Muuttujien riippumattomuus

- Jos esimerkkiin lisätään muuttuja Säätila, jolla on 4 mahdollista arvoa, on edellä ollut yhteisjakauman taulukko monistettava neljästi
- Koska hammasongelmat eivät vaikuta säätilaan, pätee:

  P(Säätila = pilveä | särky, osuma, reikä) = P(Säätila = pilveä)

- Tulosäännön ja yo. havainnon perusteella

  P(särky, osuma, reikä, Säätila = pilveä) =
  P(Säätila = pilveä) P(särky, osuma, reikä)

- Sama pätee myös muille muuttujan Säätila arvoille, joten

  $\mathbf{P}$(Särky, Osuma, Reikä, Säätila) =
  $\mathbf{P}$(Säätila) $\mathbf{P}$(Särky, Osuma, Reikä)

- Tarkasteltavat yhteisjakauma taulukot ovat siis 8 ja 4 alkioiset
- Propositiot a ja b ovat *riippumattomia* jos

  $P(a \mid b) = P(a) \Leftrightarrow P(b \mid a) = P(b) \Leftrightarrow P(a \wedge b) = P(a)\,P(b)$

- Vastaavasti muuttujat X ja Y ovat toisistaan riippumattomia jos

  $\mathbf{P}(X \mid Y) = \mathbf{P}(X) \Leftrightarrow \mathbf{P}(Y \mid X) = \mathbf{P}(Y) \Leftrightarrow \mathbf{P}(X, Y) = \mathbf{P}(X)\mathbf{P}(Y)$

- Rippumattomat kolikonheitot:

  $\mathbf{P}(C_1, \ldots, C_n)$:n sijaan tarkasteltava n:n $P(C_i)$:n tuloa

Bayesin sääntö

- Tulosäännön perusteella $P(a \wedge b) = P(a \mid b)\, P(b)$ ja konjunktion kommutatiivisuuden perusteella $P(a \wedge b) = P(b \mid a)\, P(a)$
- Oikeat puolet ekvivalisoimalla ja jakamalla $P(a)$:lla saadaan Bayesin sääntö

$$P(b \mid a) = P(a \mid b)\, P(b) \,/\, P(a)$$

- Yleisemmässä muodossa, muuttujille X ja Y ja taustatietämykselle $\mathbf{e}$

$$\underline{P}(Y \mid X, \mathbf{e}) = \underline{P}(X \mid Y, \mathbf{e})\, \underline{P}(Y \mid \mathbf{e}) \,/\, \underline{P}(X \mid \mathbf{e})$$

- Normalisointia käyttäen Bayesin sääntö voidaan kirjoittaa muotoon

$$\underline{P}(Y \mid X) = \alpha\, \underline{P}(X \mid Y)\, \underline{P}(Y)$$

- Puolella aivokalvontulehduspotilaista on jäykkä niska
  $P(j \mid a) = 0.5$
- Aivokalvontulehdusta esiintyy 1 / 50 000 tapauksessa
  $P(a) = 1/50\,000$
- Jäykkää niskaa valittaa joka 20:s potilas
  $P(j) = 1/20$
- Mikä on todennäköisyys, että jäykkää niskaansa valittavalla potilaalla on aivokalvontulehdus?
  $P(a \mid j) = P(j \mid a)\, P(a) \,/\, P(j)$
  $= 20 \,/\, (2 \cdot 50\,000) = 0.0002$

- Lääkärillä toki voisi olla suoraan tieto, että jäykkä niska tietää aivonkalvontulehdusta yhdessä tapauksessa 5 000:sta
- Lääkärillä siis olisi kvantitatiivista tietoa *diagnostiseen* suuntaan – oireista aiheuttajiin, eikä tarvetta Bayesin säännön käyttöön
- Diagnostinen tietämys valitettavasti vain on usein arempaa kuin kausaalinen tietämys
- Aivonkalvontulehdusepidemian tapauksessa ehdollistamaton todennäköisyys $P(a)$ nousee
- Sen sijaan ehdollisen tn:n $P(j \mid a)$ arvo pysyy ennallaan
- Diagnostisen todennäköisyyden $P(a \mid j)$ suoraan epidemiaa edeltäneisiin havaintoihin perustanut lääkäri ei osaa päivittää arvoa mitenkään
- Sen sijaan kausaaliseen päättelyyn nojaava lääkäri tietää, että $P(a \mid j)$:n tulee nousta $P(a)$:n suhteessa

- Bayesin sääntö on kaikkien modernien probabilististen päättelyjärjestelmien pohjana
- Sinänsä yksinkertainen sääntö ei pintapuolisesti katsoen näyttäisi antavan paljon päättelymahdollisuuksia, mutta kuten edellinen esimerkki osoittaa kyse on mahdollisuudesta soveltaa olemassa olevaa tietoa
- Nimittäjässä olevan ehdon todennäköisyyksien laskemisesta voidaan päästä laskemalla kyselymuuttujan posteriori-todennäköisyys kaikilla sen mahdollisilla arvoilla
- Esim. $\underline{\mathbf{P}}(A \mid j) = \alpha\, [\, P(j \mid a)\, P(a),\, P(j \mid \neg a)\, P(\neg a)\,]$
- Joskus helpompaa, joskus ei

- Kun probabilistinen kysely on ehdollistettu useammalle kuin yhdelle havainnolle

  $\underline{P}(Reikä \mid särky \wedge osuma)$

  täyteen yhteisjakaumaan perustuva menetelmä ei käy
- Bayesin säännön soveltaminenkaan ei yleisesti ottaen lievennä eksponentiaalista vaativuutta
- Tarvittaisiin muuttujien riippumattomuutta, mutta havaintoja särky ja osuma vastaavat muuttujat ovat selvästi yhteydessä toisiinsa
- Molempien muuttujien suora vaikuttaja on reikä hampaassa, mutta ilman muuttujaa Reikä ne ovat *ehdollisesti riippumattomia* (conditionally independent) — Särky ja Osuma eivät ole suoraan toisistaan riippuvia

- Ehdollinen riippumattomuus:

  $\underline{P}(särky \wedge osuma \mid Reikä) = \underline{P}(särky \mid Reikä)\ \underline{P}(osuma \mid Reikä)$
- Yhdessä Bayes säännön kanssa saamme lopulta

  $\underline{P}(Reikä \mid särky \wedge osuma) =$

  $\alpha\ \underline{P}(Reikä)\ \underline{P}(särky \mid Reikä)\ \underline{P}(osuma \mid Reikä)$
- Nyt siis tarvitsee enää käsitellä kolmea erillistä jakaumaa
- Muuttujien X ja Y ehdollinen riippumattomuus annettuna Z on tarkkaan ottaen

  $\underline{P}(X, Y \mid Z) = \underline{P}(X \mid Z)\ \underline{P}(Y \mid Z)$
- Ekvivalentisti $\underline{P}(X \mid Y, Z) = \underline{P}(X \mid Z)$ ja $\underline{P}(Y \mid X, Z) = \underline{P}(Y \mid Z)$

- Jos kaikki havaintomuuttujat ovat ehdollisesti riippumattomia annettuna kyselymuuttuja, niin eksponentiaalinen tietämyksenesityksen kasvuvauhti putoaa lineaariseksi
- Tämä on naiivin Bayes -mallin ajatus, kaikki seuraukset $S_1, \ldots, S_n$ ovat ehdollisesti riippumattomia annettuna yksi aiheuttaja A
- Tällöin niiden täysi yhteisjakauma

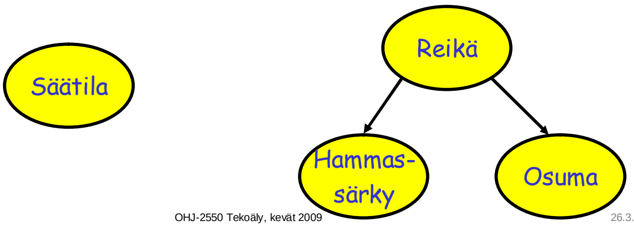
$$\underline{P}(A, S_1, \ldots, S_n) = \underline{P}(A) \prod_i \underline{P}(S_i \mid A)$$

- Mallia käytetään usein yksinkertaistamiseen, vaikka seuraukset eivät olisi ehdollisesti riippumattomia
- Saattaa toimia yllättävän hyvin kaikesta huolimatta

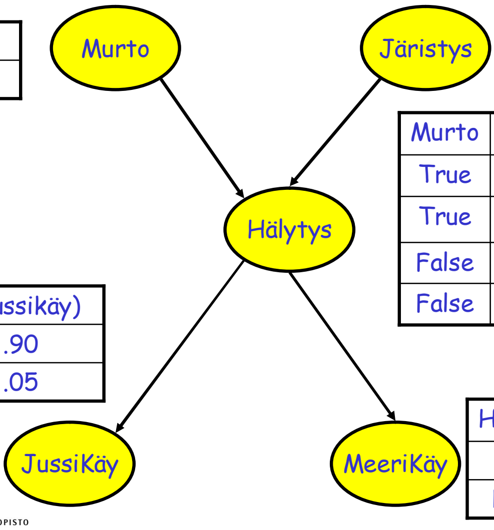
4.1 PROBABILISTINEN PÄÄTTELY

- Bayes-verkko on suunnattu verkko, jonka kuhunkin solmuun liittyy todennäköisyys
  1. Satunnaismuuttujien joukko muodostaa verkon solmut. Muuttujat voivat olla diskreetti- tai jatkuva-arvoisia
  2. Solmujen välillä on suunnattuja linkkejä (nuolia). Jos solmusta X on nuoli solmuun Y, niin X on Y:n *vanhempi*
  3. Solmulla X_i on ehdollinen todennäköisyysjakauma $\underline{P}(X_i \mid Vanhemmat(X_i))$
  4. Verkossa ei ole suunnattuja syklejä. Se on siis suunnattu syklitön verkko DAG

- Nuolen $X \rightarrow Y$ intuitiivinen merkitys on, että X vaikuttaa suoraan Y:hyn
- Verkon topologia — solmut ja nuolet — määräävät muuttujien ehdolliset riippumattomuudet
- Kun topologia on kiinnitetty, pitää vielä määrätä kullekin muuttujalle ehdollinen todennäköisyysjakauma annettuna sen vanhemmat
- Verkon topologia ja ehdolliset todennäköisyysjakaumat määräävät implisiittisesti täyden yhteisjakauman muuttujille



- Taloon on asennettu varashälytin, joka on luotettava murtoyritysten tunnistaja, mutta reagoi myös lieviin maanjäristyksiin
- Naapurit Jussi ja Meeri ovat luvanneet tarkistaa tilanteen hälytyksen kuullessaan
- Jussi kuulee jokaisen hälytyksen, mutta sen lisäksi joskus erehtyy tarkistamaan tilanteen myös puhelimen soidessa
- Meeri puolestaan kuuntelee musiikkia lujalla, eikä aina kuule hälytystä
- Tavoitteenamme on arvioida murron todennäköisyys annettuna tieto siitä kuka on käynyt tai ei ole käynyt tarkistamassa tilannetta



P(murto)
.001

P(järistys)
.002

Murto	Järistys	P(hälytys)
True	True	.95
True	False	.94
False	True	.29
False	False	.001

Hälytys	P(jussikäy)
True	.90
False	.05

Hälytys	P(meerikäy)
True	.70
False	.01

- Verkon topologiaan sisältyvät seuraavat ajatukset:
  - Murto ja maanjäristys vaikuttavat suoraan hälyttimen laukeamiseen,
  - Jussin ja Meerin vierailut riippuvat suoraan vain hälyttimestä
  - he eivät huomaa murtoa tai maanjäristystä suoraan tai keskustele keskenään
- Musiikinkuuntelu ja puhelimen soiminen ovat verkosta luettavissa vain implisiittisesti hälytyksen aiheuttaman tarkistuskäynnin epävarmuutena
- Todennäköisyydet tiivistävät äärettömän mahdollisten vaikuttajien joukon
- Esim. hälytin voi jäädä laukeamatta koska ilmankosteus on korkea, on sähkökatko, varapatteri on tyhjentynyt, johdot on katkaistu, kuollut hiiri estää kelloa soimasta, jne.

- Ehdollisten todennäköisyyksien taulukot luettelevat muuttujan arvojen todennäköisyydet riippuen solmun vanhempien arvoasetuksesta
- Kunkin rivin todennäköisyyksien on summauduttava arvoon 1
- Edellisessä esimerkissä kaikki muuttujat ovat totuusarvoisia, joten riittää todennäköisyyden p tapauksessa, että muuttujan arvo on tosi, koska arvon epätosi tn. on $1-p$
- Yleisesti ottaen, jos solmulla on k vanhempaa, niin erikseen määrättäviä todennäköisyyksiä on 2^k kappaletta
- Jos solmulla ei ole vanhempia on taulukossa vain yksi rivi

Bayes-verkon semantiikka

- Jokaisen yhteistodennäköisyysjakauman solun arvo voidaan laskea Bayes-verkon sisältämästä informaatiosta
- Solun arvo on muuttujien tietty arvoasetus
  $P(X_1 = x_1 \wedge \ldots \wedge X_n = x_n)$, jota merkitään $P(x_1, \ldots, x_n)$
- Solun arvo on $P(x_1, \ldots, x_n) = \prod_{i=1,\ldots,n} P(x_i \mid vanhemmat(X_i))$, missä $vanhemmat(X_i)$ viittaa muuttujien $Vanhemmat(X_i)$ määrättyihin arvoihin
- $P(jussiKäy \wedge meeriKäy \wedge hälytys \wedge \neg murto \wedge \neg järistys)$
  $= P(jussiKäy \mid hälytys)\ P(meeriKäy \mid hälytys)$
  $P(hälytys \mid \neg murto \wedge \neg järistys)\ P(\neg murto)\ P(\neg järistys)$
  $= .90 \times .70 \times .001 \times .999 \times .998$
  $= .00062$

- Yhteisjakauma $P(x_1, \ldots, x_n)$ voidaan kirjoittaa tulosäännöllä muotoon $P(x_n \mid x_{n-1}, \ldots, x_1)\, P(x_{n-1}, \ldots, x_1)$
- Samaa purkamista voidaan toistaa, jolloin lopulta saadaan kaava

$$P(x_n \mid x_{n-1}, \ldots, x_1)\, P(x_{n-1} \mid x_{n-2}, \ldots, x_1) \ldots P(x_2 \mid x_1)\, P(x_1) = \prod_{i=1,\ldots,n} P(x_i \mid x_{i-1}, \ldots, x_1)$$

- Tämä *ketjusääntö* (chain rule) pätee mille tahansa satunnaismuuttujien joukolle
- Jakauma on siis yhtäpitävä väittämän $\mathbf{P}(X_i \mid X_{i-1}, \ldots, X_1) = \mathbf{P}(X_i \mid Vanhemmat(X_i))$ kanssa kunhan $Vanhemmat(X_i) \subseteq \{ X_{i-1}, \ldots, X_1 \}$

- Kun solmut indeksoidaan järjestyksessä, joka on konsistentti verkkostruktuurin implisiittisen osittaisjärjestyksen kanssa, niin ehto pätee
- Solmuilta siis vaaditaan ehdollista riippumattomuutta edeltäjäsolmuista annettuna vanhemmat
- Solmun vanhemmiksi tulee valita ne solmut, jotka vaikuttavat muuttujan arvoon suoraan
- Jos esimerkissä ajatellaan Meerin tarkastuskäyntisolmua, niin murto ja järistys vaikuttavat sen arvoon, mutta vain hälytyksen kautta
- Uskomme mukaan

$\mathbf{P}(MeeriKäy \mid JussiKäy, Hälytys, Järistys, Murto) =$
$\mathbf{P}(MeeriKäy \mid Hälytys)$

- Bayes-verkon etu täyteen yhteisjakaumaan verrattuna on kompaktius, joka seuraa sen lokaalisuudesta
- Alikomponentit ovat tekemisissä vain rajoitetun määrän muita komponentteja kanssa riippumatta komponenttien kokonaismäärästä
- Tästä seuraa kompleksisuuden lineaarinen, ei eksponentiaalinen, kasvu
- Bayes-verkko sovellukselle, jossa kuhunkin n:stä muuttujasta vaikuttaa suoraan korkeintaan k muuttujaa, vaatii ehdollisten todennäköisyystaulukoiden esittämiseen 2^k lukua
- Kaikkiaan siis $n2^k$ lukua, kun yhteisjakaumassa on 2^n solua
- Esim. $n = 30$ ja $k = 5$: $n2^k = 960$ ja $2^n > 10^9$

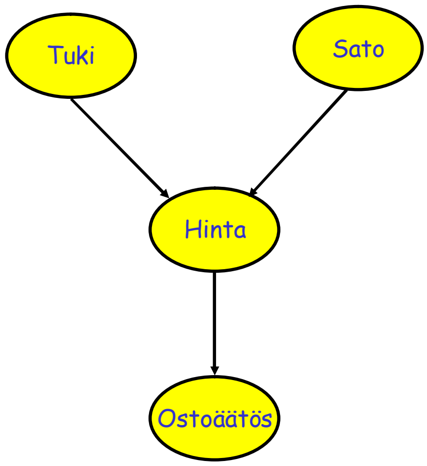
- Bayes-verkossa voidaan tarkkuutta vaihtaa kompleksisuuteen
- Heikkoja riippuvuuksia voi kannattaa jättää pois verkosta kompleksisuuden pitämiseksi kurissa, mutta tällöin verkon tarkkuus kärsii

- Verkkotopologian oikea määrääminen on vaikea ongelma
- Solmut pitäisi lisätä järjestyksessä: "juuri aiheuttajat", muuttujat, joihin ne vaikuttavat suoraan, jne., kunnes lopulta lehdiksi tulevat muuttujat jotka eivät vaikuta suoraan mihinkään muihin muuttujiin
- Väärässä järjestyksessä lisätyt solmut tekevät verkosta kompleksisemman ja epäintuitiivisen

Jatkuva-arvoiset muuttujat

- Näiden käsittely voidaan välttää *diskretoinnilla*, jossa jatkuva arvoalue jaetaan joihinkin intervalleihin
- Liian harvasta jaosta seuraa suuri tarkkuuden menetys ja liian tiheästä puolestaan suuri määrä käsiteltäviä arvoja
- Diskretointi voi kuitenkin olla jopa todistettavasti oikea lähestymistapa
- Toinen mahdollisuus on arvioida jatkuvaa arvoaluetta jollain standardilla todennäköisyyden tiheysfunktioperheellä, joilla on äärellinen määrä parametreja
- Esim. Gaussin l. normaalijakauman $N(\mu, \sigma^2)(x)$ parametrit ovat keskiarvo μ ja varianssi σ^2

- Hybridiverkossa on sekä diskreettejä että jatkuva-arvoisia muuttujia
- Tällöin on kyettävä laskemaan jatkuva-arvoisen solmun ehdollinen todennäköisyys ja diskreetin solmun todennäköisyys annettuna jatkuva-arvoiset vanhemmat
- Esim. hedelmän (jatkuva-arvoinen) Hinta riippuu (jatkuva-arvoisesta) Sadosta ja (binääriarvoisesta) Tuesta
- Asiakkaan (diskreetti) Ostopäätös puolestaan riippuu yksinomaan hinnasta
- Hinnan todennäköisyysjakauman $\mathbf{P}(Hinta \mid Sato, Tuki)$ määräämiseksi lasketaan erikseen $P(Hinta \mid Sato, tuki)$ ja $P(Hinta \mid Sato, \neg tuki)$
- Lisäksi tarkastellaan hinnan h jakauman riippuvuutta muuttujan Sato arvosta s



- Hinnan jakauman parametrit määritellään siis s:n funktiona
- Usein käytetty jakauma on lineaarinen Gaussin jakauma, jonka keskiarvo μ riippuu lineaarisesti vanhemman arvosta ja jolla on kiinnitetty hajonta σ

$$
\begin{aligned}
P(h \mid s, tuki) &= N(a_t s + b_t, \sigma_t^2)(h) \\
&= \frac{1}{\sigma_t \sqrt{2\pi}} \exp\left(-\frac{1}{2} ((h - (a_t s + b_t)) / \sigma_t)^2 \right) \\
P(h \mid s, \neg tuki) &= N(a_f s + b_f, \sigma_f^2)(h)
\end{aligned}
$$

- Näin muodostuvat yksikyttyräiset jakaumat yhdistämällä saadaan lopulta kaksikyttyräinen jakauma P(h | s)
- Lineaarista normaalijakaumaa noudattavan Bayes-verkon yhteisjakauma "käyttäytyy hyvin"

- Diskreetti Ostopäätös riippuu tuotteen hinnasta
- Hinnan ollessa korkea ei ostoja tehdä ja hinnan ollessa matala asiakas ostaa hedelmän
- Hinnan ollessa näiden väliltä asiakkaan ostopäätös muuttuu pehmeästi/sileästi
- Ehdollinen todennäköisyys on siis pehmeä kynnysarvo
- Yksi mahdollisuus on käyttää ns. probit-jakaumaa, standardin normaalijakauman integraalia

$$\Phi(x) = \int_{-\infty, \dots, x} N(0, 1)(x)\, dx$$

- Tällöin ostopäätöksen ehdollinen todennäköisyys voisi olisi

$$P(\text{ostaa} \mid \text{hinta} = h) = \Phi((-h+ \mu) / \sigma)$$

- Hintakynnys siis on noin μ, kynnysalueen leveys on σ:aan suhteellinen ja oston todennäköisyys pienenee hinnan kasvaessa
- Esim. neuraaliverkoissa paljon käytetty pehmeän kynnyksen määrittelytapa on ns. logit-jakauma, joka käyttää sigmoidi-funktiota

$$P(\text{ostaa} \mid h) = 1/(1 + \exp(-2(-h+ \mu)/\sigma))$$

- Logit-jakaumalla on pidemmät hännät kuin probit-jakaumalla
- Probit on käytännössä tarkempi kuin logit, joka taas puolestaan on helpompi käsitellä matemaattisesti ja siksi yleinen

Tarkka päättely Bayes-verkoissa

- Tavoitteemme on selvittää kyselymuuttujan X posteriorijakauma kun havaintomuuttujien $\mathbf{E} = E_1, \ldots, E_m$ havaitut arvot ovat $\mathbf{e}$ ja piilomuuttujat ovat $\mathbf{Y} = Y_1, \ldots, Y_l$
- Täydestä yhteisjakaumasta voimme vastata kyselyyn $\underline{P}(X \mid \mathbf{e})$ laskemalla

$$\alpha\, \underline{P}(X, \mathbf{e}) = \alpha \textstyle\sum_{\mathbf{y}} \underline{P}(X, \mathbf{e}, \mathbf{y})$$

- Bayes-verkko on yhteisjakauman täysi esitys, erityisesti tekijät $P(X, \mathbf{e}, \mathbf{y})$ voidaan ilmaista verkon ehdollisten todennäköisyyksien tulona
- Täten verkon perusteella voidaan vastata annettuun kyselyyn

- Kyselyn $\underline{P}(Murto \mid jussikäy, meerikäy)$ arvottaminen
- Piilomuuttujat ovat $Järistys$ ja $Hälytys$

$$\underline{P}(Murto \mid jussikäy, meerikäy) = \alpha \textstyle\sum_j \sum_h \underline{P}(Murto, j, h, jussikäy, meerikäy)$$

- Verkon ehdollisten todennäköisyystaulukoiden perusteella tarvittavat arvot voidaan laskea, esim.

$$P(murto \mid jussikäy, meerikäy) = \alpha \textstyle\sum_j \sum_h P(murto)\, P(j)\, P(h \mid murto, j)\, P(jussikäy \mid h)\, P(meerikäy \mid h)$$

- Tekijöitä järjestelemällä saadaan tehokkaammin laskettava kaava

$$\alpha\, P(murto) \textstyle\sum_j P(j) \sum_h P(h \mid murto, j)\, P(jussikäy \mid h)\, P(meerikäy \mid h)$$

- Laskemalla P(murto | jussikäy, meerikäy) ja P(¬murto | jussikäy, meerikäy), saadaan aiemmin käyttämillämme arvoilla

  P(Murto | jussikäy, meerikäy) ≈ [0.284, 0.716]
- Bayes-verkon DAGin evaluoiminen vastaa puun syvyyssuuntaista läpikäyntiä, joten sen tilavaativuus on vain lineaarinen muuttujien lukumäärän suhteen
- Aikavaativuus n:n muuttujan verkolle on aina $O(2^n)$
- Esim. tulo P(jussikäy | h) P(meerikäy | h) joudutaan laskemaan uudestaan jokaiselle arvolle j
- Toistuvien alilauseiden uudelleenevaluoinnin välttämisellä voidaan saavuttaa lisää säästöä