

1. Käytetään yksinkertaisesti Bayesin kaavaa:

$$P(A | B) = \frac{P(A)P(B | A)}{P(B)}.$$

Tapauksessa  $B = 1$  saadaan

$$P(A = 0 | B = 1) = \frac{P(A = 0)P(B = 1 | A = 0)}{P(B = 1)} = \frac{0,3 \cdot 0,6}{P(B = 1)} \quad \text{ja}$$
$$P(A = 1 | B = 1) = \frac{P(A = 1)P(B = 1 | A = 1)}{P(B = 1)} = \frac{0,7 \cdot 0,2}{P(B = 1)}.$$

Koska näiden todennäköisyyksien summa on välttämättä 1, voidaan ratkaista jaka-  
ja  $P(B = 1)$ :

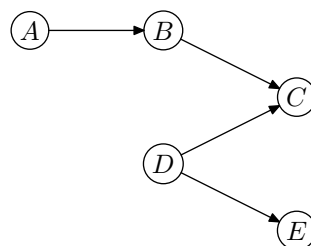
$$(0,18 + 0,14)/P(B = 1) = 1 \implies P(B = 1) = 0,32.$$

Siis  $P(A = 0 | B = 1) = 0,18/0,32 \approx 0,56$  ja  $P(A = 1 | B = 1) = 0,14/0,32 \approx 0,44$ .

Tapauksessa  $B = 0$  saadaan vastaavalla tavalla laskien  $P(A = 0 | B = 0) \approx 0,18$  ja  
 $P(A = 1 | B = 0) \approx 0,82$ .

2. Jos satunnaismuuttujan  $X$  arvo tunnetaan, muuttujat  $A, B, C$  ja  $D$  ovat toisistaan  
pareittain riippumattomia. Jos  $X$  on tuntematon, verkkorakenteesta ei seuraa mitään  
riippumattomuuksia. (Bayes-verkon paikalliset jakaumat voivat toki olla sellaiset,  
että muitakin riippumattomuuksia on voimassa.)

- 3.



Ensinnäkin  $P(A | E = 0) = P(A)$ . Intuitiivisesti sanottuna tieto muuttujan  $E$   
arvosta ei pääse kulkemaan verkon kohdan  $B \rightarrow C \leftarrow D$  läpi, kun  $C$ :n arvo on  
tuntematon. Tämän perustelun voi esittää täsmällisesti d-separaation avulla, mutta

tulos seuraa melko helposti myös yhteisjakauman hajotelmasta:

$$\begin{aligned}
 P(A, E = 0) &= \sum_{B,C,D} P(A)P(D)P(B | A)P(C | B, D)P(E = 0 | D) \\
 &= P(A) \sum_D P(D)P(E = 0 | D) \underbrace{\sum_B P(B | A) \sum_C P(C | B, D)}_{=1} \\
 &= P(A) \sum_D P(D)P(E = 0 | D) \underbrace{\sum_B P(B | A)}_{=1} \\
 &= P(A) \sum_D P(D, E = 0) = P(A)P(E = 0),
 \end{aligned}$$

joten  $P(A | E = 0) = P(A, E = 0)/P(E = 0) = P(A)$ .

Jakauma  $P(C | E = 0)$  sen sijaan ei ole sama kuin priorijakauma, koska rakenteen  $C \leftarrow D \rightarrow E$  läpi tieto voi kulkea (kun  $D$ :n arvo on tuntematon). Lasketaan ensin Bayesin kaavalla jakauma  $P(D | E = 0)$ :

$$\begin{aligned}
 P(D = 0 | E = 0) &\propto P(D = 0)P(E = 0 | D = 0) = 0,3 \cdot 0,1 = 0,03, \\
 P(D = 1 | E = 0) &\propto P(D = 1)P(E = 0 | D = 1) = 0,7 \cdot 0,4 = 0,28, \\
 P(D = 0 | E = 0) + P(D = 1 | E = 0) &= 1 \\
 \implies P(D = 0 | E = 0) &= 3/31 \approx 0,097, \quad P(D = 1 | E = 0) = 28/31 \approx 0,903.
 \end{aligned}$$

Jotta pystytään laskemaan  $C$ :n jakauma, tarvitaan myös tieto  $B$ :stä. Tämä on helppo laskea:

$$\begin{aligned}
 P(B | E = 0) &= \sum_A P(A | E = 0)P(B | A, E = 0) = \sum_A P(A)P(B | A) \\
 &= P(A = 0)P(B | A = 0) + P(A = 1)P(B | A = 1),
 \end{aligned}$$

siis  $P(B = 0 | E = 0) = 0,1 \cdot 0,1 + 0,9 \cdot 0,2 = 0,19$  ja  $P(B = 1 | E = 0) = 0,1 \cdot 0,9 + 0,9 \cdot 0,8 = 0,81$ . Nämä tulokset yhdistämällä saadaan jakauma  $P(B, D | E = 0)$ :

$b$	$d$	$P(B = b, D = d   E = 0)$
0	0	$0,19 \cdot 3/31 \approx 0,0184$
0	1	$0,19 \cdot 28/31 \approx 0,1726$
1	0	$0,81 \cdot 3/31 \approx 0,0784$
1	1	$0,81 \cdot 28/31 \approx 0,7316$

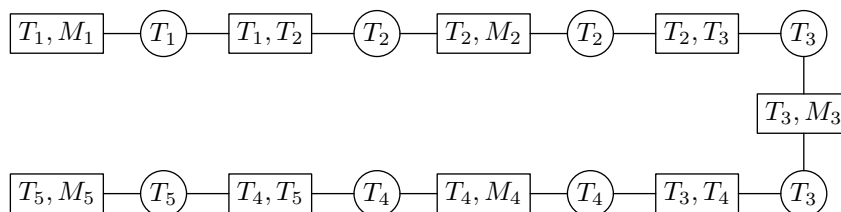
Koska Bayes-verkon Markov-ominaisuuksien perusteella  $P(C | B, D, E) = P(C | B, D)$ , joten nyt voidaan laskea  $P(C | E = 0) = \sum_{B,D} P(C | B, D, E = 0) =$

$$\sum_{B,D} P(C | B, D)P(B | E = 0)P(D | E = 0):$$

$$P(C = 0 | E = 0) = 0,19(3/31 \cdot 0,1 + 28/31 \cdot 0,2) + 0,81(3/31 \cdot 0,2 + 28/31 \cdot 0,5) \approx 0,42$$

$$P(C = 1 | E = 0) = 0,19(3/31 \cdot 0,9 + 28/31 \cdot 0,8) + 0,81(3/31 \cdot 0,8 + 28/31 \cdot 0,5) \approx 0,58$$

4. Koska jokaisella solmulla on enintään yksi vanhempi, moralisointi tapahtuu yksinkertaisesti jättämällä kaarien suunnat pois. Moralisoitussa verkossa ei ole yhtään silmukkaa, joten kolmiointiakaan ei tarvitse tehdä. Saatu verkko on sinällään puu, joten periaatteessa liittymäpuuesitystä ei tarvittaisi, mutta tehdään se nyt opiskelutarkoituksessa. Saadun verkon jokaisessa klikissä on kaksi solmua, ja klikkejä on yhdeksän: neljä muotoa  $\{T_i, T_{i+1}\}$  ( $i = 1, \dots, 4$ ) ja viisi muotoa  $\{T_i, M_i\}$  ( $i = 1, \dots, 5$ ). Kun klikit yhdistetään puuksi, on huomattava, että klikit  $\{T_i, T_{i+1}\}$ ,  $\{T_{i+1}, M_{i+1}\}$  ja  $\{T_{i+1}, T_{i+2}\}$  sisältävät yhteisen alkion  $T_{i+1}$  ( $i = 1, \dots, 3$ ). Seuraavassa kuvassa suorakulmiot merkitsevät klikkejä ja ympyrät niiden leikkauksia. (Tämä ei ole ainoa mahdollinen puu.)



Kysymys potentiaalifunktioista oli hieman huonosti muotoiltu. Alustuksen ja propagoinnin jälkeen potentiaalifunktiot ovat joka tapauksessa marginaalijakaumia. Kun mitään evidenssiä ei ole nähty, symmetriasyyistä  $P(T_i = 1) = P(M_i = 1) = 0,5$  kaikilla  $i$ , mistä on helppo laskea potentiaalit:

$t$	$t'$	$P(T_i = t, T_{i+1} = t')$	$t$	$m$	$P(T_i = t, M_i = m)$
0	0	0,4	0	0	0,45
0	1	0,1	0	1	0,05
1	0	0,1	1	0	0,05
1	1	0,4	1	1	0,45

5. Algoritmista on vihjeenä annettu mainio kuvaus. Liitteinä ovat algoritmin tässä erikoistapauksessa toteuttavat Matlab- ja Python-skriptit, joista edellinen käyttää Bayes Network Toolboxia. Python-skriptistä saa kommenttimerkkejä poistamalla myös välitulosteita.

Lopputulokset on seuraava.

$$\begin{aligned} P(T_1 = 0) &\approx 0.9667, & P(T_1 = 1) &\approx 0.0333 \\ P(T_2 = 0) &\approx 0.9778, & P(T_2 = 1) &\approx 0.0222 \\ P(T_3 = 0) &\approx 0.9000, & P(T_3 = 1) &\approx 0.1000 \\ P(T_4 = 0) &\approx 0.1225, & P(T_4 = 1) &\approx 0.8775 \\ P(T_5 = 0) &\approx 0.0614, & P(T_5 = 1) &\approx 0.9386 \end{aligned}$$

(Vertaa myös kätkeytyneiden Markov-mallien "forward-backward"-algoritmiin.)

6. Koska

$$\frac{d^2}{dx^2} f(x) = \frac{d}{dx} \frac{1}{2} x^{-1/2} = -\frac{1}{4} x^{-3/2} < 0$$

kaikilla  $x > 0$ , funktio  $f$  on konkaavi.

Konjugaattifunktio on  $f^*(\lambda) = \min_x \{ \lambda x - \sqrt{x} \}$ .

Minimin löytämiseksi lasketaan derivaatta

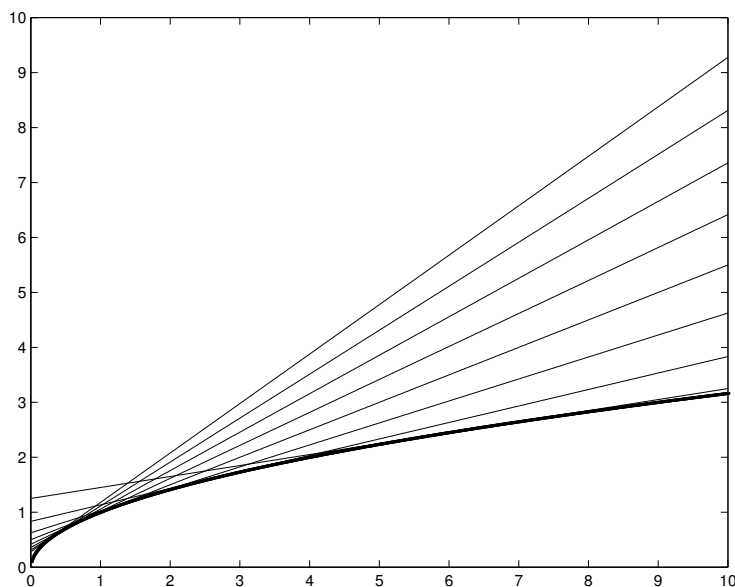
$$\frac{d}{dx} (\lambda x - \sqrt{x}) = \lambda - \frac{1}{2} x^{-1/2}.$$

Derivaatta on nolla vain, kun  $x = (2\lambda)^{-2}$ . Pienemmillä  $x$  derivaatta on negatiivinen ja suuremmilla positiivinen, joten kyseessä on todella minimi. Siis

$$f^*(\lambda) = \lambda \frac{1}{4\lambda^2} - \frac{1}{2\lambda} = -\frac{1}{4\lambda},$$

joten on saatu esitys

$$f(x) = \min_{\lambda} \left\{ \lambda x + \frac{1}{4\lambda} \right\}.$$



7. Tuotetaan satunnaislukuja jakaumasta  $P(A, B | C = c)$ . Bayesin kaava antaa

$$P(A, B | C = c) = \frac{P(C = c | A, B)P(A, B)}{P(C = c)},$$

ja verkkorakenteesta saadaan  $P(C = c | A, B) = P(C = c | B)$  ja  $P(A, B) = P(A)P(B | A)$ . Lukua  $P(C = c) = \int_{A,B} P(A, B, C = c)$  ei ehkä osata laskea, mutta ei se mitään: funktion

$$P^*(A, B) = P(A)P(B | A)P(C = c | B)$$

arvoja osataan laskea millä tahansa  $A$ :n ja  $B$ :n arvoilla.

MCMC-menetelmillä saadaan siis otos jakaumasta  $P(A, B | C = c)$ . Haluttu jakauma oli  $P(B | C = c) = \int_A P(A, B | C = c)$ , ja otoksen marginalisointihan on helppoa: otetaan otoksista  $B$ :n arvot ja unohdetaan  $A$ :n arvot.

8. (a) Tapauksessa 2 on hyvin suuri osa ehdotetuista arvoista hylätään, joten ketjua pitäisi simuloida pitkään luotettavan otoksen saamiseksi. Toisin sanottuna autokorrelaatio on kovin suuri. Jakaumalla on selvästi (ainakin) kaksi moodia, mutta niiden suhteellisesta osuudesta on vaikea sanoa mitään – Markov-ketju on hyppinyt niiden välillä pari kertaa, mutta näin lyhyessä simulaatiossa on sattuman varassa, kuinka luotettavan kuvan simulaatio antaa moodien suhteellisista osuuksista.

Tapauksessa 3 on myös suuri autokorrelaatio, tässä tapauksessa satunnaiskävelyn eikä hylkäysten takia. Alkupisteen valinta näkyy ketjussa hyvin pitkälle.

Ensimmäisenkään simulaation konvergenssista ei voi sanoa mitään varmaa. Tämä on MCMC-menetelmien yleinen ongelma.

(b) Kolmannessa simulaatiossa on selvästi pienin ehdotusjakauman leveys – ketjun ottamat askeleet ovat hyvin pieniä. Leveys on suurin ehkä hieman yllättävästi toisessa simulaatiossa. Otetut askeleet ovat jokseenkin yhtä pitkiä kuin ensimmäisessä, mutta koska simulaatioiden ainoa ero on ehdotusjakauman leveydessä, suuren hylättyjen askelten osuuden on jollain tapaa liityttävä tähän. Ehdotus hyväksymistodennäköisyys riippuu sen todennäköisyydestä kuulua simuloitavaan jakaumaan, ja liian leveä ehdotusjakauma tuottaa paljon epätodennäköisiä ja siis hylättäviä ehdotuksia.

Näiden tarkastelujen valossa ensimmäisen ajon tulos näyttää parhaalta. [Todellinen jakauma, johon simulaatio perustui, koostui kahdesta normaalijakaumasta:  $N(2, 0; 1, 0)$  priorilla 0,8 ja  $N(6, 0; 0, 4)$  priorilla 0,2.]