Bayes-verkkojen rakenteen bayesiläinen tarkastelu

Mikko Koivisto

15. maaliskuuta 2001

1 Johdanto

Bayes-verkko on graafinen esitys äärellisen satunnaismuuttujajoukon yhteistodennäköisyysjakaumalle. Se koostuu satunnaismuuttujien tilastolliset riippuvuudet määrittelevästä *verkkorakenteesta* sekä joukosta reaaliarvoisia parametreja, *parametrivektorista*. Bayes-verkot tarjoavat siten yleisen malliperheen, jota voidaan soveltaa äärellistä havaintoaineistoa hyödyntävään tilastolliseen päättelyyn.

Bayes-verkkojen soveltajat ovat yleensä bayesiläisiä. Niinpä tilastollisia päätelmiä ei yritetä perustaa sellaiseen yhteen verkkoon, joka antaa havannoille suurimman todennäköisyyden.[1] Sen sijaan pyritään keskiarvoistamaan vaihtoehtoisten Bayes-verkkojen yli painotettuna verkkoihin liitetyillä prioriuskomuksilla.

Koska keskiarvoistaminen on kuitenkin yleensä laskennallisesti erittäin vaativaa, täydellisestä bayesiläisestä päättelystä usein luovutaan. Karkeimmillaan Bayes-verkon rakenne kiinnitetään perustuen asiantuntijatietämykseen tai muihin kriteereihin. Tällöin keskiarvoistamista sovelletaan vain verkon parametrien arvoihin. On myös kehitelty menetelmiä, joilla havaintoaineiston ja prioriuskomusten valossa "todennäköisin" verkkorakenne voidaan tehokkaasti oppia – sitä siis ei tarvitse ennalta kiinnittää. Tällöinkin keskiarvoistamista sovelletaan vain verkon parametrien arvoihin. Tällaiset menetelmät soveltunevat uusien havaintojen ennustamiseen mutta eivät välttämättä muuttujien välisten riippuvuuksien päättelemiseen [3]. Ongelmia aiheutuu, jos verkkorakenteiden joukosta löytyy aivan erilaisia mutta keskenään yhtä hyviä rakenteita. Näin on tyypillisesti silloin, kun muuttujia on useita mutta aineistoa on vähän [3].

[1] Tämä on klassisessa tilastotieteessä suosittu suurimman uskottavuuden periaate (engl. maximum likelihood principle).

Friedman ja Koller mainitsevat esimerkillisenä sovellusongelmana geeniekspressioanalyysiin [3, 4]. Siinä havaintoaineisto muodostuu suurelle geenijoukolle yhtaikaisesti mitatuista ekspressiotasoista vaihtelevissa olosuhteissa. Olosuhde voi olla esimerkiksi solutyyppi (lihassolu, hermosolu jne.) tai jonkin biologisen prosessin vaihe. Muuttujat koostuvat geeneistä (tarkemmin niiden ekspressiotasoista) ja olosuhdetta kuvaavista suureista. Mitattavia geenejä on tyypillisesti paljon (tuhansia), mutta erilaisia mittausolosuhteita on käytännän syiden (aika ja raha) vuoksi vain muutamia, korkeintaa satoja.

Viimeaikoina on suunnattu kohti täydellisempää bayesiläistä päättelyä – pyritään keskiarvoistamaan (myös) yli verkkorakenteiden. Vaikka kokonaisen verkkorakenteen päättely ei olisikaan mielekästä, keskiarvoistamalla yli rakenteiden voidaan yrittää päätellä joitakin "pienempiä piirteitä", kuten ovatko tietyt kaksi muuttujaa vahvasti korreloituneita [3]. Laskennallisten ongelmien voittamiseksi on ehdotettu erilaisia rajoituksia ja likimääräismenetelmiä. Madigan ja York [7] esittivät Markovin ketju Monte Carlo -simulaatiota (lyh. engl. MCMC) hyödyntävän menetelmän; parannusta tähän tarjoavat Friedman ja Koller [3].

Tämän kirjoitelman keskeisin tavoite on selvittää lukijalle, mistä Friedmanin ja Kollerin artikkelissa [3] on kysymys. Tätä varten ensin tarkastellaan bayesiläistä lähestymistapaa yleisesti (osio 2) sekä selvitetään MCMC-simulaation perusajatus (osio 3). Kirjoitelman loppuosassa siirrytään varsinaisesti Bayes-verkkoihin. Osiossa 4 muotoillaan Friedmanin ja Kollerin tarkastelema ongelma, minkä jälkeen osiossa 5 esitetään heidän tarjoamansa ratkaisu. Friedmanin ja Kollerin kokeellisia tuloksia selvitetään tiivistäen osiossa 6. Viimeinen osio (osio 7) sisältää kirjoittajan omia yhteenvetäviä näkemyksiä.

2 Bayesiläisyys

Klassinen eli frekventistinen tilastotiede perustuu todennäköisyyden frekvenssitulkintaan ja siten siinä todennäköisyyden käsite liitetään ainoastaan tapahtumiin, jotka voidaan (luonnollisesti) ajatella toistettavan satunnaiskokeen tuloksiksi. Erityisesti hypoteesien todennäköisyydestä puhuminen ei tämän koulukunnan edustajien mielestä ole mielekästä.

Bayesiläinen tilastotiede nojaa todennäköisyyden subjektiiviseen tulkintaan: maailmasta kertovan lauseen todennäköisyys on luku, joka kertoo, missä määrin tarkasteltava agentti uskoo lauseen paikkansapitävyyteen. Bayesiläinen vertailee vaihtoehtoisten hypoteesien hyvyyttä niiden todennäköisyyksiin nojautuen.

Bayesiläinen lähestymistapa kiteytyy bayesiläisessä päätösteoriassa. Sen mukaan optimaalinen päätös tai teko (epävarmuuden vallitessa) on se, joka tuottaa odotusarvoisesti suurimman hyödyn eli utiliteetin (engl. utility). Odotusarvo otetaan yli subjektiivisen todennäköisyysjakauman. Näin saadaan toimintaperiaate ”älykkäälle” agentille. Kääntäen tämä periaate antaa sisällön bayesiläisen todennäköisyyden käsitteelle. Esimerkiksi kahden kilpailevan hypoteesin (Mika vai Schumi?) todennäköisyydet voidaan jossakin tapauksessa nähdä siitä, millä vedonlyöntisuhteella agentti on valmis sijoittamaan jompaan kumpaan hypoteesiin.

Bayesiläisessä mallintamisessa pyritään tuottamaan järkevä todennäköisyysjakauma päätöksenteon pohjaksi. Tätä varten mallintaja valitsee tarkasteltavakseen joukon sopivia hypoteeseja eli malleja sekä kerää havaintoja, jotka auttavat arvioimaan hypoteesien uskottavuutta. Kuten klassisessa tilastotieteessä kukin hypoteesi M määrittelee havaintojen D todennäköisyyden $P(D \mid M)$. Ennakkouskomusta kuvaava prioritodennäköisyys $P(M)$ voidaan Bayesin kaavaa käyttäen päivittää posterioritodennäköisyydeksi $P(M \mid D)$. Näin subjektiivisuus saadaan palautettua priorien valintaan – todennäköisyyspäättely itsessään noudattaa yleisesti hyväksyttyjä todennäköisyysmatematiikan ”objektiivisia” sääntöjä.

Usein bayesiläinen päättely nojautuu suoraan mallien posteriorijakaumaan. Merkitään posteriorijakaumaa kirjaimella π, siis $\pi(M) \equiv P(M \mid D)$. Bayesiläisessä päättelyssä tarkasteltavat posteriorijakauman piirteet, kuten momentit, kvantiilit ja suuren todennäköisyyden alueet sekä parametrien marginaalijakaumat, voidaan esittää mallien funktioiden odotusarvona. Näin päädytään ongelmaan, jossa on laskettavana odotusarvo

$$E_\pi(f) = \sum_M \pi(M) f(M), \tag{1}$$

missä funktio f määrää kiinnostuksen kohteen. Usein odotusarvo ei ole analyyttisesti laskettavissa tai laskenta on hyvin vaikeaa, minkä vuoksi turvaudutaan numeerisiin menetelmiin.

3 MCMC-menetelmä

MCMC:n keskeisiä käsitteitä ovat Monte Carlo -integrointi ja Markovin ketju. Seuraavassa selvitetään näitä käsitteitä erityisesti MCMC-menetelmän bayesiläistä soveltamista ajatellen. Esitys nojautuu pääosin lähteisiin [5, s. 1–19] ja [8].

3.1 Monte-Carlo -integrointi

Moniulotteisten integraalien numeeriseen arvioimiseen voidaan käyttää satunnaiseen otantaan perustuvaa Monte Carlo -integrointia. Odotusarvoon (1) sovellettuna Monte Carlo -integroinnissa muodostetaan jakaumasta π satunnaisotos $\{X_i \mid i = 1, \dots, n\}$ ja lasketaan likiarvo

$$E_\pi(f) \approx \frac{1}{n} \sum_{i=1}^{n} f(X_i).$$

Otoksen satunnaisuuden vuoksi likiarvo on satunnaismuuttuja. Jos otokset ovat riippumattomia, saadaan likiarvo suurten lukujen lain nojalla mielivaltaisen tarkaksi kasvattamalla otoksen kokoa n.

Useissa tapauksissa riippumattomien otosten muodostaminen ei ole käytännöllistä. Tyypillisesti jakauma π on epätasainen keskittyen joihinkin moniulotteisen avaruuden osiin. Pisteitä satunnaisesti arpomalla vain harvoin osutaan niihin osiin, joissa π on merkittävästi nollaa suurempi. Riippumattomuus ei kuitenkaan ole välttämätöntä: väljästi sanoen riittää, että otokset muodostava prosessi tuottaa arvoja jakauman π mukaisesti sopivissa suhteissa. Yksi keino sopivan otoksen muodostamiseen hyödyntää Markovin ketjua, jonka *tasapainojakauma* on π.

3.2 Markovin ketjut

Markovin ketju on diskreettiaikainen stokastinen prosessi $\{X_0, X_1, \dots\}$, jolle kaikilla $t = 1, 2, \dots$ satunnaismuuttujan X_t jakauma ehdollistettuna prosessin historialle, eli muuttujille $X_0, X_1, \dots, X_{t-1}$, riippuu ainoastaan muuttujasta X_{t-1}. Matemaattisemmin tämä *Markov-ominaisuus* kirjoitetaan

$$P(X_t \mid X_0, X_1, \dots, X_{t-1}) = P(X_t \mid X_{t-1}).$$

Satunnaismuuttujaa X_t kutsutaan prosessin *tilaksi askeleella* t. Markovin ketju on täysin määrätty, kun annetaan alkutilan X_0 jakauma ja tilasiirtymien jakaumat $P(X_t \mid X_{t-1})$. Sovelluksissa suositaan stationäärisiä Markovin ketjuja eli ketjuja, joille todennäköisyys siirtyä tilasta x tilaan y on kiinteä.

Markovin ketjun sanotaan unohtavan alkutilansa, jos tila X_m "tulee riippumattomaksi" alkutilasta X_0, kun m kasvaa rajatta. Tällöin on olemassa yksikäsitteinen *tasapainojakauma* ψ, jolle

$$\lim_{m \to \infty} P(X_m = y \mid X_0 = x) = \psi(y)$$

kaikilla tiloilla x ja y. Bayesiläisissä MCMC-menetelmissä pyritään rakentamaan Markovin ketju, jolla on yksikäsitteisenä tasapainojakaumanaan posteriorijakauma π. Tähän päästään valitsemalla sopiva tilasiirtymäjakauma.

Tilasiirtymäjakauman konstruoimista Metropolisin ja Hastingsin algoritmilla [6] käsitellään jäljempänä osiossa 3.3.

Tasapainojakauma saavutetaan teoriassa vasta äärettömän monen askeleen jälkeen. Monte-Carlo -integrointia ajatellen kohtuullisen hyvä otos tasapainojaukaumasta on kuitenkin $\{X_m, X_{m+1}, \ldots, X_n\}$, mikäli *konvergointijakson* pituus m ja *otantajakson* pituus n ovat riittävän suuria. Jos m on pieni, seurauksena voi olla otoksen vinoutuminen ja siten huono likiarvo. Pieni n voi puolestaan aiheuttaa likimääräistysvirheen (Monte Carlo -virhe). Peräkkäisten näytteiden riippuvuutta on luvallista heikentää ottamalla otokseen mukaan vain joka k:nnes tila jollakin sopivalla k.

3.3 Metropolisin ja Hastingsin algoritmi

Metropolisin ja Hastingsin algoritmi kuvaa tilasiirtymät kaksivaiheisesti. Ensimmäisessä vaiheessa ehdotetaan uutta tilaa; toisessa vaiheessa päätetään, hyväksytäänkö ehdotettu tila vai pysytäänkö edellisessä tilassa. Molemmat vaiheet ovat stokastisia. Tarkastellaan algoritmia tarkemmin kuvassa 1 esitettyyn muotoiluun nojautuen. Algoritmi saa syötteenään jakauman π, aloitustilan X_0 jakauman ϕ ja ehdotusjakauman q sekä iteraatioiden lukumäärän n. Rivillä 1 arvotaan aloitustila X_0 jakauman ϕ mukaisesti. Tämän jälkeen siirrytään kaksivaiheisiin tilasiirtymiin. Rivillä 3 ehdokastila Y valitaan ehdotusjakauman q mukaisesti; askeleella t ehdokastila on ehdollinen tilalle X_t. Kun ollaan tilassa x, saadaan ehdokastilaksi y todennäköisyydellä

$$P(Y_t = y \mid X_t = x) = q(y \mid x).$$

Rivit 4–8 ilmaisevat, että ehdokastila hyväksytään todennäköisyydellä $\alpha(Y \mid X_t)$, jonka Hastings [6] määrittelee asettamalla

$$\alpha(Y \mid X_t) = \min\left\{1, \frac{\pi(Y)q(X_t \mid Y)}{\pi(X_t)q(Y \mid X_t)}\right\}.$$

Jos ehdokastila hylätään, pysytään edellisessä tilassa (rivi 8).

Voidaan osoittaa, että algoritmi konstruoi Markovin ketjun, jonka yksikäsitteinen tasapainojakauma on π. Sovellusten kannalta algorimilla on toinenkin tärkeä ominaisuus: jakauma π tarvitsee tuntea tai osata laskea vain vakiokerrointa vaille, sillä tuntematon kerroin supistuu hyväksymistodennäköisyyttä $\alpha(Y \mid X_t)$ laskettaessa. Tämä ominaisuus osoittautuu hyödylliseksi myöhemmin, kun käsitellään MCMC-menetelmön soveltamista Bayes-verkkojen tapauksessa.

METROPOLIS-HASTINGS(π, ϕ, q, n)
1. valitse $X_0 \sim \phi(.)$
2. **jokaiselle** $t := 1, \ldots, n$ **:**
3. valitse $Y \sim q(\cdot \mid X_t)$
4. valitse $U \sim \mathrm{Tas}(0,1)$
5. **jos** $U \leq \alpha(Y \mid X_t)$ **niin**
6. $X_{t+1} := Y$
7. **muuten**
8. $X_{t+1} := X_t$
9. **palauta** $(X_0, X_1, \ldots, X_n)$

Kuva 1: Metropolisin ja Hastingsin algoritmi. Algoritmi saa syötteenään tasapainojakauman π, aloitusjakauman ϕ, ehdotusjakauman q ja iteraatioiden lukumäärän n. Algoritmi palauttaa simuloinnin tuottaman tilojen jonon.

4 Bayes-verkkojen rakenteen päättely

Tarkastellaan seuraavaa tilastollisen päättelyn ongelmaa. Olkoon annettuna joukko $\{V_1, V_2, \ldots, V_n\}$ ongelmakenttää kuvaavia muuttujia ja joukko D muuttujien havaittuja arvoja. Olkoon mallintajan valitsema hypoteesien joukko Bayes-verkkojen malliperhe. Muistutetaan, että Bayes-verkko on nyt pari $M = (G, \theta_G)$, missä G on n-solmuisen verkon rakenne ja θ_G reaaliarvoisten parametrien vektori. Oletetaan sitten, että mallintaja on kiinnostunut jostakin Bayes-verkkojen rakenteellisesta piirteestä f.

Jos mallintaja on bayesiläinen, hän on halukan laskemaan piirteen posterioritodennäköisyyden

$$P(f \mid D) = \sum_G P(G \mid D) f(G), \tag{2}$$

missä $f(G) = 1$, jos rakenteella G on kyseinen piirre, muuten $f(G) = 0$. Tämä on Fiedmanin ja Kollerin [3] tarkastelema laskentaongelma. Huomataan, että (2) on erikoistapaus yleisestä bayesiläisestä keskiarvoistamisesta (kaava (1)): Tässä f ei riipu parametrivektoreista θ_G lainkaan. Lisäksi f on erikoisesti juuri indikaattorifunktio.

Odotusarvon (2) tarkka laskeminen on vaativaa, sillä n-solmuisten verkkorakenteiden lukumäärä on ylieksponentiaalinen. Verkkorakenteiden läpikäyntiä edellyttää paitsi kaavassa (2) eksplisiittisesti esitetty summaus, niin myös verkkorakenteiden posterioritodennäköisyyksien laskeminen käytettäessä Bayesin kaavaa

$$P(G \mid D) = \frac{P(G)P(D \mid G)}{\sum_{G'} P(G')P(D \mid G')}.$$

Verkkorakenteiden täydellinen läpikäynti on käytännössä mahdollista vain, kun muuttujia (siis solmuja) on enintään kolme. (Summauksen vaativuutta voidaan jonkin verran pienentää rajoittamalla solmujen vanhempien määrä jollakin sopivalla vakiolla. Tällöinkin täydellinen läpikäynti on mahdollista vain, kun solmuja on enintään viisi.) On tyydyttävä likimääräiseen laskentaan!

MCMC-menetelmää voidaan soveltaa suoraviivaisesti, kuten Madigan ja York [7] esittivät. Tätä varten riittää, että posterioritodennäköisyydet $P(G \mid D)$ osataan laskea vakiokerrointa vaille ja riittävän tehokkaasti. Tämä onnistuukin, kun Bayes-verkkojen (G, θ_G) priorit rakennetaan sopivasti modulaarisesti. Tällöin on voimassa $P(G) = \prod_i P(G_i)$ ja $P(D \mid G) = \prod_i P(D_i \mid G_i)$, missä i indeksoi verkkorakenteen G solmut V_i, G_i on solmun V_i vanhempaisjoukko (eli solmun vanhempien joukko) ja D_i käsittää sen osan havantoaineistosta D_i, joka – väljästi sanoen – liittyy solmuperheeseen G_i.[2] (Yksityiskohtien tarkastelu sivuutetaan tässä, ks. [3].)

Friedmanin ja Kollerin [3] mukaan suoraviivainen MCMC-menetelmä on kuitenkin ongelmallinen. He arvelevat, että simulaation sekoittuminen ja konvergoituminen voi olla hidasta, kun verkkorakenteiden avaruus on suuri. Lienee vaikea keksiä ehdotusjakaumaa, jolle posteriorijakauma ei olisi hyvin piikikäs ehdotusjakauman tuottamassa naapurustoavaruudessa.

5 Ehdollistaminen järjestyksellä

Friedman ja Koller [3] esittävät menetelmän, jolla edellä mainittua suoraviivaista MCMC-menetelmää voidaan parantaa. Tätä varten he ensin osoittavat, että rajoituttaessa sopivaan verkkorakenteiden osajoukkoon, eräiden piirteiden posterioritodennäköisyys voidaan laskea tehokkaasti ”suljetussa muodossa”. He näyttävät, että minkä tahansa kiinnitetyn solmujärjestyksen kanssa konsistentit rakenteet muodostavat tällaisen sopivan verkkorakenteiden osajoukon. Seuraavassa tarkastellaan Friedmanin ja Kollerin lähestymistapaa tarkemmin.

[2] Tarkasti ottaen G_i käsittää vanhempaissolmujen lisäksi myös solmun V_i. Tässä asiayhteyden mukaan puhutaan joko vanhempaisjoukosta tai perheestä.

5.1 Suljettu muoto – tehokas laskenta

Olkoon $\prec$ jokin lukujen $1, 2, \ldots, n$ lineaarijärjestys (toisinsanoen jokin lukujen permutaatio). Verkkorakenteen G sanotaan noudattavan järjestystä $\prec$, mikäli rakenteessa G solmu V_i on solmun V_j vanhempi vain jos $i \prec j$. Tehtävänä on nyt laskea piirteen f posterioritodennäköisyys, kun malliperhe koostuu niistä Bayes-verkoista, joiden verkkorakenne noudattaa kiinnitettyä järjestystä $\prec$.

Ehdollisen todennäköisyyden määritelmän nojalla posterioritodennäköisyys voidaan ilmaista osamääränä:

$$P(f \mid D, \prec) = \frac{P(f, D \mid \prec)}{P(D \mid \prec)}.$$

Osoittaja voidaan kirjoittaa summana yli verkkorakenteiden:

$$P(f, D \mid \prec) = \sum_G P(G \mid \prec) P(D \mid G) f(G). \tag{3}$$

Nimittäjä saadaan tämän erikoistapauksena, kun piirteeksi valitaan triviaalisti $f(G) \equiv 1$. Laskennan vaativuutta ajatellen ei vielä ole voitettu mitään: summausten suoraviivainen laskenta vie ylieksponentiaalisen ajan suhteessa muuttujien eli solmujen lukumäärään n.

Järjestyksellä ehdollistaminen tuo mukanaan kuitenkin merkittävän modulaarisuusominaisuuden, jota voidaan tehokkaasti hyödyntää lausekkeen (3) laskemisessa, mikäli piirre f on sopivan yksinkertainen. Oleellinen havainto on, että kun solmujärjestys on kiinnitetty, yhden solmun vanhempien valinta ei mitenkään rajoita sitä, mitkä solmut voidaan valita jonkin toisen solmun vanhemmiksi. Tästä seuraa, että summaus yli verkkorakenteiden voidaan kirjoittaa riippumattomina summauksina yli kunkin solmun valinnaisten vanhempaisjoukkojen: $\sum_G = \sum_{G_1} \sum_{G_2} \cdots \sum_{G_n}$.

Oletetaan nyt, että piirteelle f on erikoisesti voimassa

$$f(G) = \prod_i f_i(G_i). \tag{4}$$

Toisin sanoen rakenteella G on piirre f jos ja vain jos sen kullakin perheellä G_i on piirre f_i. Lauseke (3) saadaan tällöin tehokkaammin laskettavaan muotoon käyttämällä klassista "summa-tulo -temppua" (ks. [1]):

$$\begin{aligned} P(f, D \mid \prec) &= \sum_G \prod_i \Big(P(G_i \mid \prec) P(D_i \mid G_i) f_i(G_i) \Big) \\ &= \prod_i \sum_{G_i} \Big(P(G_i \mid \prec) P(D_i \mid G_i) f_i(G_i) \Big). \end{aligned} \tag{5}$$

Huomautettakoon, että tässä, kuten jo aiemmin, priorien oletaan olevan modulaarisia. Laskennan aikavaatimukseen vaikuttaa oleellisesti se, kuinka monta valinnaista vanhempaisjoukkoa G_i solmuilla V_i on. Jos vanhempien määrälle ei aseteta ylärajaa, vaihtoehtoisia joukkoja voi (yhdelle solmulle) olla enimmillään 2^{n-1}. Asettamalla vanhempien lukumäärälle yläraja k, mahdollisten vanhempaisjoukkojen määrän ylärajaksi saadaan $\binom{n-1}{0} + \binom{n-1}{1} + \ldots + \binom{n-1}{k}$.[3] Summauksen laskenta onnistuu tällöin polynomisessa ajassa $O(n^k)$.

Edellä johdettua suljettua muotoa voidaan hyödyntää vain joillekin verkkorakenteiden piirteille. Heti huomataan, että triviaalissa tapauksessa $f(G) \equiv 1$ ehto (4) on voimassa, ja siten $P(D \mid \prec)$ on tehokkaasti laskettavissa. Yksinkertaisille piirteille $f_{V_i \to V_j}$ eli kaarille $V_i \to V_j$ kyseinen ehto on myös voimassa (asetetaan $f_i(G_i) = 1$ joss $V_j \in G_i$). Edelleen yksittäistä vanhempaissolmua laajemmin voidaan tarkastella kokonaista joukkoa: $f_{V_i \to U}(G) = 1$ joss joukon U solmut ovat solmun V_i vanhempia. Näille piirteille $P(f, D \mid \prec)$ ja siten $P(f \mid D, \prec)$ voidaan laskea tehokkaasti kuten edellä esitettiin.

Jo mainittujen piirteiden lisäksi Friedman ja Koller [3] tarkastelevat piirrettä $f_{V_i \sim V_j}$, joka on rakenteella G, jos siinä solmu V_i on solmun V_j *Markovin peitteessä* (engl. Markov blanket); näin on jos G sisältää kaaren $V_i \to V_j$, kaaren $V_j \to V_i$ tai sellaisen solmun V_k, jolle sekä $V_i \to V_k$ että $V_j \to V_k$ ovat G:n kaaria. Friedman ja Koller näyttävät, että tämän piirteen posterioritodennäköisyys voidaan esittää edellä mainittujen yksinkertaisempien piirteiden posterioritodennäköisyyksien avulla (yksityiskohdat sivuutetaan tässä).

Monimutkaisempien piirteiden, kuten kahden solmun välisen suunnatun polun, posterioritodennäköisyyden laskemiseen Friedman ja Koller ehdottavat simulointimenetelmää, jossa verkkorakenteita generoidaan satunnaisesti niiden posteriorijakaumasta. Ehdollistettuna solmujärjestyksellä generoiminen on helppoa, koska vanhempaisjoukot voidaan kullekin solmulle generoida riippumattomasti. Mielivaltaisten piirteiden posterioritodennäköisyyttä voidaan arvioida generoidusta otoksesta laskettavana Monte-Carlo -keskiarvona.

5.2 Keskiarvoistus yli järjestysten: MCMC-simulaatio

Edellä esitettiin, kuinka verkkorakenteiden piirteiden posterioritodennäköisyyksiä voidaan laskea tehokkaasti, kun solmujärjestys on kiinnitetty. Käytännön sovelluksia ajatellen ennalta kiinnitetty järjestys on kuitenkin liian rajoittava [3]. Ratkaisuksi Friedman ja Koller esittävät MCMC-menetelmää. Toisin kuin Madigan ja Yorkin [7] suoraviivaisessa lähestymistavassa Markovin ketjun tila-avaruudeksi otetaan nyt solmujärjestysten joukko.

[3] Friedman ja Koller antavat virheellisesti ylärajan $\binom{n}{k}$.

Monte-Carlo -integrointia varten piirteen posterioritodennäköisyys esitetään odotusarvona yli järjestysten $\prec$ posteriorijakauman:

$$P(f \mid D) = \sum_{\prec} P(\prec \mid D) P(f \mid D, \prec).$$

Kun Markovin ketjua simuloimalla on tuotettu otos $\{\prec_1, \ldots, \prec_T\}$ järjestysten posteriorijakaumasta, piirteen posteritodennäköisyyttä voidaan arvioida ehdollisten posterioritodennäköisyyksien $P(f \mid D, \prec)$ otoskeskiarvona.

Jos MCMC-simulaatiossa käytetään Metropolisin ja Hastingsin algoritmia, kuten Friedman ja Koller ehdottavat, riittää että posterioritodennäköisyydet $P(\prec \mid D)$ osataan laskea vakiokerrointa vaille. Tämä onnistuukin, sillä $P(\prec \mid D) \propto P(\prec) P(D \mid \prec)$, missä $P(D \mid \prec)$ voidaan laskea aiemmin esitetyllä tavalla. Priorit $P(\prec)$ Friedman ja Koller asettivat tasaisiksi.

Simulaation sekoittumisen ja konvergoinnin kannalta sopivan ehdotusjakauman valitseminen on oleellista. Friedman ja Koller yhdistivät kaksi yksinkertaista ehdotusjakaumaa. *Vaihto-operaatiossa* (engl. "flip") valitaan kaksi lukua, j ja k, satunnaisesti joukosta $\{1, 2, \ldots, n\}$ ja uudeksi järjestykseksi ehdotetaan sitä, joka saadaan vaihtamalla edellisessä järjestyksessä $(i_1 i_2 \ldots i_n)$ lukujen i_j ja i_k paikat:

$$(i_1 \ldots i_j \ldots i_k \ldots, i_n) \mapsto (i_1 \ldots i_k \ldots i_j \ldots, i_n).$$

Kierto-operaatiossa (engl. "cutting the deck") valitaan satunnaisesti kohta, josta jonoa "kierretään":

$$(i_1 \ldots i_j i_{j+1} \ldots i_n) \mapsto (i_{j+1} \ldots i_n i_1 \ldots i_j).$$

Nämä operaatiot ovat vaikutukseltaan hyvin erilaisia. Vaihto-operaatio muuttaa järjestystä noudattavien verkkorakenteiden joukkoa vain vähän verrattuna kierto-operaatioon. Vaihto-operaatio ehdottaa pieniä askelia ja siten ehdotukset hyväksytään usein. Operaatiot voidaan yhdistää siten, että vaihto valitaan aina eräällä todennäköisyydellä p ja kierto todennäköisyydellä $1-p$. (Tässä p on säätelyparametri, jolle määrätään sopiva arvo tapauskohtaisesti.)

6 Kokeellisia tuloksia

Friedman ja Koller testasivat esittämiään menetelmiä monipuolisesti. Seuraavassa referoidaan lyhyesti koejärjestelyt, koetulokset ja tehdyt johtopäätökset.

Ensimmäisessä asetelmassa verrattiin eksaktia posteriorin laskemista summaamalla yli järjestysten eksaktiin posteriorin laskemiseen summaamalla yli verkkorakenteiden. Tarkoituksena oli selvittää oliko käytettyjen prioritodennäköisyyksien vaikutus posterioritodennäköisyyksiin merkittävä. Kahdesta datajoukosta, *Vote* ja *Flare* (UCI-repository) valittiin viisi muuttujaa ja posterioritodennäköisyydet eräille Markov-piirteille laskettiin. Friedmanin ja Kollerin mukaan priorien vaikutus oli pieni.

Toisessa asetelmassa verrattiin järjestysten yli integroivaa MCMC-menetelmää eksaktiin laskentaan. Havaintoaineistoksi valittiin yhdeksän muuttujaa aineistosta *Flare*. Posterioritodennäköisyydet laskettiin Markov-piirteille ja kaarille käyttäen erittäin pieniä otoskokoja 5, 20 ja 50. Erityisesti Markov-piirteillä arvioidut posterioritodennäköisyydet olivat erittäin tarkkoja jo otoskoolla 5. Tässä näkyy järjestys-MCMC:n teho: yksi järjestys pitää sisällään eksponentiaalisen määrän rakenteita!

Kolmannessa asetelmassa verrattiin järjestys-MCMC:tä ja Madiganin ja Yorkin [7] ehdottamaa suoraviivaista rakenne-MCMC:tä. Havaintoaineisto oli keinotekoinen, 37 muuttujaa. Kilpailevien MCMC-menetelmien konvergointia oikeaan tasapainojakaumaan verrattiin generoimalla havaintoaineistot kokoa 100, 500, 1000 ja käynnistämällä useita simulaatioita eri aloitusarvoilla. Pienellä aineistolla molemmat simulaatiot vaikuttivat konvergoituvan, keskikokoisella ja suurella aineistolla järjestys-MCMC vaikutti konvergoituvan selvästi luotettavammin. Piirteiden posterioritodennäköisyyksien laskemiseksi molemmista simulaatioista poimittiin 50 solmujärjestyksen, ja vastaavasti rakenteen, otos. Friedman ja Koller havaitsivat, että järjestys-MCMC:n antamat arviot eivät eri simulaatiossa juurikaan vaihtele kuten rakenne-MCMC:llä. Tästä he päättelevät, että järjestys-MCMC tuottaa otoksen todellakin posteriorijakaumasta. Tämä päätelmä vaikuttaa hieman hätäiseltä: 50 verkkorakennetta on ilman muuta liian vähän luotettavaan arviointiin. Miksi ei otettu suurempaa otosta, kun iteraatioiden määrä rakenne-MCMC -simulaatioissa oli jopa 600 000? (Samassa ajassa järjestys-MCMC simuloi noin 60 000 iteraatioita.)

Friedman ja Koller toteavat, että simulaatioajassa mitattuna järjestys-MCMC sekoittuu ja konvergoi nopeammin kuin rakenne-MCMC. He selittävän tämän sillä, että järjestysten avaruudessa posteriorimaasto on sileämpi kuin rakenteiden avaruudessa. Sekoittumiseen vaikuttaa myös havaintoaineiston koko: mitä enemmän on havaintoja, sitä terävämpiä ovat posteriorimaaston huiput ja harjanteet.

7 Yhteenveto

Friedmanin ja Kollerin tarkastelema Bayes-verkkojen rakenteen päättelyn ongelma on kiinnostava bayesiläisen päättelyn ja MCMC-menetelmien käytön näkökulmista. ”Puolibayesiläisten” optimointimenetelmien sijaan pyrkimys täydelliseen bayesiläiseen päättelyyn on (kirjoittajan mielestä) kannatettavaa. Keskiarvoistusperiaatteen noudattaminen näyttää tulevan mahdolliseksi myös käytännössä, kun tietokoneiden laskentakapasiteetti kasvaa, ja numeerisia menetelmiä, kuten MCMC-tekniikoita, opitaan hyödyntämään tehokkaasti.

Friedmanin ja Kollerin esittämä menetelmä edustaa lähestymistapaa, jossa MCMC-menetelmä otetaan käyttöön vasta, kun se on aivan välttämätöntä. Nojaten bayesiläiseen keskiarvoistamiseen naiivi, ”raaka” MCMC operoisi Bayes-verkkojen tila-avaruudessa. Koska tarkasteltavat piirteet riippuvat vain verkkorakenteesta ja koska tarvittavat todennäköisyysintegroinnit liittyen verkkorakenteiden parametreihin voidaan suorittaa tehokkaasti, naiivi MCMC seikkailisi tarpeettoman suuressa tila-avaruudessa. Madiganin ja Yorkin [7] ehdottama suoraviivainen, ”integroiva” MCMC operoikin verkkorakenteiden avaruudessa. Kuten Friedman ja Koller koetulosten tuella osoittavat, rakenteiden tila-avaruuskin on liian ”hienojakoinen”.

Heidän oma versionsa on äärimmäinen, ”yli-integroiva” MCMC, jossa Markovin ketjun tilat vastaavat sopivia verkkorakenteiden osajoukkoja, kooltaan eksponentiaalisia (suhteessa solmujen määrään). Friedman ja Koller osoittavat, että kiinnitetyn solmujärjestysten kanssa konsistentit rakenteet muodostavat sopivan osajoukon, jossa erilaisten piirteiden posterioritodennäköisyydet ovat tehokkaasti laskettavissa.

Järjestyksellä ehdollistaminen ja suljetun muodon (kaava (5)) hyväksikäyttö eivät sinänsä ole uusia ajatuksia, kuten Friedman ja Koller itsekin myöntävät, vaan niitä on esitetty Bayes-verkkojen yhteydessä jo aiemmin [2]. Kyseisen lausekkeen johdossa käytetty ”summa-tulo -temppu” on lopulta vain osittelulain $ab + ac = a(b + c)$ sovellus. Lain yleistäminen tarjoaa nopeita laskentamenetelmiä moniin vastaaviin laskentaongelmiin [1]. Friedman ja Kollerin tapa soveltaa tehokasta suljettua muotoa bayesiläisessä keskiarvoistamisessa ja MCMC-menetelmän yhteydessä lienee kuitenkin uusi.

Tässä kirjoitelmassa on voitu käsitellä vain joitakin Friedmanin ja Kollerin artikkelissa [3] esiin nousseita kysymyksiä. Erityisesti on jouduttu sivuuttamaan priorien valintaan ja baeysiläiseen lähestymistapaan itsessään liittyviä ongelmia. Tärkeä kysymys on esimerkiki, missä tapauksissa ja minkälaiset piirteet verkkorakenteiden posteriorijakauma todella kykenee identifioimaan.

Viitteet

[1] Aji, S. M., McEliece, R. J., The generalized distributive law. *IEEE Transactions of Information Theory*, 46 (2000): 325–343.

[2] Buntine, W., Theory refinement of Bayesian networks. *Proc. of 7'th Annual Conf. on Uncertainty in Artificial Intelligence (UAI-91)*, Los Angeles, California, USA, July, 1991, 52–60.

[3] Friedman, N., Koller, D., Being Bayesian about network structure. *Proc. of 16'th Annual Conf. on Uncertainty in Artificial Intelligence (UAI-00)*, Stanford, California, USA, June, 2000, 201–210.

[4] Friedman, N., Linial, M., Nachman, I., Pe'er, D., Bayesian networks to analyze expression data. *RECOMB* (2000), 206–215.

[5] Gilks, W. R., Richardson, S., Spiegelhalter, D. J., *Markov Chain Monte Carlo in Practice.* Chapman & Hall, Lontoo, Iso-Britannia, 1996.

[6] Hastings, W. K., Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57 (1970), 97–109.

[7] Madigan, D., York, J., Bayesian graphical models for discrete data. *Inter. Stat, Rev.* 63 (1995), 215–232.

[8] Roberts, G. O., Markov chain concepts related to sampling algorithms. Teoksessa [5], 45–57, 1995.