(X_i))$. Sellest saab omakorda järeldada, et iga $\mathcal{N}d(X_i)$ alamhulga $\mathcal{P}a(X_i) \subseteq B$ korral $p(X_i|B) = p(X_i|\mathcal{P}a(X_i))$.

Seda arvestades on lihtne veenduda, et me saame graafi muuta Bayesi võrguks lisades vajalikud tõenäosused tippudesse. Kuna Bayesi võrgu moraal on sõltumatuste graaf, siis d-eraldatuse definitsioonist saame koheselt järeldada tipuhulkade A, B tinglikku sõltumatust C suhtes. □

Bayesi võrgu konstrueerimisel kasutatakse tingimust (DL)—otsesteks eelasteks loetakse need juhuslikud suurused, mille fikseerimisel tõenäosusjaotus ei muutu. Teoreem ütleb, et see on teatud mõttes optimaalne viis sõltuvuste avastamiseks. Viimase tulemusena anname d-eraldatuse kirjelduse.

Theorem 6. *Bayesi võrgu tipuhulgad A ja C on d-eraldatud tipuhulga C suhtes, kui iga tee korral alusgraafis hulgast A hulka B on täidetud üks kahest tingimusest:*
(D1) teel leidub tipp $x_i \in C$ nii, et nooled pole vastakuti;

(D2) teel leidub tipp x_i, mis pole C eellane ja nooled on vastakuti.

Korrektuur. Tingimus (D1) välistab sama tee graafi moraalis. Tingimus (D2) kindlustab, et tippu x_i ei vaadelda sõltumatuse kontrollil. Lihtne on taibata, et ka vastupidine implikatsioon kehtib. □

8 Triangulatsioon ning klikipuu

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

Theorem 7. *Sõltumatuste graafi servade lisamisel jääb graaf ikkagi sõltumatuste graafiks.*

Korrektuur. Tõestus saadakse induktsiooniga üle lisatud servade.

□

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

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

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. Seda üldistades loeme faktoriseeringu triviaalseks kui indeksite hulgad saab järjestada nii, et $\mathcal{I}_1 \subseteq \mathcal{I}_2 \subseteq \ldots \subseteq \mathcal{I}_r$.

Kerge on taibata, et sellise Bayesi võrgu moraal on täielik graaf. Sõltumatuste graafis mängivad klikid analoogset rolli—nad viitavad juhuslike suuruste komplektidele, mille ühisjaotusi pole mõtet faktoriseerida. Lõikeomaduse ja minimaalse sõltumatuste graafi korral on hinnang täpne, kuid moraali korral võib osa komplekt siiski edukalt faktoriseeruda.

Ilmneb, et kui tihedusfunktsioon on igas punktis positiivne, saab ühisjaotust alati esitada klikkidele väärtustustele vastavate potensiaalide korrutisena.

Theorem 8 (Grimmett). *Kui ühisjaotuse tihedusfunktsioon $p(X_1, \ldots, X_n)$ on rangelt positiivne, siis leiduvad potensiaalifunktsioonid $\phi_C : X_C \to \mathbb{R},\ C \in \mathcal{C}$ nii, et*

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

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

Korrektuur. Täielik tõestus on toodud lisas ?? ning vastab Grimmetti originaal tõestusele [**?**, **?**]. $\square$

Sõltumatuste graafi koos klikkide potensiaalifunktsioonidega nimetatakse Markovi võrguks. Seetõttu tuntakse tulemust tuntakse Gibbsi jaotuse ja Markovi väljade ekvivalentsiteoreemina, mille tõestasid esimesena Hammersley ja Clifford [**?**]. Tõestus annab küll ekplisiitse potenssiaalifunktsioonide valemi, kuid sellest hoolimata pole ϕ_C kergesti hoomatavat tähendust. Seetõttu vähendame veelgi moraali täpsust, et saaks faktoritele anda konkreetse tähenduse. Selleks vatame erikujulisi moraali graafe.

Definition 11. Graafi klikipuuks nimetatakse maksimaalsete klikkide jadale $C_1, C_2, \ldots, C_m$ vastavate tippudega puud, kus iga C_k korral leidub klikk C_j, mis sisaldab tippude ühisosa $C_k \cap (C_1 \cup \cdots \cup C_{k-1})$ (jagamatu lõike tingimus) ning see vastab servale C_k ja C_i vahel. Servaldele vastavaid ühisosasid S_{ij} nimetatakse eraldushulkadeks (*sepset*) ja hulki $R_i = C_i \setminus S_i$ jääkhulkadeks (*residues*).

Igal graafil pole klikipuud, väikseimaks näiteks on neljane tsükkel. Ilmneb, et pikemate neljaste tsüklite lõhkumine kõõludega on tarvilik ja piisav.

Definition 12. Graaf on trianguleeritud, kui iga tsükkel, mille pikkus on suurem kui kolm, omab kõõlu.

Theorem 9. *Graafil on kilikipuu parajasti siis, kui ta on trianguleeritud.*

Korrektuur. Kui graafis on ilma kõõluta tsükkel, mille pikkus on suuremn kui 3, siis tekkib tsüklit sisaldavatest klikkidest tsükkel. Ning klikipuud pole võimalik konstrueerida.

Kui klikkidest pole võimalik konstrueerida puud, siis peab erinevate lõigetega klikkide jada moodustama vähemalt neljase tsükkli (kolmene tsükkel vastab klikile ja vastoulu ei teki!). Vaatame lühimat sellist klikkide tsükklit, kuid siis on sama pikk tsükkel ka tippude vahel. Kuna graaf on trianguleeritud, siis leidub kõõl ja me oleme saanud vastuolu tsükli minimaalsusega. □

Niisiis selleks, et saada Bayesi võrgu moraalist graafi, millel on klikipuu tuleb seda trianguleerida. Loomulikult soovime leida minimaalse servade arvuga trianguleeritud graafi. Kahjuks on minimaalse trianguleeritud graafi leidmine $\mathcal{NP}$-raske ülesanne ning sellepärast kasutatakse erinevaid ahneid algoritme. Enamasti kasutatakse Kj{aerulff'i välja tõõtatud heuristilisi algoritme[**?**]. Vaatleme nüüd klikipuu täieliku lahutuse $\mathcal{T} = \mathcal{T}_1 \dot{\cup} \mathcal{T}_2$ peamist omadust.

Lemma 1. *Iga klikipuu täielikule lahutusele vastavate juhuslike suuruste hulgad A_1 ja A_2 on tingimuslikult sõltumatud alampuude vahelise eraldushulga S_{ij} suhtes.*

Korrektuur. Oletame vastuväiteliselt, et tipuhulkade A_1 ja A_2 vahel on tee, mis ei läbi S_{ij}. Siis peab leiduma üks serv, mis on A_1 ja A_2 vahel sest $S_{ij} \cup A_1 \cup A_2 = \mathcal{X}$. Seega on veel kaks klikki $C_k \in \mathcal{T}_1$ ja $C_l \in \mathcal{T}_2$ nii, et $C_k \cap C_l \not\subseteq C_j \cap C_j$. Kuid see on vastuolus klikipuu definitsiooniga, tekib klikkidest tsükkel. □

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

Theorem 10. *Kui sõltumatuste graafile vastab klikipuu, siis faktoriseerub ühendjaotus järgnevalt*

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

Korrektuur. Üldsust kitsendamata vaatame sidusaid sõltumatuste graafe. Tõestuseks piisab kui näidata võrdust $p(R_k|R_{k-1},\ldots,R_1) = p(R_k|S_k)$. Vaatleme nüüd vastavat klikipuuud. Lemmast 2 saame, et $I(C_k, C_1 \cup \cdots C_{k-1}|S_k)$, seega $p(C_k, C_1 \cup \cdots C_{k-1}) = p(S_k)p(C_1 \cup \cdots C_{k-1}|S_k)p(C_k|S_k) = p(C_1 \cup \cdots C_{k-1})p(R_k|S_k)$. □

Tulemus ütleb, et teades klikkidele vastavaid potensiaale on lihtne leida suvalistele tippuhulkadele vastavaid marginaaljaotusi, mis on kuuluvad ühte klikki. Mitme kliki korral on see veidi keerulisem, kuid siiski võrratult lihtsam kui esialgse ühisjaotuse keskmistamine.

9 Klikkidele vastavate ühisjaotuste leidmine

Eelnevad tulemused kindlustavad klikipuu olemasolu. Kuigi maksimaalsete klikkide ledmine on üldjuhul $\mathcal{NP}$-raske, muudab graafi trianguleerimine ülesande oluliselt lihtsamaks. Selleks kasutatakse Golumbic'u [**?**] poolt välja töötatud algoritme, mis põhinevad tippude hulda erilisl järjestusel[**?**, **?**]. Ka klikipuu enda konstrueerimisel on mõned nipid, mis vähendavad edasist arvutusmahtu. Vastavad algoritmid on toodud ülevaate materjalides [**?**, **?**] ning nendega tuvumise rõõmu jätame lugejale.

Klikkidele vastavad potensiaalid rahuldavad järgmisi tingimusi ning ilmneb, et need on piisavad ning tarvilikud.

Theorem 11. *Olgu $\mathcal{T}$ ühisjaotusele vastav klikipuu, siis klikipuu potensiaalid rahuldavad järgmisi tingimusi:*

*(LK) lokaalne kooskõla (*local consistency*)—iga kliki C ja sellega lõikuva serva S korral*

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

*(GK) globaalne kooskõla (*global consistency*)—potensiaalid vastavad ühisjaotusele*

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

Need tingimused on tarvilikud ja piisavad potensiaalifunktsioonide fikseerimiseks.

Korrektuur. Tarvilikuseks panema tähele, et $\phi_C(X) = p(X_C)$ saame keskmistades

$$\int p(X_C)dX_R = p(X_S), \qquad R = C \setminus S;$$

ning globaalne kooskõla on ilmne. Kuna potensiaalifunktsioonid, mis rahuldavad lokaalset kooskõskõla defineerivad ühe ühisjaotuse vastavalt valemile (**??**). Siis globaalne kooskõla tingimus kindlustab õige ühisjaotuse tekke. □

Kui muutujad $\mathcal{X}$ on diskreetsed, siis kõdub integraal summa märgiks. Sellel tulemusel põhineb järgmine klikipuu potensiaalide määramise algoritm, mis koosneb kolmest järjestikusest operatsioonist: potensiaalide initsialiseerimine ja lokaalse kooskõla tagamine.

Kui juhuslike sauuruste väärtused on diskreetsed, siis saab tõenäosusjaotusi hoida tabelites ning vajalik algoritm on lihtsasti realiseeritav. Seetõttu vaatame edaspidi vaid diskreeteid juhuslikke suurusi.

Initsialiseerimiseks võetakse $\phi_C \equiv 1$ ning iga tipu X_i korral korrutatakse vastava kliki potensiaali $\phi_C(X) = \phi_C(X)p(X_i|\mathcal{P}a(X_i))$. Moraali konstruktsioon kindlustab, tipu eelaste ühte klikki kuulumist. Eraldajahulkade potensiaalid võetakse $\phi_S \equiv 1$. Pärast initsialiseerimist kehtib globaalne kooskõla kui lokaalne kooskõla on rikutud.

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

1. **Keskmistamine**. Salvetakse eraldushuga potensiaali tabel ϕ_S^{old} ning leitakse uus tabel keskmistamisel
$$\int \phi(X_C)dX_R = \phi(X_S), \qquad R = C \setminus S;$$
2. **Neelamine**. Kliki D potensiaali muudetakse
$$\phi_D = \phi_D \frac{\phi_S}{\phi_S^{old}}$$

Erandina loetakse $0/0 = 0$, see kindlustab korrektse tulemuse [**?**]. Kuna ühe teate edastamisel muutub vaid eraldajahulga ning teadet vastuvõtva kliki potensiaal, siis võrdus
$$\frac{\phi_D^{old}}{\phi_S^{old}} = \frac{\phi_D^{old}}{\phi_S^{old}} \frac{\phi_S}{\phi_S^{old}} \frac{\phi_S^{old}}{\phi_S} = \frac{\phi_D}{\phi_S}$$
kindlustab globaalse kooskõla säilimise.

Sõnumite saatmise järjekorra määramiseks valitakse puus juurtipp ning tõestusmaterjali (*evidence*) kogumiseks saadavad esmalt lehed teated vahetippudele ning seejärel vahetipud järgmistele vahetippudele, kuni jõutakse juureni. Oluline on sealjuures vaid see, et iga tipp saadab teate pärast seda kui kõik tema alluvad on talle teated edastanud. Teises faasis saadab juur oma alluvatele teate ning need saadavad seda edasi oma alluvatele. Siin on oluline, et iga tipp saadaks teate edasi ainult pärast seda kui talle on teade tulnud.

Theorem 12. *Pärast tõestusmaterjalik kogumist ning selle laialisaatmist on rahuldatud ni globaalne kui lokaalne kooskõla tingimus.*

Korrektuur. Tõestuseks on indukstsioon üle klikipuu tippude arvu. Vaatame vaid tõestuse olulisemat fragmenti—kahe tipu korral viib sõnumite vahetus tulemuseni. Olgu meil

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

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

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

Täielikus tõestuseks tuleb tähele panna kahte asja. Esiteks uue ϕ_S poolt tekitatud muutus on tipust väljuvate servade suhtes konstant. Teiseks tõestus läheb läbi suvaliste tipuhulkade $C \cap D = S$ korral, see tähendab nee ei pea olam klikid. Kuid täpsemad detailid jäteme lugejale [**?**]. □

Saadud tulemused võimaldavad meil optimeerida eksperdi poolt kirjeldatud eelteadmisele vastava meid huvitava tõenaäosusjaotuse arvutamist. Muuseas kui teateid õnnestub töödelda sümbolarvutuse võtetega, siis saab sama algoritmi rakendada ka pidevatele jaotustele. Eriti heas seisus on normaal ehk gaussi jaotus, mis on kinnine korrutamise ja keskmistamise suhtes!

10 Bayesi võrgu treenimine

Mõningates olukordades puudum meil piisav teadmus probleemist ning me soovime andmeid kasutada õppimiseks. Näiteks osakab ekspert koostada vaid põhjuslikke seosed kandva Bayesi võrgu kuid jääb hätta tinglike tõenäosuste määramisel. Või mis veelgi hullem pole teada isegi Bayesi võrgu täpset struktuuri. Vastavalt Bayesi statistika põhi printsiibile tuleb leida igale struktuurile vastav tõenäosus ning keskmistada üle kõigi mudelite vastavalt leitud kaaludele. Reegline pole selline lähenemine arvutuslikult võimalik ning seetõttu kasutatakse MAP (*maximum a posteriori*) hinnangut, see tähendab leitakse mudel mille tõenäosus peale andmete registyreerimist on suurim. Kuid ka selliste mudelite leidmine on mittetriviallne ning seetõttu kasutatakse ligikaudseid MonteCarlo meetodeid, mis genereerivad mudeleid vastavalt nende tõenäosusjaotustele. Tuntumad meetodid põhinevad Gibbsi ja Metropolise[**?**] algoritmidel. Teiseks edukaks lähenemiseks Bayesi võrgu parameetrite fikseerimisel on EM-algoritm [**?**]

Teiseks huvitavaks küsimuseks on tõestusmaterjalik kasutamine. Bayesi võrk kodeerib teatavasti meie usku juhuslike suuruste erinevate väärtuste esinemise tõenäosuse kohta. Vastavalt Bayesi printsiibile saame me parandada hinnangut meid huvitava suuruse tõenäosusjaotuse kohta, kui me registreerime teiste suuruste väärtusi. Näiteks saab erinevate sümptomite esinemisest järeldada haiguse tõenäosust.

Tähistame registreritud suuruste hulka $\mathcal{X}_{obs}$ ja neile vastavaid väärtusi x_{obs}. Kui meid huvitab tõenäosus $p(V|\mathcal{X}_{obs} = x_{obs})$, siis vastavalt Bayesi printsiibile

$$p(V|\mathcal{X}_{obs} = x_{obs}) = \frac{p(V, \mathcal{X}_{obs} = x^{obs})}{p(\mathcal{X}_{obs} = x_{obs})}$$

ning selle arvutamise saab kergesti lülitada kliki potensiaalide leidmise algoritmi. Ainus mis muutub on initsialisatsiooni samm. Kui $X_i \in \mathcal{X}_{obs}$, siis loeme kliki $C \ni X_i$ potensiaali

$$\phi_C(X) = 0, \qquad X_i \neq x_i^{obs}.$$

Päärast seda muutust rahuldab algjaotus globaalne kooskõla tingimust $p(\mathcal{X}, \mathcal{X}_obs = x^{obs})$ ning pärast lokaalse tingimuse rahuldamist saame jaotusele $p(\mathcal{X}_{unobs}, \mathcal{X}_obs = x^{obs})$ vastavad potensiaalid. Nüüd

$$p(V|\mathcal{X}_{obs} = x^{obs}) = \frac{p(V, \mathcal{X}_obs = x^{obs})}{\int p(V, \mathcal{X}_obs = x^{obs} dV)}$$

ehk mei tuleb teha kaks kesmistamist üle klikkide.

Seda algoritmi saab kergelt modifitseerida nii, et arvutuste vahetulemused *cache*-takse ja iga uue vaatlustulemuse jaoks pole tarvis otsast peale alustada. Lisaks sellele on olemas veel mitmeid konkreetseid nippe mis muudavad arvutuseeskirja veelgi efektiivsemsks täpsemalt vaata Huang ja Darwiche ülevaate artiklit [**?**]

11 Lisa 1: Elemente graafiteooriast

12 Lisa 2: Mööbiuse inversioon

Kuigi tavaliselt räägitakse Mööbiuse inversioonivalemis vaid alamhulkade ja sisalduvusseose kontekstis, siis tegelikult saab Mööbiuse inversiooni kasutada igasuguse osalise järjestuse korral.

Theorem 13. *Olgu S osaliselt järjestatud hulk, siis funktsioonid $F, G : S \to \mathbb{R}$ rahuldavad tingimust*

$$\forall y \in S: \quad F(y) = \sum_{x \le y} G(y) \quad \Longleftrightarrow \quad \forall y \in S: \quad G(y) = \sum_{x \in S} \mu(y,x) F(y),$$

kus μ on üheselt määratud Mööbiuse funktsioon.

Theorem 14. *Igas osaliselt järjestatud hulga S korral saab Mööbise funktsiooni μ üheselt määrata seosestest $\mu(x,y) = 0$, kui $x \not\geq y$ ning*

$$\sum_{x \le y \le z} \mu(z,y) = \begin{cases} 1, & \textit{kui } x = z, \\ 0, & \textit{kui } x \neq z. \end{cases}$$

Korrektuur. Teoreemide tõestused võib leida kombinatoorika õpikuist[**?**] □

Vaatleme nüüd meile olulist erijuhtu, kus hulga S moodustab graafi klikkide hulk $\mathcal{C}$ koos kõigi tippude hulgaga $\mathcal{X}$. Lihtne on mõista, et sisalduvusseos indutseerib osalise järjestuse hulgal S. Kuna iga kliki alamhulk on klikk, siis iga ahela $A \subseteq C$ korral

$$\sum_{A \subseteq B \subseteq C} (-1)^{|C|-|B|} = \sum_{k=0}^{n} (-1)^k \binom{n}{k} = (1-1)^n,$$

kus $n = |C \setminus A|$, seega iga klikipaari $B \subseteq A$ korral $\mu(A,B) = (-1)^{|A|-|B|}$. Olgu $\mathcal{C}^\circ$ maksimaalsete klikkide hulk, siis on lihtne veenduda $\mu(\mathcal{X},\mathcal{X}) = 1$ ja $\mu(\mathcal{X}, C) = -1,\ C \in C^\circ$. Kerge on taibatA

Ning seega võtab Mööbiuse inversioonivalem kuju

$$G(A) = \sum_{\substack{B \in \mathcal{C} \\ B \subseteq A}} (-1)^{|A|-|B|} F(B), \quad A \in \mathcal{C}.$$

13 Lisa 3: Grimmetti teoreemi tõestus

Korrektuur. Tõestuseks, toome sisse logaritmilised potensiaalid $v_A : X_A \to \mathbb{R}$, kus $A \subseteq S$ ja $S = \{1, \ldots, n\}$. Selleks fikseerime ühe võimaliku väärtustuse x^0 ja defineerime projektsioonid $\pi_A(X) = (X'_1, \ldots, X'_n)$ nii, et $X'_i = X_i$ kui $i \in A$ ja $X'_i = x^0_i$ kui $i \notin A$. See võimaldab otsida võrdustele

$$\log p(\pi_A(X)) = \sum_{B \subseteq A} v_B(X), \qquad A \subseteq S$$

vastavaid potensiaale. Kuna alamhulkadehulk moodustab osalise järjestuse, siis süsteemi lahend on määratud Mööbiuse inversioonivalemiga (**??**)

$$v_A(X) = \sum_{B \subseteq A} (-1)^{|A|-|B|} \log p(\pi_B(X)).$$

Projektsiooni omaduste tõttu on v_A tegelikult määratud X_A väärtustega. Tõestuse lõpetamiseks tuleb näidata $v_A \equiv 0$, kui A pole graafi klikk. Selleks pole vaja midagi muud kui toorest jõudu. Kui A pole klikk, siis leiduvad tipud X_i ja X_j, mille korral $I(X_i, X_j | \mathcal{X} \setminus \{X_i, X_j\})$. Olgu $A^\circ = A \setminus \{i, j\}$, siis

$$\begin{aligned} v_A(X) = & \sum_{B \subseteq A^\circ} (-1)^{|A^\circ|-|B|} \log p(\pi_B(X)) - \sum_{B \subseteq A^\circ} (-1)^{|A^\circ|-|B|} \log p(\pi_{B \cup \{i\}}(X)) \\ & - \sum_{B \subseteq A^\circ} (-1)^{|A^\circ|-|B|} \log p(\pi_{B \cup \{j\}}(X)) + \sum_{B \subseteq A^\circ} (-1)^{|A^\circ|-|B|} \log p(\pi_{B \cup \{i,j\}}(X)) \\ = & \sum_{B \subseteq A^\circ} (-1)^{|A^\circ|-|B|} \log \frac{p(\pi_B(X)) p(\pi_{B \cup \{i,j\}}(X))}{p(\pi_{B \cup \{i\}}(X)) p(\pi_{B \cup \{j\}}(X))}. \end{aligned}$$

On ülimalt oluline veenduda, et logarimi võetakse alati ühest. Selleks tuleb hoolikalt dekodeerida krüptiline kirjapilt. Selguse mõttes olgu $B = \{1, \ldots, m\}$, $i = m + 1$ ja $j = m + 2$, siit saame

$$\begin{aligned} p(\pi_B(X)) &= p(X_1, \ldots, X_m, x^0_{m+1}, x^0_{m+2}, \ldots, x^0_n) \\ p(\pi_{B \cup \{i\}}(X)) &= p(X_1, \ldots, X_m, X_{m+1}, x^0_{m+2}, \ldots, x^0_n) \\ p(\pi_{B \cup \{j\}}(X)) &= p(X_1, \ldots, X_m, x^0_{m+1}, X_{m+2}, \ldots, x^0_n) \\ p(\pi_{B \cup \{i,j\}}(X)) &= p(X_1, \ldots, X_m, X_{m+1}, X_{m+2}, \ldots, x^0_n) \end{aligned}$$

Kõiki liikmeid saab jagada läbi $p(X_1, \ldots, X_m, x^3_{m+3} \ldots, x^0_n)$ ning järele jäävad järele korrutised

$$\begin{aligned} & p(x^0_{m+1} | X_1, \ldots, X_m, x^0_{m+3} \ldots, x^0_n) p(x^0_{m+2} | X_1, \ldots, X_m, x^0_{m+3} \ldots, x^0_n) \\ & p(X_{m+1} | X_1, \ldots, X_m, x^0_{m+3} \ldots, x^0_n) p(x^0_{m+2} | X_1, \ldots, X_m, x^0_{m+3} \ldots, x^0_n) \\ & p(x^0_{m+1} | X_1, \ldots, X_m, x^0_{m+3} \ldots, x^0_n) p(X_{m+2} | X_1, \ldots, X_m, x^0_{m+3} \ldots, x^0_n) \\ & p(X_{m+1} | X_1, \ldots, X_m, x^0_{m+3} \ldots, x^0_n) p(X_{m+2} | X_1, \ldots, X_m, x^0_{m+3} \ldots, x^0_n), \end{aligned}$$

mis taanduvad nimetajast lugejast välja. Sama tõestus läheb läbi suvalise B korral ning seega tõesti $v_A \equiv 0$. See aga tähendab

$$p(X_1, X_2, \ldots, X_n) = \prod_{c \in \mathcal{C}} \exp(v_C(X_C)).$$

Viimaseks näitame, et summa üle kõigi klikkide $\mathcal{C}$ võib asendada summaga üle maksimaalsete klikkide $\mathcal{C}_{max}$ Võrduse hoolikamast lahtikirjutamisest

$$\log p(X) = \sum_{C \in \mathcal{C}} \sum_{D \subseteq C} (-1)^{|C|-|D|} \log p(\pi_D(X)) = \sum_{D \in \mathcal{C}} \sum_{\substack{C \in \mathcal{C}: \\ D \subseteq C}} (-1)^{|C|-|D|} \log p(\pi_D(X))$$

on näha, et D panus summasse on 0, kui D pole masimaalne klikk, sest

$$\sum_{\substack{C \in \mathcal{C}: \\ D \subseteq C}} (-1)^{|C|-|D|} = (1-1)^n,$$

kus $n = |E \setminus C|$ ja $C \subseteq E$ onmaksimaalne klikk.