Tartu Ülikool
Loodus- ja täppisteaduste valdkond
Matemaatika ja statistika instituut

Lisette Pajula

Kompuutertomograafia piltide hindamine magnetresonantstomograafia piltide põhjal varjatud Markovi mudeli abil

Matemaatika ja statistika eriala

Magistritöö (30 EAP)

Juhendaja: PhD Kristi Kuljus

Tartu 2021

KOMPUUTERTOMOGRAAFIA PILTIDE HINDAMINE MAGNETRESONANTSTOMOGRAAFIA PILTIDE PÕHJAL VARJATUD MARKOVI MUDELI ABIL

Magistritöö

Lisette Pajula

Lühikokkuvõte

Magnetresonatstomograafia (MRT) ja kompuutertomograafia (KT) on kaks erinevat diagnostilise uuringu tüüpi, mis võimaldavad keha eri piirkondadest kujutisi saada. Magistritöö eesmärk on KT vaatluste hindamine MRT vaatluste põhjal ehk nii-öelda substituut-KT-pildi leidmine. Selleks kasutatakse varjatud Markovi mudeli erijuhtu, kus eeldatakse MRT ja KT vaatluste tinglikku sõltumatust, kui vastavate varjatud tunnuste väärtused ehk kudede klassid on teada. Töös tuuakse ülevaade nii klassikalisest varjatud Markovi mudelist kui ka kirjeldatud erijuhust, mida töös nimetatakse tingliku sõltumatuse mudeliks. Substituut-KT-pildi leidmist tingliku sõltumatuse mudeli abil näitlikustatakse viie pea andmete põhjal. Parameetrite hindamiseks kasutatakse EM-algoritmi.

CERCS teaduseriala: P160 Statistika, operatsioonianalüüs, programmeerimine, finants- ja kindlustusmatemaatika.

Märksõnad: varjatud Markovi mudel, Markovi ahel, EM-algoritm, kompuutertomograafia, magnetresonantstomograafia.

ESTIMATION OF COMPUTED TOMOGRAPHY IMAGES BASED ON MAGNETIC RESONANCE IMAGES USING A HIDDEN MARKOV MODEL

Master's thesis

Lisette Pajula

Abstract

Magnetic resonance imaging (MRI) and computed tomography (CT) are two different methods of medical diagnostics that enable to image the anatomy of the human body. The aim of this Master's thesis is to estimate CT observations based on MRI observations – that is, to calculate a so-called pseudo-CT image. For that a special case of the hidden Markov model is investigated, which assumes conditional independence of MRI and CT observations if the corresponding hidden states – tissue classes – are known. The thesis provides an overview of the classical hidden Markov model, followed by a description of the forementioned special case which is named as the conditional independence model. Finding pseudo-CT images using the conditional independence model is demonstrated on data consisting of CT and MRI measurements of five heads. For parameter estimation the EM algorithm is used.

CERCS research specialisation: P160 Statistics, operations research, programming, financial and actuarial mathematics.

Key Words: hidden Markov model, Markov chain, EM algorithm, computed tomography, magnetic resonance imaging.

Sisukord

Sissejuhatus

Magnetresonatstomograafia (MRT) ja kompuutertomograafia (KT) on kaks erinevat diagnostilise uuringu tüüpi, mis võimaldavad keha eri piirkondadest saada nii kahe- kui ka kolmemõõtmelisi kujutisi. Mõlemal on omad eelised ja puudused. MRT-pilt saadakse tugeva magnetvälja abil ning tegemist on väga hea meetodiga pehmete kudede, sealhulgas kasvajate uurimiseks ja eristamiseks. Seevastu luude jäädvustamiseks eelistatakse pigem KT-uuringuid, kuna MRT nende kohta samaväärset informatsiooni ei anna.

KT-pildi tegemine kujutab endast aga patsiendi tervisele suuremat riski, kuna põhineb röntgenkiirguse kasutamisel. Seega pakub huvi olukord, kus saaksime MRT-piltidelt samaväärse informatsiooni kätte ilma KT-pilti tegelikult tegemata. Taoline lähenemine on aktuaalne näiteks kiiritusravis dooside planeerimisel. Kui patsiendile on juba MRT-uuring tehtud, siis rohkemate lisaprotseduuride ja kõrvalmõjude vältimiseks oleks hea kasutada nii-öelda substituut-KT-pilti, mis on leitud olemasolevate MRT-piltide põhjal.

Töös uuritakse ühte võimalust substituut-KT-piltide leidmiseks, mis põhineb varjatud Markovi mudeli kasutamisel. Kudede klasse käsitletakse kui varjatud tunnuseid, millest sõltuvad nii MRT kui ka KT vaatlused. Lisaks tavapärastele varjatud Markovi mudeli eeldustele käsitletakse MRT ja KT vaatluseid tinglikult sõltumatutena ehk eeldatakse, et need sõltuvad teineteisest vaid varjatud tunnuste kaudu. Sellise eelduse tegemine on tingitud praktilistest kaalutlustest, sest tegelikult ongi MRT ja KT niivõrd erinevad tehnikad, et vaatlused võiksid üksteisest sõltuda vaid varjatud kudede klasside kaudu. Teatud lisaeeldustel minnakse üle mudelile, mis lubab MRT ja KT vaatluste jaoks vaadelda erinevat arvu seisundeid ehk kudede klasse. Kirjeldatud mudelit nimetatakse töös tingliku sõltumatuse mudeliks.

Magistritöö eesmärk on KT vaatluste hindamine MRT vaatluste põhjal, kasutades selleks tingliku sõltumatuse mudelit. Töö koosneb kolmest peatükist. Alustuseks tehakse tutvust klassikalise varjatud Markovi mudeliga ning selle parameetrite hindamiseks kasutatava EM-algoritmiga. Teises peatükis antakse ülevaade tingliku sõltumatuse mudelist ning tuuakse selle paremaks mõistmiseks ka lihtne näide simuleeritud andmetega. Viimases peatükis rakendatakse mudelit viie erineva pea andmetele, kasutades nelja neist treeningandmestikuna, et leida viiendale peale substituut-KT-pilt. Tulemusi võrreldakse ka artikliga Kuljus *et al.* (2018), kus on samale probleemile lähenetud teisi varjatud tunnustega mudeleid kasutades.

Töö on vormistatud tekstitöötlusprogrammiga LaTeX. Tingliku sõltumatuse mudeli rakendamiseks ja jooniste tegemiseks on kasutatud statistikatarkvara R (versioon 3.6.3).

Autor tänab juhendajat Kristi Kuljust töösse panustatud aja ning arvukate harivate selgituste eest.

1 Varjatud Markovi mudel

Tegeleme töös MRT- ja KT-piltide mõõtmistega, mille väärtused sõltuvad teadupärast uuritava piirkonna kudede klassidest. Milliste kudedega keha eri piirkondades täpsemalt tegemist on, me otseselt mõõta ei saa. Üks võimalus selliste andmete modelleerimiseks on kasutada varjatud Markovi mudeleid. Tutvume alustuseks klassikalise varjatud Markovi mudeliga, mille kirjeldus tugineb artiklile Rabiner (1989). Et MRT ja KT mõõtmiste puhul on tegemist pidevate tunnustega, keskendume pidevate emissioonijaotustega juhule.

Vaatleme juhuslike suuruste jada $Z^T = (Z_1, \ldots, Z_T)$, kus iga Z_t, $t = 1, \ldots, T$, võtab väärtusi hulgast $S = \{1, \ldots, K\}$. Hulga S elemente nimetame seisunditeks. Kehtigu

$$P(Z_t = z_t | Z_1 = z_1, \ldots, Z_{t-1} = z_{t-1}) = P(Z_t = z_t | Z_{t-1} = z_{t-1}),$$
$$z_1, \ldots, z_t \in S, \quad t = 2, \ldots, T,$$

siis juhuslike suuruste jada Z^T nimetatakse (esimest järku) Markovi ahelaks. Tõenäosuseid $\pi = (\pi_1, \ldots, \pi_K)$, kus

$$\pi_i = P(Z_1 = i), \quad i \in S,$$

nimetatakse algtõenäosusteks. Need näitavad, kui suure tõenäosusega on mingi seisund jada algseisundiks, seega $\sum_{i=1}^{K} \pi_i = 1$. Ühest seisundist teise liikumise tõenäosusi nimetatakse üleminekutõenäosusteks. Vaatleme homogeenset Markovi ahelat, kus üleminekutõenäosused p_{ij} ei sõltu t väärtusest:

$$p_{ij} = P(Z_t = j | Z_{t-1} = i), \quad i, j \in S, \ \forall t \in \{2, \ldots, T\}.$$

Üleminekutõenäosuseid on tavaline esitada maatriksina $\mathrm{P} = (p_{ij})$, seega iga rea i korral kehtib $\sum_{j=1}^{K} p_{ij} = 1$.

Eeldame, et Z^T realisatsioonid ei ole otseselt mõõdetavad ehk seisundid on varjatud. Olgu $X^T = (X_1, \ldots, X_T)$ vaadeldud andmed, mille väärtused $x^T = (x_1, \ldots, x_T)$ sõltuvad varjatud ahelast Z^T järgmiselt:

1) antud $z^T = (z_1, \ldots, z_T)$ korral on $X_1, \ldots, X_T$ tinglikult sõltumatud ehk tõepära $p(x_1, \ldots, x_T | Z_1 = z_1, \ldots, Z_T = z_T)$ avaldub kujul

$$\begin{aligned} &p(x_1, \ldots, x_T | Z_1 = z_1, \ldots, Z_T = z_T) \\ &= \prod_{t=1}^{T} p(x_t | Z_1 = z_1, \ldots, Z_T = z_T), \end{aligned}$$

2) iga t korral sõltub X_t tinglik jaotus vaid seisundist $Z_t = z_t$ ehk

$$p(x_t | Z_1 = z_1, \ldots, Z_T = z_T) = p(x_t | Z_t = z_t).$$

Vaatluste jaotust seisundi $Z_t = i$ korral kirjeldab tihedusfunktsioon:

$$f_i(x|\theta_i) = p(x|Z_t = i, \theta_i), \quad i \in S.$$

Tinglikke vaatluste jaotusi nimetatakse emissioonijaotusteks ning mudelit (X^T, Z^T) tervikuna varjatud Markovi mudeliks, lühendatult HMM (ingl *hidden Markov model*). Mudeli kõiki parameetreid kokku tähistame

$$\theta = (\pi, \mathrm{P}, \theta_1, \ldots, \theta_K).$$

Antud temaatika kontekstis saame seega kudede klasse käsitleda kui varjatud tunnuseid Z^T, mis on vaadeldavad vaid kaudselt näiteks MRT või KT vaatluste kaudu. HMM parameetrite hindamiseks kasutatakse tavaliselt EM-algoritmi.

1.1 EM-algoritm

Peatükk põhineb allikatel Bilmes (1998) ja Izenman (2008).

EM-algoritm on iteratiivne meetod suurima tõepära hinnangute leidmiseks. Algoritmi nimetus tuleneb iteratsiooni kahest sammust: E-samm (ingl *expectation step*) ja M-samm (ingl *maximization step*). Algoritm võimaldab leida hinnangud mudeli parameetritele, kui vaadeldud andmed sõltuvad varjatud tunnustest või sisaldavad puuduvaid väärtuseid.

Olgu $\mathcal{X} = (x_1, \ldots, x_n)$ mingi jaotuse poolt genereeritud andmed, mille tõepärafunktsiooniks on

$$\mathcal{L}(\Theta|\mathcal{X}) = p(\mathcal{X}|\Theta),$$

kus Θ tähistab kõiki mudeli parameetreid, mida soovime hinnata. Oletame, et vaatlused $\mathcal{X}$ sõltuvad varjatud tunnustest $\mathcal{Z} = (z_1, \ldots, z_n)$. Andmestikku $(\mathcal{X}, \mathcal{Z})$ nimetame täielikuks. Vaatluste ja varjatud tunnuste ühistihedus avaldub kujul

$$p(\mathbf{x}, \mathbf{z}|\Theta) = p(\mathbf{z}|\mathbf{x}, \Theta)p(\mathbf{x}|\Theta),$$

kus $\mathbf{x}$ tähistab mingit andmete komplekti ning $\mathbf{z}$ andmetega seotud varjatud tunnuste väärtuseid. Vastavat tõepärafunktsiooni nimetame samuti täielikuks:

$$\mathcal{L}(\Theta|\mathcal{X}, \mathcal{Z}) = p(\mathcal{X}, \mathcal{Z}|\Theta) = p(x_1, \ldots, x_n, z_1, \ldots, z_n|\Theta).$$

Kuna tunnused $\mathcal{Z}$ on varjatud, ei saa me neid otse mõõta. Teada on aga $\mathcal{Z}$ tinglik jaotus $p(\mathbf{z}|\mathcal{X},\Theta)$, mis võimaldab meil täieliku tõepärafunktsiooni keskväärtust leida.

Olgu $\mathscr{Z}$ kõikvõimalike $\mathbf{z}$ väärtuste hulk. EM-algoritmi esimeseks sammuks (E-samm) on täielike andmete log-tõepära tingliku keskväärtuse leidmine varjatud andmete $\mathcal{Z}$ tingliku jaotuse suhtes, kasutades alglähendeid Θ' ja vaadeldud andmete $\mathcal{X}$ väärtuseid. Defineerime selleks Q-funktsiooni:

$$Q(\Theta,\Theta') = E\left[\log \mathcal{L}(\Theta|\mathcal{X},\mathcal{Z})|\mathcal{X},\Theta'\right] = \int_{\mathbf{z}\in\mathscr{Z}} \log p(\mathcal{X},\mathbf{z}|\Theta)p(\mathbf{z}|\mathcal{X},\Theta')d\mathbf{z}.$$

Algoritmi teiseks sammuks (M-samm) on leitud keskväärtuse ehk Q-funktsiooni maksimiseerimine hinnatavate parameetrite suhtes:

$$\Theta'' = \operatorname*{argmax}_{\Theta} Q(\Theta,\Theta').$$

Järgmisel E-sammul kasutatakse alglähendite asemel eelmise iteratsiooni M-sammul leitud hinnanguid. Neid kahte sammu korratakse algoritmi koondumiseni. Näitame, et probleemile selliselt Q-funktsiooni kaudu lähenemine annab soovitud tulemuse ehk vaatluste tõepära kasvab iga iteratsiooniga (täpsemini, ei kahane).

Lause 1. Olgu $\Theta^{(i)}$ ja $\Theta^{(i-1)}$ iteratsioonidel i ja $i-1$ leitud parameetrite hinnangud ehk $\Theta^{(i)} = \operatorname{argmax}_{\Theta} Q(\Theta,\Theta^{(i-1)})$, siis

$$\mathcal{L}(\Theta^{(i)}|\mathcal{X}) \geqslant \mathcal{L}(\Theta^{(i-1)}|\mathcal{X}).$$

Tõestus. Tähistame vaatluste log-tõepära $\ell(\Theta|\mathcal{X}) = \log \mathcal{L}(\Theta|\mathcal{X})$. Kehtib

$$p(\mathcal{Z}|\mathcal{X}, \Theta) = \frac{p(\mathcal{X}, \mathcal{Z}|\Theta)}{p(\mathcal{X}|\Theta)},$$

seega vaatluste log-tõepära avaldub kujul

$$\ell(\Theta|\mathcal{X}) = \log p(\mathcal{X}|\Theta) = \log p(\mathcal{X}, \mathcal{Z}|\Theta) - \log p(\mathcal{Z}|\mathcal{X}, \Theta).$$

Võttes avaldisest keskväärtuse jaotuse $p(\mathbf{z}|\mathcal{X}, \Theta')$ suhtes, saame

$$\ell(\Theta|\mathcal{X}) = Q(\Theta, \Theta') - H(\Theta, \Theta'),$$

kus $H(\Theta, \Theta') = E[\log p(\mathcal{Z}|\mathcal{X}, \Theta)|\mathcal{X}, \Theta']$. Võtame kasutusele abifunktsiooni:

$$h(\mathcal{Z}) := \frac{p(\mathcal{Z}|\mathcal{X}, \Theta)}{p(\mathcal{Z}|\mathcal{X}, \Theta')},$$

siis

$$\begin{aligned} H(\Theta, \Theta') - H(\Theta', \Theta') &= E[\log h(\mathcal{Z})|\mathcal{X}, \Theta'] \\ &\leqslant E[h(\mathcal{Z})|\mathcal{X}, \Theta'] - 1 \\ &= 0, \end{aligned}$$

kus kasutasime võrratust $\log x \leqslant x - 1$. Järelikult $H(\Theta, \Theta') \leqslant H(\Theta', \Theta')$. Vaatleme andmete log-tõepärade vahet iteratsioonidel i ja $i-1$ leitud parameetritega:

$$\begin{aligned} \ell(\Theta^{(i)}|\mathcal{X}) - \ell(\Theta^{(i-1)}|\mathcal{X}) = Q(\Theta^{(i)}, \Theta^{(i-1)}) - H(\Theta^{(i)}, \Theta^{(i-1)}) \\ - Q(\Theta^{(i-1)}, \Theta^{(i-1)}) + H(\Theta^{(i-1)}, \Theta^{(i-1)}) \end{aligned}$$

$$\geqslant Q(\Theta^{(i)}, \Theta^{(i-1)}) - Q(\Theta^{(i-1)}, \Theta^{(i-1)})$$
$$\geqslant 0.$$

Viimane võrratus kehtib, kuna EM-algoritm leiab sellise $\Theta^{(i)}$, mis maksimiseerib Q-funktsiooni. Järelikult, $\ell(\Theta^{(i)}|\mathcal{X}) \geqslant \ell(\Theta^{(i-1)}|\mathcal{X})$ ehk ka $\mathcal{L}(\Theta^{(i)}|\mathcal{X}) \geqslant \mathcal{L}(\Theta^{(i-1)}|\mathcal{X})$.

$\square$

EM-algoritm koondub lokaalseks maksimumiks. Tulemus sõltub alglähendite valikust, mistõttu globaalse maksimumi leidmiseks tuleks läbi proovida mitmeid erinevaid algparameetrite komplekte.

1.2 HMM parameetrite hinnangute leidmine

Tuletame meelde, et vaatleme varjatud Markovi mudelit, kus $X^T = (X_1, \ldots, X_T)$ tähistab vaadeldud andmeid ning $Z^T = (Z_1, \ldots, Z_T)$ vastavat varjatud olekute järjendit. Varjatud seisundite hulk on $S = \{1, \ldots, K\}$. Algtõenäosuseid tähistame

$$\pi_i = P(Z_1 = i|\theta), \quad i \in S,$$

ja üleminekutõenäosuseid iga $t \in \{1, \ldots, T\}$ korral

$$p_{ij} = P(Z_{t+1} = j|Z_t = i, \theta), \quad i, j \in S.$$

Vaatluste emissioonijaotusteks olgu $f_i(x|\theta_i)$, $i \in S$. Näitame, kuidas avalduvad EM-algoritmi hinnangud varjatud Markovi mudeli parameetritele, kui on teada eelmisel sammul leitud parameetrite hinnangud (või alglähendid) θ'.

Hinnangute avaldiste tuletamisel on kasutatud ideid allikast Bilmes (1998).

Tuletatud parameetrite hinnanguid, mis maksimiseerivad Q-funktsiooni, tähistame esialgu samamoodi kui funktsiooni argumente. Tähistuste ühtlustamiseks kasutame edaspidi arvutustes ka tõenäosuste jaoks üldist tõepära tähistust $p(\cdot)$, sama kehtib ka peatükis 2.

Meie mudelile vastav Q-funktsioon on kujul

$$Q(\theta, \theta') = \sum_{\mathcal{Z}} \log p(x^T, z^T|\theta) p(z^T|x^T, \theta'),$$

kus θ tähistab kõiki mudeli parameetreid, θ' on meie esialgne parameetrite hinnang ning summeerimine toimub üle $z^T \in \mathcal{Z}$ ehk üle kõigi pikkusega T varjatud olekute jadade. Konkreetse järjendi z^T korral avaldub $p(x^T, z^T|\theta)$ järgmiselt:

$$p(x^T, z^T|\theta) = p(z^T|\theta)p(x^T|z^T, \theta) = \pi_{z_1} \prod_{t=1}^{T-1} p_{z_t z_{t+1}} \prod_{t=1}^{T} f_{z_t}(x_t|\theta_{z_t}).$$

Seega Q-funktsioon avaldub kujul

$$\begin{aligned} Q(\theta, \theta') = &\sum_{\mathcal{Z}} \log \pi_{z_1} p(z^T|x^T, \theta') + \sum_{\mathcal{Z}} \sum_{t=1}^{T-1} \log p_{z_t z_{t+1}} p(z^T|x^T, \theta') \\ &+ \sum_{\mathcal{Z}} \sum_{t=1}^{T} \log f_{z_t}(x_t|\theta_{z_t}) p(z^T|x^T, \theta'). \end{aligned}$$

Näeme, et kõik parameetrid, üle mille soovime funktsiooni Q maksimiseerida, on avaldunud üksteisest eraldi sõltumatutes liidetavates, mida saame ükshaaval maksimiseerida.

Algtõenäosuste ja üleminekutõenäosuste hindamine

Alg- ja üleminekutõenäosuste hindamisel tegeleme Q-funktsiooni kahe esimese liidetavaga. Vaatleme esimest liidetavat:

$$\begin{aligned}
\sum_{\mathcal{Z}} \log \pi_{z_1} p(z^T|x^T,\theta') &= \sum_{z_1=1}^{K}\sum_{z_2=1}^{K}\dots\sum_{z_T=1}^{K} \log \pi_{z_1} p(z_1\dots,z_T|x^T,\theta') \\
&= \sum_{z_1=1}^{K} \log \pi_{z_1} \sum_{z_2=1}^{K}\dots\sum_{z_T=1}^{K} p(z_1,z_2\dots,z_T|x^T,\theta') \\
&= \sum_{z_1=1}^{K} \log \pi_{z_1} p(z_1|x^T,\theta') = \sum_{i=1}^{K} \log \pi_i p(Z_1=i|x^T,\theta').
\end{aligned}$$

Parameetrite π_i hinnangute leidmisel kasutame kitsenduse $\sum_{i=1}^{K}\pi_i = 1$ arvesse võtmiseks Lagrange'i kordajate meetodit:

$$\frac{\partial}{\partial \pi_i}\left(\sum_{i=1}^{K} \log \pi_i p(Z_1=i|x^T,\theta') + \gamma\Big(\sum_{i=1}^{K}\pi_i - 1\Big)\right) = 0,$$
$$\frac{p(Z_1=i|x^T,\theta')}{\pi_i} + \gamma = 0 \quad\Rightarrow\quad \gamma\pi_i = -p(Z_1=i|x^T,\theta').$$

Summeerime üle i, et leida Lagrange'i kordaja γ:

$$\sum_{i=1}^{K}\gamma\pi_i = \sum_{i=1}^{K} -p(Z_1=i|x^T,\theta') \quad\Rightarrow\quad \gamma = -1.$$

Seega π_i avaldub kujul

$$\pi_i = p(Z_1=i|x^T,\theta').$$

Vaatleme nüüd teist liidetavat. Fikseerime $t \in \{2,\dots,T-2\}$ (tuletuskäigu lihtsustamiseks väldime juhte $t=1$ ja $t=T-1$, kuid on ilmne, et tulemus kehtib ka nende korral) ja summerime esmalt üle $\mathcal{Z}$, siis:

$$
\begin{aligned}
&\sum_{\mathcal{Z}} \log p_{z_t z_{t+1}} p(z^T|x^T,\theta') \\
&= \sum_{z_1=1}^{K} \dots \sum_{z_t=1}^{K} \sum_{z_{t+1}=1}^{K} \dots \sum_{z_T=1}^{K} \log p_{z_t z_{t+1}} p(z_1,\dots,z_t,z_{t+1},\dots,z_T|x^T,\theta') \\
&= \sum_{z_t=1}^{K} \sum_{z_{t+1}=1}^{K} \log p_{z_t z_{t+1}} p(z_t,z_{t+1}|x^T,\theta') \\
&= \sum_{i=1}^{K} \sum_{j=1}^{K} \log p_{ij} p(Z_t=i, Z_{t+1}=j|x^T,\theta').
\end{aligned}
$$

Seega teine liidetav on kujul:

$$
\begin{aligned}
\sum_{\mathcal{Z}} \sum_{t=1}^{T-1} \log p_{z_t z_{t+1}} p(z^T|x^T,\theta') &= \sum_{t=1}^{T-1} \sum_{\mathcal{Z}} \log p_{z_t z_{t+1}} p(z^T|x^T,\theta') \\
&= \sum_{t=1}^{T-1} \sum_{i=1}^{K} \sum_{j=1}^{K} \log p_{ij} p(Z_t=i, Z_{t+1}=j|x^T,\theta').
\end{aligned}
$$

Kasutame jällegi p_{ij} avaldiste leidmisel igale üleminekumaatriksi reale kehtiva kitsenduse $\sum_{j=1}^{K} p_{ij} = 1$ arvesse võtmiseks Lagrange'i kordajate meetodit:

$$
\begin{aligned}
\frac{\partial}{\partial p_{ij}} \left(\sum_{t=1}^{T-1} \sum_{i=1}^{K} \sum_{j=1}^{K} \log p_{ij} p(Z_t=i, Z_{t+1}=j|x^T,\theta') + \gamma_i \Big(\sum_{j=1}^{K} p_{ij} - 1 \Big) \right) &= 0, \\
\sum_{t=1}^{T-1} \frac{p(Z_t=i, Z_{t+1}=j|x^T,\theta')}{p_{ij}} + \gamma_i &= 0, \\
\gamma_i p_{ij} &= -\sum_{t=1}^{T-1} p(Z_t=i, Z_{t+1}=j|x^T,\theta').
\end{aligned}
$$

Summeerime üle j, et leida Lagrange'i kordaja γ_i:

$$\sum_{j=1}^{K} \gamma_i p_{ij} = -\sum_{j=1}^{K} \sum_{t=1}^{T-1} p(Z_t = i, Z_{t+1} = j | x^T, \theta'),$$
$$\gamma_i = -\sum_{t=1}^{T-1} p(Z_t = i | x^T, \theta').$$

Järelikult, p_{ij} avaldub kujul

$$p_{ij} = \frac{\sum_{t=1}^{T-1} p(Z_t = i, Z_{t+1} = j | x^T, \theta')}{\sum_{t=1}^{T-1} p(Z_t = i | x^T, \theta')}.$$

Emissioonijaotuste parameetrite hindamiseks tuleb vaadata kolmandat Q-funktsiooni liidetavat, mis lihtsustub järgmiselt:

$$\sum_{\mathcal{Z}} \sum_{t=1}^{T} \log f_{z_t}(x_t | \theta_{z_t}) p(z^T | x^T, \theta')$$
$$= \sum_{t=1}^{T} \sum_{z_1=1}^{K} \cdots \sum_{z_t=1}^{K} \cdots \sum_{z_T=1}^{K} \log f_{z_t}(x_t | \theta_{z_t}) p(z_1, \ldots, z_t, \ldots, z_T | x^T, \theta')$$
$$= \sum_{t=1}^{T} \sum_{z_t=1}^{K} \log f_{z_t}(x_t | \theta_{z_t}) p(z_t | x^T, \theta') = \sum_{t=1}^{T} \sum_{i=1}^{K} \log f_i(x_t | \theta_i) p(Z_t = i | x^T, \theta').$$

Emissioonijaotuste parameetrite hindamine normaaljaotuste segu korral

On selge, et f_i, $i \in S$, parameetrite hindamine sõltub sellest, millise jaotuste klassiga on tegemist. Vaatleme täpsemalt olukorda, kus emissioonijaotuseks on iga oleku korral M komponendiga Gaussi segumudel:

$$f_i(x | \theta_i) = \sum_{l=1}^{M} c_{il} f_{il}(x | \theta_{i\ell}) = \sum_{l=1}^{M} c_{il} \phi(x | \mu_{il}, \Sigma_{il}),$$

seega $c_{i\ell}$ tähistab komponendi ℓ kaalu tiheduses f_i ning $\theta_i = (\theta_{i1}, \dots, \theta_{iM})$, kus $\theta_{i\ell} = (\mu_{i\ell}, \Sigma_{i\ell})$ tähistab f_i komponendi $f_{i\ell}$ parameetreid, $\ell \in \{1, \dots, M\}$. Kuna nüüd on meil kaks varjatud tunnuste jada, peame ka Q-funktsiooni vastavalt modifitseerima:

$$Q(\theta, \theta') = \sum_{\mathcal{Z}} \sum_{\mathcal{M}} \log p(x^T, z^T, m^T|\theta) p(z^T, m^T|x^T, \theta'),$$

kus varjatud seisundite jada $M^T = (M_1, \dots, M_T)$ väärtus $m^T = (m_1, \dots, m_T)$ näitab, millise segujaotuse komponendiga iga ajahetke korral parasjagu tegemist on. Summeerimine toimub nüüd ka üle kõikvõimalike m^T väärtuste hulga $\mathcal{M}$. Võttes arvesse, et $c_{z_t m_t} = P(M_t = m_t|Z_t = z_t)$, saame $p(x^T, z^T, m^T|\theta)$ esitada konkreetsete z^T ja m^T korral järgmiselt:

$$\begin{aligned} p(x^T, z^T, m^T|\theta) &= p(z^T|\theta) p(m^T|z^T, \theta) p(x^T|m^T, z^T, \theta) \\ &= \pi_{z_1} \prod_{t=1}^{T-1} p_{z_t z_{t+1}} \prod_{t=1}^{T} c_{z_t m_t} f_{z_t m_t}(x_t|\theta_{z_t m_t}). \end{aligned}$$

Varjatud tunnuste jaotuse kohta kehtib

$$\sum_{\mathcal{M}} p(z^T, m^T|x^T, \theta) = p(z^T|x^T, \theta).$$

Seega, kui emissioonijaotusteks on normaaljaotuste segu ja meil on kaks varjatud seisundite ahelat, avaldub Q-funktsioon kujul

$$\begin{aligned} Q(\theta, \theta') = &\sum_{\mathcal{Z}} \sum_{\mathcal{M}} \log \pi_{z_1} p(z^T, m^T|x^T, \theta') + \sum_{\mathcal{Z}} \sum_{\mathcal{M}} \sum_{t=1}^{T-1} \log p_{z_t z_{t+1}} p(z^T, m^T|x^T, \theta') \\ &+ \sum_{\mathcal{Z}} \sum_{\mathcal{M}} \sum_{t=1}^{T} \log(c_{z_t m_t} f_{z_t m_t}(x_t|\theta_{z_t m_t})) p(z^T, m^T|x^T, \theta') \end{aligned}$$

$$= \sum_{\mathcal{Z}} \log \pi_{z_1} p(z^T|x^T, \theta') + \sum_{\mathcal{Z}} \sum_{t=1}^{T-1} \log p_{z_t z_{t+1}} p(z^T|x^T, \theta')$$

$$+ \sum_{\mathcal{Z}} \sum_{\mathcal{M}} \sum_{t=1}^{T} \log(c_{z_t m_t} f_{z_t m_t}(x_t|\theta_{z_t m_t})) p(z^T, m^T|x^T, \theta').$$

Näeme, et esimene ja teine liidetav on samaks jäänud, kuna nendes olevad parameetrid ei sõltu m^T väärtusest. Kolmas liidetav avaldub aga järgmiselt:

$$\sum_{\mathcal{Z}} \sum_{\mathcal{M}} \sum_{t=1}^{T} \log(c_{z_t m_t} f_{z_t m_t}(x_t|\theta_{z_t m_t})) p(z^T, m^T|x^T, \theta')$$

$$= \sum_{i=1}^{K} \sum_{\ell=1}^{M} \sum_{t=1}^{T} \log(c_{i\ell} f_{i\ell}(x_t|\theta_{i\ell})) p(Z_t = i, M_t = \ell|x^T, \theta')$$

$$= \sum_{i=1}^{K} \sum_{\ell=1}^{M} \sum_{t=1}^{T} \log c_{i\ell} p(Z_t = i, M_t = \ell|x^T, \theta')$$

$$+ \sum_{i=1}^{K} \sum_{\ell=1}^{M} \sum_{t=1}^{T} \log f_{i\ell}(x_t|\theta_{i\ell}) p(Z_t = i, M_t = \ell|x^T, \theta').$$

Saame jällegi kahe liidetavaga eraldi tegeleda. Kasutame $c_{i\ell}$ optimaalse väärtuse leidmiseks Lagrange'i kordajate meetodit kitsendusega $\sum_{\ell=1}^{M} c_{i\ell} = 1$, siis:

$$\frac{\partial}{\partial c_{i\ell}} \left(\sum_{i=1}^{K} \sum_{\ell=1}^{M} \sum_{t=1}^{T} \log c_{i\ell} p(Z_t = i, M_t = \ell|x^T, \theta') + \gamma_i \Big(\sum_{\ell=1}^{M} c_{i\ell} - 1 \Big) \right) = 0,$$

$$\sum_{t=1}^{T} \frac{p(Z_t = i, M_t = \ell|x^T, \theta')}{c_{i\ell}} + \gamma_i = 0,$$

$$\gamma_i c_{i\ell} = - \sum_{t=1}^{T} p(Z_t = i, M_t = \ell|x^T, \theta').$$

Summeerime üle kõigi ℓ võimalike väärtuste:

$$\sum_{\ell=1}^{M} \gamma_i c_{i\ell} = -\sum_{\ell=1}^{M}\sum_{t=1}^{T} p(Z_t = i, M_t = \ell | x^T, \theta'),$$
$$\gamma_i = -\sum_{t=1}^{T} p(Z_t = i | x^T, \theta').$$

Seega

$$c_{i\ell} = \frac{\sum_{t=1}^{T} p(Z_t = i, M_t = \ell | x^T, \theta')}{\sum_{t=1}^{T} p(Z_t = i | x^T, \theta')}.$$

Leiame ka emissioonijaotuse f_i komponendi ℓ parameetrid $\theta_{i\ell}$. Vastava (d-mõõtmelise) normaaljaotuse tihedus on kujul:

$$f_{i\ell}(x|\theta_{i\ell}) = \frac{1}{(2\pi)^{d/2}|\Sigma_{i\ell}|^{1/2}} \cdot e^{-\frac{1}{2}(x-\mu_{i\ell})^T \Sigma_{i\ell}^{-1}(x-\mu_{i\ell})}.$$

Võttes logaritmi, saame

$$\log f_{i\ell}(x|\theta_{i\ell}) = -\frac{d}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma_{i\ell}| - \frac{1}{2}(x-\mu_{i\ell})^T \Sigma_{i\ell}^{-1}(x-\mu_{i\ell}).$$

Paneme tähele, et avaldise alguses olev konstant on üle $\mu_{i\ell}$ ja $\Sigma_{i\ell}$ maksimiseerimisel ebaoluline. Kokkuvõttes on Q-funktsiooni vastava liidetava maksimiseeritav osa kujul

$$\sum_{i=1}^{K}\sum_{\ell=1}^{M}\sum_{t=1}^{T}\left(-\frac{1}{2}\log|\Sigma_{i\ell}| - \frac{1}{2}(x_t-\mu_{i\ell})^T\Sigma_{i\ell}^{-1}(x_t-\mu_{i\ell})\right) p(Z_t = i, M_t = \ell | x^T, \theta').$$

Keskväärtuste hinnangute leidmiseks võtame sellest $\mu_{i\ell}$ järgi tuletise ja võrdsustame nulliga, mis annab meile:

$$\sum_{t=1}^{T} \frac{1}{2} \left(\Sigma_{i\ell}^{-1} + (\Sigma_{i\ell}^{-1})^T \right) (x_t - \mu_{i\ell}) p(Z_t = i, M_t = \ell | x^T, \theta') = 0,$$

$$\sum_{t=1}^{T} \Sigma_{i\ell}^{-1} (x_t - \mu_{i\ell}) p(Z_t = i, M_t = \ell | x^T, \theta') = 0.$$

Avaldame viimasest $\mu_{i\ell}$:

$$\mu_{i\ell} = \frac{\sum_{t=1}^{T} x_t p(Z_t = i, M_t = \ell | x^T, \theta')}{\sum_{t=1}^{T} p(Z_t = i, M_t = \ell | x^T, \theta')}.$$

Kovariatsioonimaatriksite $\Sigma_{i\ell}$ leidmiseks kirjutame vaadeldava avaldise ümber:

$$\begin{aligned}
\sum_{i=1}^{K} \sum_{\ell=1}^{M} \sum_{t=1}^{T} & \left(-\frac{1}{2} \log |\Sigma_{i\ell}| - \frac{1}{2} (x_t - \mu_{i\ell})^T \Sigma_{i\ell}^{-1} (x_t - \mu_{i\ell}) \right) p(Z_t = i, M_t = \ell | x^T, \theta') \\
&= \sum_{i=1}^{K} \sum_{\ell=1}^{M} \Bigg(\frac{1}{2} \log |\Sigma_{i\ell}^{-1}| \sum_{t=1}^{T} p(Z_t = i, M_t = \ell | x^T, \theta') \\
&\quad - \frac{1}{2} \sum_{t=1}^{T} p(Z_t = i, M_t = \ell | x^T, \theta') \operatorname{tr} \left(\Sigma_{i\ell}^{-1} N_{t,i\ell} \right) \Bigg),
\end{aligned}$$

kus $N_{t,i\ell} = (x_t - \mu_{i\ell})(x_t - \mu_{i\ell})^T$. Võttes $\Sigma_{i\ell}^{-1}$ järgi tuletise, saame:

$$\begin{aligned}
&\frac{1}{2} \sum_{t=1}^{T} p(Z_t = i, M_t = \ell | x^T, \theta') (2\Sigma_{i\ell} - \operatorname{diag}(\Sigma_{i\ell})) \\
&- \frac{1}{2} \sum_{t=1}^{T} p(Z_t = i, M_t = \ell | x^T, \theta') (2N_{t,i\ell} - \operatorname{diag}(N_{t,i\ell})) = 2S_{i\ell} - \operatorname{diag}(S_{i\ell}),
\end{aligned}$$

kus $S_{i\ell} = \frac{1}{2}\sum_{t=1}^{T} p(Z_t = i, M_t = \ell | x^T, \theta')(\Sigma_{i\ell} - N_{t,i\ell})$. Võrdsustades nulliga saame, et $2S_{i\ell} - \mathrm{diag}(S_{i\ell}) = 0$ ehk järelikult $S_{i\ell} = 0$. Seega

$$\sum_{t=1}^{T} p(Z_t = i, M_t = \ell | x^T, \theta')(\Sigma_{i\ell} - (x_t - \mu_{i\ell})(x_t - \mu_{i\ell})^T) = 0,$$

kust saame avaldada $\Sigma_{i\ell}$:

$$\Sigma_{i\ell} = \frac{\sum_{t=1}^{T}(x_t - \mu_{i\ell})(x_t - \mu_{i\ell})^T p(Z_t = i, M_t = \ell | x^T, \theta')}{\sum_{t=1}^{T} p(Z_t = i, M_t = \ell | x^T, \theta')}.$$

Paneme tähele, et $\Sigma_{i\ell}$ hindamiseks tuleb eelnevalt hinnata $\mu_{i\ell}$.

1.3 Edasi- ja tagasi-tõenäosused

Eelmises alapeatükis leitud parameetrite hinnangute arvutamiseks kasutatakse praktikas suuruseid, mida kutsutakse edasi- ja tagasi-tõenäosusteks. Kõikide parameetrite hinnangud on avaldatavad nende kaudu. Alapeatüki kirjutamisel on tuginetud allikale Bishop (2006), kust muuhulgas on rekursiivsete valemite tuletamisel kasutatud konkreetseid tingliku sõltumatuse omadusi leheküljelt 619. Kasutatud omaduste numbrid on toodud vastavate võrdusmärkide peal.

Vaatleme klassikalist varjatud Markovi mudelit (X^T, Z^T) parameetritega $\theta = (\pi, \mathrm{P}, \theta_1, \ldots, \theta_K)$, kus $\theta_1, \ldots, \theta_K$ on emissioonijaotuste $f_1, \ldots, f_K$ parameetrid. Edasi-tõenäosused $\alpha_t(i)$ näitavad, kui suur on tõepära, et emiteeritakse osaline vaatluste jada kuni hetkeni t ning et samal hetkel on varjatud ahel seisundis i:

$$\alpha_t(i) := p(x_1, \ldots, x_t, Z_t = i | \theta), \quad i \in S, \quad t \in \{1, \ldots, T\}.$$

Parameetri θ jätame järgnevates arvutustes märkimata, kuna see on fikseeritud. Hetkel $t = 1$ saame seega

$$\alpha_1(i) = p(x_1, Z_1 = i) = p(Z_1 = i)p(x_1|Z_1 = i) = \pi_i f_i(x_1).$$

Muutuja $\alpha_t(i)$ järgnevad väärtused saame leida rekursiivselt:

$$\begin{aligned}
\alpha_{t+1}(j) = p(x_1, \ldots, x_{t+1}, Z_{t+1} = j) &= \sum_{i=1}^{K} p(x_1, \ldots, x_{t+1}, Z_t = i, Z_{t+1} = j) \\
&= \sum_{i=1}^{K} p(x_1, \ldots, x_t, Z_t = i)\, p(x_{t+1}, Z_{t+1} = j|x_1, \ldots, x_t, Z_t = i) \\
&= \sum_{i=1}^{K} \alpha_t(i)\, p(Z_{t+1} = j|x_1, \ldots, x_t, Z_t = i) \\
&\quad \cdot p(x_{t+1}|x_1, \ldots, x_t, Z_t = i, Z_{t+1} = j) \\
&\overset{\substack{(13.30)\\(13.31)}}{=} \sum_{i=1}^{K} \alpha_t(i)\, p(Z_{t+1} = j|Z_t = i)\, p(x_{t+1}|Z_{t+1} = j) \\
&= \left(\sum_{i=1}^{K} \alpha_t(i)\, p_{ij} \right) f_j(x_{t+1}), \quad t = 1, \ldots, T-1.
\end{aligned}$$

Analoogiliselt defineeritakse tagasi-tõenäosused

$$\beta_t(i) := p(x_{t+1}, \ldots, x_T|Z_t = i, \theta), \quad i \in S, \quad t \in \{1, \ldots, T\},$$

mis näitavad, kui suure tõepäraga emiteeritakse osaline vaatluste jada alates hetkest $t+1$ kuni hetkeni T, kui hetkel t on varjatud ahel seisundis i. Hetkel T defineeritakse $\beta_T(i) = 1$. Lihtsustamiseks jätame edasises jällegi θ avaldistesse märkimata. Ka tagasitõenäosused saab arvutada rekursiivselt:

$$\beta_t(i) = p(x_{t+1}, \ldots, x_T|Z_t = i) = \frac{p(x_{t+1}, \ldots, x_T, Z_t = i)}{p(Z_t = i)}$$

$$
\begin{aligned}
&= \sum_{j=1}^{K} \frac{p(x_{t+1},\dots,x_T, Z_t = i, Z_{t+1} = j)}{p(Z_t = i)} \\
&= \sum_{j=1}^{K} p(x_{t+1},\dots,x_T | Z_t = i, Z_{t+1} = j) \cdot \frac{p(Z_t = i, Z_{t+1} = j)}{p(Z_t = i)} \\
&\overset{(13.27)}{=} \sum_{j=1}^{K} p(x_{t+1},\dots,x_T | Z_{t+1} = j)\, p(Z_{t+1} = j | Z_t = i) \\
&= \sum_{j=1}^{K} p(x_{t+1},\dots,x_T | Z_{t+1} = j)\, p_{ij} \\
&\overset{(13.24)}{=} \sum_{j=1}^{K} p(x_{t+1} | Z_{t+1} = j)\, p(x_{t+2},\dots,x_T | Z_{t+1} = j)\, p_{ij} \\
&= \sum_{j=1}^{K} f_j(x_{t+1})\, \beta_{t+1}(j)\, p_{ij}, \quad t = T-1,\dots,1.
\end{aligned}
$$

Paneme tähele, et ajahetkele t ja seisundile i vastava edasi- ja tagasi-tõenäosuse omavaheline korrutis annab meile tõepära

$$
p(x^T, Z_t = i) = \alpha_t(i)\beta_t(i), \quad i \in S, \quad t \in \{1,\dots,T\}. \tag{1}
$$

Näeme, et edasi- ja tagasi-tõenäosuste rekursiivsel arvutamisel on meil jada pikkuse kasvamisel tegemist väga väikeste numbriliste suurustega. See toob meid probleemini, kus suure T korral lähenevad $\alpha_t(i)$ ja $\beta_t(i)$ t kasvades nullile. Seepärast kasutatakse praktikas arvutamiseks skaleeritud suuruseid $\hat{\alpha}_t(i)$ ja $\hat{\beta}_t(i)$:

$$
\begin{aligned}
\hat{\alpha}_t(i) = p(Z_t = i | x_1,\dots,x_t) = \frac{p(x_1,\dots,x_t, Z_t = i)}{p(x_1,\dots,x_t)} &= \frac{\alpha_t(i)}{\prod_{n=1}^{t} c_n}, \\
\hat{\beta}_t(i) = \frac{p(x_{t+1},\dots,x_T | Z_t = i)}{p(x_{t+1},\dots,x_T | x_1,\dots,x_t)} &= \frac{\beta_t(i)}{\prod_{n=t+1}^{T} c_n},
\end{aligned}
$$

kus $c_1 = p(x_1)$ ning $c_n = p(x_n|x_1, \dots, x_{n-1})$, $n = 2, \dots, T$, nimetatakse skaleerimiskonstantideks. Näeme, et kehtib

$$p(x^T) = p(x_1, \dots, x_t)\, p(x_{t+1}, \dots, x_T|x_1, \dots, x_t) = \prod_{n=1}^{T} c_n. \tag{2}$$

Skaleeritud edasi-tõenäosuste korral kehtib järgnev rekursiivne seos:

$$c_{t+1}\hat{\alpha}_{t+1}(j) = c_{t+1}\frac{\alpha_{t+1}(j)}{\prod_{n=1}^{t+1} c_n} = \frac{\sum_{i=1}^{K} \alpha_t(i) p_{ij} f_j(x_{t+1})}{\prod_{n=1}^{t} c_n} = \sum_{i=1}^{K} \hat{\alpha}_t(i) p_{ij} f_j(x_{t+1}).$$

Kuna $\sum_{j=1}^{K} \hat{\alpha}_{t+1}(j) = \sum_{j=1}^{K} p(Z_{t+1} = j|x_1, \dots, x_{t+1}) = 1$, siis eelneva avaldise põhjal saame skaleerimiskonstandid arvutada kujul:

$$c_{t+1} = \sum_{j=1}^{K} c_{t+1}\hat{\alpha}_{t+1}(j) = \sum_{j=1}^{K} \left(\sum_{i=1}^{K} \hat{\alpha}_t(i) p_{ij} \right) f_j(x_{t+1}).$$

Skaleeritud tagasi-tõenäosuste puhul kehtib analoogselt rekursiivne seos:

$$c_{t+1}\hat{\beta}_t(i) = \sum_{j=1}^{K} f_j(x_{t+1})\hat{\beta}_{t+1}(j) p_{ij},$$

kus saame kasutada juba eelnevalt edasi-tõenäosuste kaudu välja arvutatud skaleerimiskonstante.

Parameetrite hinnangute arvutusvalemite lihtsustamiseks defineerime veel muutujad $\gamma_t(i)$ ja $\xi_t(i,j)$, mida nimetatakse vastavalt silumistõenäosusteks ning paariviisilisteks silumistõenäosusteks:

$$\gamma_t(i) := P(Z_t = i|x^T), \quad i \in S,\ t \in \{1, \dots, T\},$$
$$\xi_t(i,j) := P(Z_t = i, Z_{t+1} = j|x^T), \quad i, j \in S,\ t \in \{1, \dots, T-1\}.$$

Eelnevalt toodud tulemuste (1) ja (2) põhjal näeme, et

$$\gamma_t(i) = p(Z_t = i|x^T) = \frac{p(Z_t = i, x^T)}{p(x^T)} = \frac{\alpha_t(i)\beta_t(i)}{\prod_{n=1}^{T} c_n} = \hat{\alpha}_t(i)\hat{\beta}_t(i).$$

Paariviisilised silumistõenäosused avalduvad aga kujul:

$$\begin{aligned}
\xi_t(i,j) = p(Z_t = i, Z_{t+1} = j|x^T) &= \frac{p(Z_t = i, Z_{t+1} = j, x^T)}{p(x^T)} \\
&= \frac{\alpha_t(i)p_{ij}f_j(x_{t+1})\beta_{t+1}(j)}{\prod_{n=1}^{T} c_n} = \frac{\alpha_t(i)p_{ij}f_j(x_{t+1})\beta_{t+1}(j)}{\left(\prod_{n=1}^{t} c_n\right) c_{t+1} \prod_{n=t+2}^{T} c_n} \\
&= \frac{\hat{\alpha}_t(i)p_{ij}f_j(x_{t+1})\hat{\beta}_{t+1}(j)}{c_{t+1}}.
\end{aligned}$$

Kasutades edasi- ja tagasi-tõenäosuste põhjal arvutatavaid silumistõenäosuseid, saame nüüd kõikide HMM parameetrite hinnangute arvutusvalemid avaldada nende kaudu. Näiteks alg- ning üleminekutõenäosuste hinnangud saame arvutada valemitega:

$$\begin{aligned}
\hat{\pi}_i &= p(Z_1 = i|x^T) = \gamma_1(i), \quad i \in S, \\
\hat{p}_{ij} &= \frac{\sum_{t=1}^{T-1} p(Z_t = i, Z_{t+1} = j|x^T)}{\sum_{t=1}^{T-1} p(Z_t = i|x^T)} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}, \quad i,j \in S.
\end{aligned}$$

2 Tingliku sõltumatuse mudel

Olgu meil antud MRT mõõtmiste jada $X^T = (X_1, \ldots, X_T)$ väärtustega $x^T = (x_1, \ldots, x_T)$ ja vastavate KT mõõtmiste jada $Y^T = (Y_1, \ldots, Y_T)$ väärtustega $y^T = (y_1, \ldots, y_T)$. Teame, et mõlema väärtused sõltuvad uuritava piirkonna kudede klassidest, mida käsitleme varjatud tunnustena $Z^T = (Z_1, \ldots, Z_T)$. Eeldame, et vaadeldavate kudede klasse on kokku K ehk varjatud tunnuste seisundite hulk on $S_z := \{1, \ldots, K\}$. Seega, vaatleme alustuseks varjatud Markovi mudelit

$$(X^T, Y^T, Z^T) = \{(X_1, Y_1, Z_1), (X_2, Y_2, Z_2), \ldots, (X_T, Y_T, Z_T)\},$$

kus alg- ning üleminekutõenäosusteks on vastavalt π_i ja p_{ij}, $i, j \in S_z$. Emissioonijaotuseid varjatud seisundi $Z_t = i$ korral tähistame $f_i(x|\theta_{i,f})$ ja $g_i(y|\theta_{i,g})$. Varjatud Markovi mudeli eelduste kohaselt on antud $Z^T = z^T$ korral vaatlused $(X_1, Y_1), \ldots, (X_T, Y_T)$ sõltumatud. Lisaks eeldame, et vaatlused X_t ja Y_t on antud $Z_t = z_t$ korral tinglikult sõltumatud. Seega, iga t korral sõltuvad kahe jada vaatlused teineteisest vaid latentse tunnuse Z_t kaudu ehk Y_t tinglik jaotus antud x^T korral avaldub:

$$\begin{aligned}
P(Y_t \in A | X^T = x^T) &= \sum_{i=1}^{K} P(Y_t \in A | Z_t = i, x^T) P(Z_t = i | x^T) \\
&= \sum_{i=1}^{K} P(Y_t \in A | Z_t = i) P(Z_t = i | x^T),
\end{aligned}$$

kus A on mingi hulk Y_t väärtuste piirkonnas.

Kirjeldatud omadusest tulenevalt nimetame seda varjatud Markovi mudeli erijuhtu tingliku sõltumatuse mudeliks. Eesmärk on Y^T prognoosimine X^T

kaudu, mida tänu eeltoodule saame teha tingliku keskväärtuse abil:

$$E(Y_t|X^T = x^T) = \sum_{i=1}^{K} E(Y_t|Z_t = i)P(Z_t = i|x^T), \quad t = 1\ldots,T.$$

Ülaltoodud mudel eeldab, et varjatud ahela seisundid ehk kudede klassid ja nende arv on mõlema vaatluste jada jaoks samad. Kuna MRT ja KT jäädvustavad eri kudesid väga erinevalt, kusjuures mõni kude võib olla teistest väga raskesti eristatav (näiteks luu ja õhk MRT korral), siis tahame lubada olukorda, kus KT ahela varjatud seisundite arv võiks MRT ahela omast erineda. Üks võimalus selleks on modelleerida KT jaotust seisundi $Z_t = i$ korral kui M komponendiga normaaljaotuste segu:

$$p(y|Z_t = i) = \sum_{\ell=1}^{M} c_{i\ell} g_\ell(y|\theta_{\ell,g}) = \sum_{\ell=1}^{M} c_{i\ell}\phi(y|\mu_{\ell,g}, \sigma^2_{\ell,g}), \quad i \in S_z,$$

kus g_ℓ on nüüd segujaotuse komponendile ℓ ehk varjatud seisundile $R_t = \ell$ vastav KT vaatluse tihedus. Paneme tähele, et emissiooniparameetrid $\mu_{\ell,g}$ ja $\sigma^2_{\ell,g}$ ei sõltu seisundist $Z_t = i$, kaalud $c_{i\ell}$ aga küll. Lihtsustuse mõttes oleme lisaks eeldanud, et KT komponentide arv on iga MRT seisundi $Z_t = i$ korral sama.

Saame nüüd vaadelda kahekordse seisundite jadaga $W^T = (Z^T, R^T)$ varjatud Markovi mudelit:

$$\begin{aligned}(X^T, Y^T, W^T) &= (X^T, Y^T, Z^T, R^T) \\ &= \{(X_1, Y_1, Z_1, R_1), (X_2, Y_2, Z_2, R_2), \ldots, (X_T, Y_T, Z_T, R_T)\},\end{aligned}$$

kus Z_t määrab MRT vaatluse seisundi ja R_t KT segujaotuse komponendi ehk ka KT vaatluse seisundi. Seega, konkreetse seisundi $W_t = (z, r)$ korral kehtib

$z \in S_z = \{1, \ldots, K\}$ ja $r \in S_r := \{1, \ldots, M\}$. Seisundile (z, r) vastavaid vaatluste X_t ja Y_t emissioonijaotuseid tähistame $f_{(z,r)}(x|\theta_{(z,r),f})$ ja $g_{(z,r)}(y|\theta_{(z,r),g})$, mille korral kehtib

$$f_{(z,r)}(x|\theta_{(z,r),f}) = f_z(x|\theta_{z,f}), \quad g_{(z,r)}(y|\theta_{(z,r),g}) = g_r(y|\theta_{r,g}),$$

kus f_z ei sõltu KT seisundist r ega g_r MRT seisundist z. Kuna MRT vaatlused võivad olla mitmemõõtmelised (meil võivad olla mõõtmised mitme MRT-pildi kohta), siis nende jaotuseid f_i, $i \in S_z$, modelleerime kui mitmemõõtmelisi normaaljaotuseid.

Eeldame, et iga $t + 1$ korral on Z_{t+1} eelmisele ajahetkele vastavast KT seisundist R_t sõltumatu. Lisaks eeldame, et iga $t + 1$ korral sõltub R_{t+1} vaid sama ajahetke MRT vaatluse seisundist Z_{t+1} ja on sõltumatu eelmisele ajahetkele vastavatest seisunditest Z_t ning R_t. Selliselt avalduvad uuele mudelile vastavad üleminekutõenäosused kujul

$$\begin{gathered}
p_{w_t w_{t+1}} = P(W_{t+1} = (z_{t+1}, r_{t+1})|W_t = (z_t, r_t)) \\
= P(Z_{t+1} = z_{t+1}, R_{t+1} = r_{t+1}|Z_t = z_t, R_t = r_t) \\
= P(Z_{t+1} = z_{t+1}|Z_t = z_t, R_t = r_t)P(R_{t+1} = r_{t+1}|Z_{t+1} = z_{t+1}, Z_t = z_t, R_t = r_t) \\
= P(Z_{t+1} = z_{t+1}|Z_t = z_t)P(R_{t+1} = r_{t+1}|Z_{t+1} = z_{t+1}) = p_{z_t z_{t+1}} c_{z_{t+1} r_{t+1}}
\end{gathered}$$

ning algtõenäosused on kujul

$$\begin{gathered}
\pi_{w_1} = P(W_1 = w_1) = P(Z_1 = z_1, R_1 = r_1) \\
= P(Z_1 = z_1)P(R_1 = r_1|Z_1 = z_1) = \pi_{z_1} c_{z_1 r_1}.
\end{gathered}$$

2.1 Tingliku sõltumatuse mudeli parameetrite hinnangute leidmine

Tähistagu θ kõiki tingliku sõltumatuse mudeli parameetreid, mida soovime hinnata ehk alg- ja üleminekutõenäosuseid, kaale ning emissioonijaotuste parameetreid. Tuletame EM-algoritmi abil parameetrite hinnangud, kuid tähistame neid käesolevas alapeatükis samamoodi kui funktsiooni argumente. Analoogsed tuletuskäigud on põhjalikumalt läbi tehtud alapeatükis 1.2.

Tingliku sõltumatuse mudelile vastav Q-funktsioon on kujul

$$Q(\theta, \theta') = \sum_{\mathcal{W}} \log p(x^T, y^T, w^T|\theta) p(w^T|x^T, y^T, \theta'),$$

kus θ' tähistab eelmisel sammul leitud parameetrite hinnanguid ning $\mathcal{W}$ on kõigi pikkusega T varjatud olekute jadade hulk.

Parameetrite hinnangute leidmine ühe jada korral

Vaatleme esmalt olukorda, kus parameetrite hinnangud leitakse ühe vaatluste jada põhjal. Konkreetse seisundite jada $w^T = (w_1, \dots, w_T)$ korral avaldub Q-funktsioonis vaadeldav tõepära kujul:

$$\begin{aligned}
p(x^T, y^T, w^T|\theta) &= p(w^T|\theta) p(x^T, y^T|w^T, \theta) \\
&= p(w^T|\theta) p(x^T|w^T, \theta) p(y^T|w^T, \theta) \\
&= \pi_{w_1} \prod_{t=1}^{T-1} p_{w_t w_{t+1}} \prod_{t=1}^{T} f_{w_t}(x_t|\theta) g_{w_t}(y_t|\theta) \\
&= \pi_{z_1} c_{z_1 r_1} \prod_{t=1}^{T-1} p_{z_t z_{t+1}} c_{z_{t+1} r_{t+1}} \prod_{t=1}^{T} f_{z_t}(x_t|\theta_{z_t,f}) g_{r_t}(y_t|\theta_{r_t,g}).
\end{aligned}$$

Saame seega Q-funktsiooni avaldada viie liidetava summana ning seejärel neid eraldi Lagrange'i kordajate meetodil maksimiseerida:

$$\begin{aligned}Q(\theta,\theta') &= \sum_{\mathcal{W}} \log \pi_{z_1} p(w^T|x^T, y^T, \theta') + \sum_{\mathcal{W}} \sum_{t=1}^{T-1} \log p_{z_t z_{t+1}} p(w^T|x^T, y^T, \theta') \\ &+ \sum_{\mathcal{W}} \sum_{t=1}^{T} \log c_{z_t r_t} p(w^T|x^T, y^T, \theta') + \sum_{\mathcal{W}} \sum_{t=1}^{T} \log f_{z_t}(x_t|\theta_{z_t,f}) p(w^T|x^T, y^T, \theta') \\ &+ \sum_{\mathcal{W}} \sum_{t=1}^{T} \log g_{r_t}(y_t|\theta_{r_t,g}) p(w^T|x^T, y^T, \theta').\end{aligned}$$

Esimene liidetav lihtsustub järgmiselt:

$$\begin{aligned}\sum_{\mathcal{W}} \log \pi_{z_1} p(w^T|x^T, y^T, \theta') &= \sum_{z_1=1}^{K} \cdots \sum_{z_T=1}^{K} \sum_{r_1=1}^{M} \cdots \sum_{r_T=1}^{M} \log \pi_{z_1} p(z^T, r^T|x^T, y^T, \theta') \\ &= \sum_{z_1=1}^{K} \log \pi_{z_1} p(z_1|x^T, y^T, \theta') = \sum_{i=1}^{K} \log \pi_i p(Z_1 = i|x^T, y^T, \theta'),\end{aligned}$$

kust Lagrange'i kordajate meetodil kitsendusega $\sum_{i=1}^{K} \pi_i = 1$ saame

$$\pi_i = p(Z_1 = i|x^T, y^T, \theta').$$

Et leida p_{ij} hinnang, maksimeerime teist liidetavat

$$\sum_{\mathcal{W}} \sum_{t=1}^{T-1} \log p_{z_t z_{t+1}} p(w^T|x^T, y^T, \theta') = \sum_{t=1}^{T-1} \sum_{i=1}^{K} \sum_{j=1}^{K} \log p_{ij} p(Z_t = i, Z_{t+1} = j|x^T, y^T, \theta').$$

Võttes arvesse kitsendust $\sum_{j=1}^{K} p_{ij} = 1$, saame, et üleminekutõenäosused p_{ij} avalduvad kujul:

$$p_{ij} = \frac{\sum_{t=1}^{T-1} p(Z_t = i, Z_{t+1} = j|x^T, y^T, \theta')}{\sum_{t=1}^{T-1} p(Z_t = i|x^T, y^T, \theta')}.$$

Kaalude hinnangu valemi leidmiseks vaatleme kolmandat liidetavat

$$\sum_{\mathcal{W}} \sum_{t=1}^{T} \log c_{z_t r_t} p(w^T | x^T, y^T, \theta') = \sum_{t=1}^{T} \sum_{i=1}^{K} \sum_{\ell=1}^{M} \log c_{i\ell} p(Z_t = i, R_t = \ell | x^T, y^T, \theta'),$$

millest kitsendust $\sum_{\ell=1}^{M} c_{i\ell} = 1$ arvestades saame

$$c_{i\ell} = \frac{\sum_{t=1}^{T} p(Z_t = i, R_t = \ell | x^T, y^T, \theta')}{\sum_{t=1}^{T} p(Z_t = i | x^T, y^T, \theta')}.$$

MRT emissioonijaotuste hindamiseks vaatame d-mõõtmelisi normaaljaotuseid f_i parameetritega $\theta_{i,f} = (\mu_{i,f}, \Sigma_{i,f})$:

$$f_i(x|\theta_{i,f}) = \frac{1}{(2\pi)^{d/2} |\Sigma_{i,f}|^{\frac{1}{2}}} \cdot \exp\Big(-\frac{1}{2}(x - \mu_{i,f})^T \Sigma_{i,f}^{-1} (x - \mu_{i,f}) \Big), \quad i \in S_z.$$

Parameetrite hinnangu saamiseks piisab, kui leiame Q-funktsiooni neljanda liidetava tuletise vastava parameetri järgi ning võrdsustame selle nulliga ja seejärel avaldame parameetri. Seega keskväärtuste $\mu_{i,f}$ ja kovariatsioonimaatriksite $\Sigma_{i,f}$ hinnangute leidmiseks vaatame avaldist:

$$\begin{aligned} \sum_{\mathcal{W}} \sum_{t=1}^{T} \log f_{z_t}(x_t | \theta_{z_t,f}) p(w^T | x^T, y^T, \theta') &= \sum_{t=1}^{T} \sum_{i=1}^{K} \log f_i(x_t | \theta_{i,f}) p(Z_t = i | x^T, y^T, \theta') \\ &= \sum_{t=1}^{T} \sum_{i=1}^{K} \Big(-\frac{d}{2} \log(2\pi) - \frac{1}{2} \log(|\Sigma_{i,f}|) - \frac{1}{2}(x_t - \mu_{i,f})^T \Sigma_{i,f}^{-1} (x_t - \mu_{i,f}) \Big) \\ &\qquad \cdot p(Z_t = i | x^T, y^T, \theta'). \end{aligned}$$

Võttes sellest $\mu_{i,f}$ järgi tuletise ja võrdsustades nulliga, saame, et keskväärtused avalduvad järgmiselt:

$$\mu_{i,f} = \frac{\sum_{t=1}^{T} x_t p(Z_t = i | x^T, y^T, \theta')}{\sum_{t=1}^{T} p(Z_t = i | x^T, y^T, \theta')}.$$

Analoogselt $\Sigma_{i,f}$ järgi tuletist võttes saame

$$\Sigma_{i,f} = \frac{\sum_{t=1}^{T}(x_t - \mu_{i,f})(x_t - \mu_{i,f})^T p(Z_t = i|x^T, y^T, \theta')}{\sum_{t=1}^{T} p(Z_t = i|x^T, y^T, \theta')}.$$

KT emissioonijaotuste hindamiseks tuleb leida parameetrite $\theta_{\ell,g} = (\mu_{\ell,g}, \sigma_{\ell,g}^2)$ hinnangud M erineva segujaotuse komponendi jaoks, milleks on ühemõõtmelised normaaljaotused g_ℓ:

$$g_\ell(y|\theta_{\ell,g}) = \frac{1}{\sigma_{\ell,g}\sqrt{2\pi}} \cdot \exp\left(- \frac{(y - \mu_{\ell,g})^2}{2\sigma_{\ell,g}^2}\right), \quad \ell \in S_r.$$

Eelnevaga analoogse tuletuskäiguga Q-funktsiooni viienda liidetava korral saame parameetrite avaldised kujul:

$$\mu_{\ell,g} = \frac{\sum_{t=1}^{T} y_t p(R_t = \ell|x^T, y^T, \theta')}{\sum_{t=1}^{T} p(R_t = \ell|x^T, y^T, \theta')},$$
$$\sigma_{\ell,g}^2 = \frac{\sum_{t=1}^{T}(y_t - \mu_{\ell,g})^2 p(R_t = \ell|x^T, y^T, \theta')}{\sum_{t=1}^{T} p(R_t = \ell|x^T, y^T, \theta')}.$$

Parameetrite hinnangute leidmine N sõltumatu jada korral

Siiani uurisime juhtu, kus hindame parameetreid vaid ühe vaatluste jada põhjal. Meid huvitava rakenduse korral pea andmetele tekib aga palju sõltumatuid jadasid, seega vaatleme nüüd juhtu, kus meil on N erinevat sõltumatut vaatluste jada pikkustega $T_1, \ldots, T_N$ ehk $(X^{T_1}, Y^{T_1}), \ldots, (X^{T_N}, Y^{T_N})$, kusjuures $\sum_{n=1}^{N} T_n = T$. Ühe konkreetse vaatluste jada väärtuseid tähistame vastavalt $x^{T_n} = (x_1^n, \ldots, x_{T_n}^n)$ ja $y^{T_n} = (y_1^n, \ldots, y_{T_n}^n)$, $n = 1, \ldots, N$. Analoogselt kehtib $W^{T_n} = (Z^{T_n}, R^{T_n})$ korral $Z^{T_n} = (Z_1^n, \ldots, Z_{T_n}^n)$ ja $R^{T_n} = (R_1^n, \ldots, R_{T_n}^n)$, $n = 1, \ldots, N$.

Seega, $p(x^T, y^T, w^T|\theta) = \prod_{n=1}^N p(x^{T_n}, y^{T_n}, w^{T_n}|\theta)$ ja Q-funktsioonis on nüüd

$$\log p(x^T, y^T, w^T|\theta) = \sum_{n=1}^N \log p(x^{T_n}, y^{T_n}, w^{T_n}|\theta),$$

kus seisundite jada $w^{T_n} = (w_1^n, w_2^n, \ldots, w_{T_n}^n)$ korral

$$\begin{aligned}
p(x^{T_n}, y^{T_n}, w^{T_n}|\theta) &= p(w^{T_n}|\theta)p(x^{T_n}|w^{T_n},\theta)p(y^{T_n}|w^{T_n},\theta) \\
&= \pi_{w_1^n} \prod_{t=1}^{T_n-1} p_{w_t^n w_{t+1}^n} \prod_{t=1}^{T_n} f_{w_t^n}(x_t^n|\theta) g_{w_t^n}(y_t^n|\theta) \\
&= \pi_{z_1^n} c_{z_1^n r_1^n} \prod_{t=1}^{T_n-1} p_{z_t^n z_{t+1}^n} c_{z_{t+1}^n r_{t+1}^n} \prod_{t=1}^{T_n} f_{z_t^n}(x_t^n|\theta_{z_t^n,f}) g_{r_t^n}(y_t^n|\theta_{r_t^n,g}).
\end{aligned}$$

Näeme, et edasine tuletuskäik on analoogne juhuga, kus vaatlesime vaid ühte jada. Saame Q-funktsiooni esitada jällegi viie liidetava summana ning seejärel neid eraldi maksimiseerida:

$$\begin{aligned}
Q(\theta, \theta') = &\sum_{n=1}^N \sum_{\mathcal{W}^n} \log \pi_{z_1^n} p(w^{T_n}|x^{T_n}, y^{T_n}, \theta') \\
&+ \sum_{n=1}^N \sum_{\mathcal{W}^n} \sum_{t=1}^{T_n-1} \log p_{z_t^n z_{t+1}^n} p(w^{T_n}|x^{T_n}, y^{T_n}, \theta') \\
&+ \sum_{n=1}^N \sum_{\mathcal{W}^n} \sum_{t=1}^{T_n} \log c_{z_t^n r_t^n} p(w^{T_n}|x^{T_n}, y^{T_n}, \theta') \\
&+ \sum_{n=1}^N \sum_{\mathcal{W}^n} \sum_{t=1}^{T_n} \log f_{z_t^n}(x_t^n|\theta_{z_t^n,f}) p(w^{T_n}|x^{T_n}, y^{T_n}, \theta') \\
&+ \sum_{n=1}^N \sum_{\mathcal{W}^n} \sum_{t=1}^{T_n} \log g_{r_t^n}(y_t^n|\theta_{r_t^n,g}) p(w^{T_n}|x^{T_n}, y^{T_n}, \theta'),
\end{aligned}$$

kus $\mathcal{W}^n$ on kõigi pikkusega T_n varjatud olekute jadade hulk. Selliselt tuletades avaduvad N sõltumatu jada põhjal hinnatud parameetrid kujul:

$$\begin{aligned}
\pi_i &= \frac{1}{N}\sum_{n=1}^{N} p(Z_1^n = i|x^{T_n}, y^{T_n}, \theta'),\\
p_{ij} &= \frac{\sum_{n=1}^{N}\sum_{t=1}^{T_n-1} p(Z_t^n = i, Z_{t+1}^n = j|x^{T_n}, y^{T_n}, \theta')}{\sum_{n=1}^{N}\sum_{t=1}^{T_n-1} p(Z_t^n = i|x^{T_n}, y^{T_n}, \theta')},\\
c_{i\ell} &= \frac{\sum_{n=1}^{N}\sum_{t=1}^{T_n} p(Z_t^n = i, R_t^n = \ell|x^{T_n}, y^{T_n}, \theta')}{\sum_{n=1}^{N}\sum_{t=1}^{T_n} p(Z_t^n = i|x^{T_n}, y^{T_n}, \theta')},\\
\mu_{i,f} &= \frac{\sum_{n=1}^{N}\sum_{t=1}^{T_n} x_t^n p(Z_t^n = i|x^{T_n}, y^{T_n}, \theta')}{\sum_{n=1}^{N}\sum_{t=1}^{T_n} p(Z_t^n = i|x^{T_n}, y^{T_n}, \theta')},\\
\Sigma_{i,f} &= \frac{\sum_{n=1}^{N}\sum_{t=1}^{T_n} (x_t^n - \mu_{i,f})(x_t^n - \mu_{i,f})^T p(Z_t^n = i|x^{T_n}, y^{T_n}, \theta')}{\sum_{n=1}^{N}\sum_{t=1}^{T_n} p(Z_t^n = i|x^{T_n}, y^{T_n}, \theta')},\\
\mu_{\ell,g} &= \frac{\sum_{n=1}^{N}\sum_{t=1}^{T_n} y_t^n p(R_t^n = \ell|x^{T_n}, y^{T_n}, \theta')}{\sum_{n=1}^{N}\sum_{t=1}^{T_n} p(R_t^n = \ell|x^{T_n}, y^{T_n}, \theta')},\\
\sigma_{\ell,g}^2 &= \frac{\sum_{n=1}^{N}\sum_{t=1}^{T_n} (y_t^n - \mu_{\ell,g})^2 p(R_t^n = \ell|x^{T_n}, y^{T_n}, \theta')}{\sum_{n=1}^{N}\sum_{t=1}^{T_n} p(R_t^n = \ell|x^{T_n}, y^{T_n}, \theta')}.
\end{aligned}$$

2.2 Parameetrite ümberhindamise valemid EM-algoritmis

Tingliku sõltumatuse mudeli parameetrite hinnangute arvutamiseks saame sarnaselt tavalise HMM-ga kasutada edasi- ja tagasi-tõenäosuseid. Alapeatükis 1.3 defineeritud suurused avalduvad antud juhul samal kujul – selle erinevusega, et ühekordse vaatluste jada X^T asemel on meil kahekordne vaatluste jada (X^T, Y^T) emissioonijaotustega

$$p(x, y|Z_t = i, \theta) = f_i(x|\theta_{i,f}) \sum_{\ell=1}^{M} c_{i\ell} g_\ell(y|\theta_{\ell,g}), \quad i \in S_z.$$

Seega silumistõenäosused avalduvad tingliku sõltumatuse mudeli korral kujul:

$$\gamma_t(i) = p(Z_t = i|x^T, y^T, \theta') = \frac{p(x^T, y^T, Z_t = i|\theta')}{p(x^T, y^T|\theta')} = \frac{\alpha_t(i)\beta_t(i)}{p(x^T, y^T|\theta')},$$

$$\begin{aligned}\xi_t(i,j) = p(Z_t = i, Z_{t+1} = j|x^T, y^T, \theta') &= \frac{p(Z_t = i, Z_{t+1} = j, x^T, y^T|\theta')}{p(x^T, y^T|\theta')} \\ &= \frac{\alpha_t(i)p'_{ij}f_j(x_{t+1}|\theta'_{j,f})\sum_{\ell=1}^{M} c_{j\ell}g_\ell(y_{t+1}|\theta'_{\ell,g})\beta_{t+1}(j)}{p(x^T, y^T|\theta')}.\end{aligned}$$

Meenutame, et vaatluste tõepära saab avaldada skaleerimiskonstantide kaudu ning praktikas kasutataksegi silumistõenäosuste arvutamiseks skaleeritud suuruseid. Leiame ka suurused $\gamma_t(i,\ell)$:

$$\begin{aligned}\gamma_t(i,\ell) &= p(Z_t = i, R_t = \ell|x^T, y^T, \theta') \\ &= p(Z_t = i|x^T, y^T, \theta')p(R_t = \ell|Z_t = i, x^T, y^T, \theta') \\ &= \gamma_t(i)p(R_t = \ell|Z_t = i, y_t, \theta') \\ &= \gamma_t(i)\frac{p(R_t = \ell|Z_t = i, \theta')p(y_t|R_t = \ell, Z_t = i, \theta')}{p(y_t|Z_t = i, \theta')} \\ &= \gamma_t(i)\frac{c_{i\ell}p(y_t|R_t = \ell, \theta')}{p(y_t|Z_t = i, \theta')} = \gamma_t(i)\frac{c_{i\ell}g_\ell(y_t|\theta')}{\sum_{\ell=1}^{M} c_{i\ell}g_\ell(y_t|\theta')}.\end{aligned}$$

Kuna vaatleme juhtu, kus parameetrite hinnangud arvutatakse N sõltumatu jada põhjal, siis võtame kasutusele veel järgnevad tähistused:

$$\begin{aligned}\gamma_t^n(i) &= p(Z_t^n = i|x^{T_n}, y^{T_n}, \theta'), \\ \xi_t^n(i,j) &= p(Z_t^n = i, Z_{t+1}^n = j|x^{T_n}, y^{T_n}, \theta'), \\ \gamma_t^n(i,\ell) &= p(Z_t^n = i, R_t^n = \ell|x^{T_n}, y^{T_n}, \theta').\end{aligned}$$

Parameetrite ümberhindamise arvutusvalemiteks on seega:

$$
\begin{aligned}
\hat{\pi}_i &= \frac{1}{N}\sum_{n=1}^{N} p(Z_1^n = i|x^{T_n}, y^{T_n}, \theta') = \frac{1}{N}\sum_{n=1}^{N} \gamma_1^n(i), \\
\hat{p}_{ij} &= \frac{\sum_{n=1}^{N}\sum_{t=1}^{T_n-1} p(Z_t^n = i, Z_{t+1}^n = j|x^{T_n}, y^{T_n}, \theta')}{\sum_{n=1}^{N}\sum_{t=1}^{T_n-1} p(Z_t^n = i|x^{T_n}, y^{T_n}, \theta')} = \frac{\sum_{n=1}^{N}\sum_{t=1}^{T_n-1} \xi_t^n(i,j)}{\sum_{n=1}^{N}\sum_{t=1}^{T_n-1} \gamma_t^n(i)}, \\
\hat{c}_{i\ell} &= \frac{\sum_{n=1}^{N}\sum_{t=1}^{T_n} p(Z_t^n = i, R_t^n = \ell|x^{T_n}, y^{T_n}, \theta')}{\sum_{n=1}^{N}\sum_{t=1}^{T_n} p(Z_t^n = i|x^{T_n}, y^{T_n}, \theta')} = \frac{\sum_{n=1}^{N}\sum_{t=1}^{T} \gamma_t^n(i,\ell)}{\sum_{n=1}^{N}\sum_{t=1}^{T_n} \gamma_t^n(i)}, \\
\hat{\mu}_{i,f} &= \frac{\sum_{n=1}^{N}\sum_{t=1}^{T_n} x_t^n p(Z_t^n = i|x^{T_n}, y^{T_n}, \theta')}{\sum_{n=1}^{N}\sum_{t=1}^{T_n} p(Z_t^n = i|x^{T_n}, y^{T_n}, \theta')} = \frac{\sum_{n=1}^{N}\sum_{t=1}^{T_n} x_t^n \gamma_t^n(i)}{\sum_{n=1}^{N}\sum_{t=1}^{T_n} \gamma_t^n(i)}, \\
\hat{\Sigma}_{i,f} &= \frac{\sum_{n=1}^{N}\sum_{t=1}^{T_n}(x_t^n - \hat{\mu}_{i,f})(x_t^n - \hat{\mu}_{i,f})^T p(Z_t^n = i|x^{T_n}, y^{T_n}, \theta')}{\sum_{n=1}^{N}\sum_{t=1}^{T_n} p(Z_t^n = i|x^{T_n}, y^{T_n}, \theta')} \\
&= \frac{\sum_{n=1}^{N}\sum_{t=1}^{T_n}(x_t^n - \hat{\mu}_{i,f})(x_t^n - \hat{\mu}_{i,f})^T \gamma_t^n(i)}{\sum_{n=1}^{N}\sum_{t=1}^{T_n} \gamma_t^n(i)}, \\
\hat{\mu}_{\ell,g} &= \frac{\sum_{n=1}^{N}\sum_{t=1}^{T_n} y_t^n p(R_t^n = \ell|x^{T_n}, y^{T_n}, \theta')}{\sum_{n=1}^{N}\sum_{t=1}^{T_n} p(R_t^n = \ell|x^{T_n}, y^{T_n}, \theta')} = \frac{\sum_{n=1}^{N}\sum_{t=1}^{T_n} y_t^n (\sum_{i=1}^{K} \gamma_t^n(i,\ell))}{\sum_{n=1}^{N}\sum_{t=1}^{T_n} (\sum_{i=1}^{K} \gamma_t^n(i,\ell))}, \\
\hat{\sigma}_{\ell,g}^2 &= \frac{\sum_{n=1}^{N}\sum_{t=1}^{T_n}(y_t^n - \hat{\mu}_{\ell,g})^2 p(R_t^n = \ell|x^{T_n}, y^{T_n}, \theta')}{\sum_{n=1}^{N}\sum_{t=1}^{T_n} p(R_t^n = \ell|x^{T_n}, y^{T_n}, \theta')} \\
&= \frac{\sum_{n=1}^{N}\sum_{t=1}^{T_n}(y_t^n - \hat{\mu}_{\ell,g})^2 (\sum_{i=1}^{K} \gamma_t^n(i,\ell))}{\sum_{n=1}^{N}\sum_{t=1}^{T_n} (\sum_{i=1}^{K} \gamma_t^n(i,\ell))}.
\end{aligned}
$$

Paneme tähele, et juht $N = 1$ annab parameetrite ümberhindamise valemid ühe jada korral.

2.3 EM-algoritmi rakendamine simuleeritud andmestikuga

Enne mudeli rakendamist päris andmetele teeme läbi lihtsa näite, et illustreerida ja kontrollida algoritmi tööd. Simuleerime väikse andmestiku, misjärel proovime kolme erinevat alglähendite komplekti, et näha, kuidas mudel erinevates olukordades parameetreid hindab. Mudeli rakendamiseks vajalike funktsioonide programmikood on toodud lisas 1, andmete simuleerimise ja parameetrite hindamise kood ühe alglähendite komplekti korral on toodud lisas 2.

MRT seisundite arvuks määrasime $K = 2$ ja KT seisundite arvuks $M = 3$. Simuleerimisel kasutatud alg- ja üleminekutõenäosused ning emissioonijaotuste parameetrid on ära toodud tabeli 1 (ja ka tabelite 2 ning 3) esimeses veerus. Kokku simuleerisime 50 sõltumatut vaatluste jada, neist igaüks pikkusega vahemikus 500-1000. MRT vaatlused simuleerisime kahemõõtmelistest normaaljaotustest.

Tabel 1: Hinnatud parameetrid esimese alglähendite komplektiga

	tegelik parameeter	valitud alglähend	hinnatud parameeter
π_i	$\begin{pmatrix} 0,5 & 0,5 \end{pmatrix}$	$\begin{pmatrix} 0,6 & 0,4 \end{pmatrix}$	$\begin{pmatrix} 0,40 & 0,60 \end{pmatrix}$
p_{ij}	$\begin{pmatrix} 0,7 & 0,3 \\ 0,2 & 0,8 \end{pmatrix}$	$\begin{pmatrix} 0,8 & 0,2 \\ 0,3 & 0,7 \end{pmatrix}$	$\begin{pmatrix} 0,70 & 0,30 \\ 0,20 & 0,80 \end{pmatrix}$
$c_{i\ell}$	$\begin{pmatrix} 0,7 & 0,1 & 0,2 \\ 0,1 & 0,6 & 0,3 \end{pmatrix}$	$\begin{pmatrix} 0,8 & 0,1 & 0,1 \\ 0,1 & 0,8 & 0,1 \end{pmatrix}$	$\begin{pmatrix} 0,70 & 0,16 & 0,14 \\ 0,17 & 0,51 & 0,32 \end{pmatrix}$
$\mu_{i,f}$	$\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 \\ 6 \end{pmatrix}$	$\begin{pmatrix} \mathbf{3} \\ \mathbf{3} \end{pmatrix}, \begin{pmatrix} \mathbf{3} \\ \mathbf{3} \end{pmatrix}$	$\begin{pmatrix} 0,99 \\ 1,98 \end{pmatrix}, \begin{pmatrix} 3,98 \\ 5,99 \end{pmatrix}$
$\Sigma_{i,f}$	$\begin{pmatrix} 1 & 0,4 \\ 0,4 & 1 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 1,5 \end{pmatrix}$	$\begin{pmatrix} 1 & 0,5 \\ 0,5 & 1 \end{pmatrix}, \begin{pmatrix} 2,5 & 1,2 \\ 1,2 & 2,3 \end{pmatrix}$	$\begin{pmatrix} 1,00 & 0,38 \\ 0,38 & 0,97 \end{pmatrix}, \begin{pmatrix} 2,03 & 1,01 \\ 1,01 & 1,51 \end{pmatrix}$
$\mu_{\ell,g}$	1, 3, 5	**3**, **3**, **3**	1,00, 3,30, 4,94
$\sigma^2_{\ell,g}$	1, 4, 6,25	1,2, 3,8, 6,7	1,07, 4,17, 6,78

Kahel esimesel juhul valisime teadlikult enamiku alglähenditest tegelikele parameetrite väärtustele küllaltki lähedased. Esimesel juhul (vt tabel 1) määrasime aga seisundite keskväärtuste alglähendid tegelikega võrreldes suhteliselt suvalised ning teisel juhul (vt tabel 2) tegime samamoodi kovariatsioonimaatriksite ja dispersioonidega, vastavad kohad on tabelis märgitud paksus kirjas.

Näeme, et esimesel kahel juhul on EM-algoritmiga saadud keskväärtuste hinnangud tegelike keskväärtustega väga sarnased. Lisaks, MRT kovariatsioonimaatriksite hinnangud on mõlemal juhul täpsed. KT seisundite dispersioonid tulevad esimesel juhul samuti üsna täpsed, teisel juhul on erinevus tegelike parameetri väärtustega veidi suurem. Muud parameetrid näivad olevat üsna hästi hinnatud ega tundu eriti sõltuvat keskväärtuste ja dispersioonide algländenditest.

Tabel 2: Hinnatud parameetrid teise alglähendite komplektiga

	tegelik parameeter	valitud alglähend	hinnatud parameeter
π_i	$\begin{pmatrix} 0{,}5 & 0{,}5 \end{pmatrix}$	$\begin{pmatrix} 0{,}4 & 0{,}6 \end{pmatrix}$	$\begin{pmatrix} 0{,}40 & 0{,}60 \end{pmatrix}$
p_{ij}	$\begin{pmatrix} 0{,}7 & 0{,}3 \\ 0{,}2 & 0{,}8 \end{pmatrix}$	$\begin{pmatrix} 0{,}8 & 0{,}2 \\ 0{,}3 & 0{,}7 \end{pmatrix}$	$\begin{pmatrix} 0{,}70 & 0{,}30 \\ 0{,}20 & 0{,}80 \end{pmatrix}$
$c_{i\ell}$	$\begin{pmatrix} 0{,}7 & 0{,}1 & 0{,}2 \\ 0{,}1 & 0{,}6 & 0{,}3 \end{pmatrix}$	$\begin{pmatrix} 0{,}8 & 0{,}1 & 0{,}1 \\ 0{,}1 & 0{,}8 & 0{,}1 \end{pmatrix}$	$\begin{pmatrix} 0{,}73 & 0{,}14 & 0{,}13 \\ 0{,}21 & 0{,}52 & 0{,}27 \end{pmatrix}$
$\mu_{i,f}$	$\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 \\ 6 \end{pmatrix}$	$\begin{pmatrix} 1{,}2 \\ 1{,}9 \end{pmatrix}, \begin{pmatrix} 3 \\ 5 \end{pmatrix}$	$\begin{pmatrix} 0{,}99 \\ 1{,}99 \end{pmatrix}, \begin{pmatrix} 3{,}99 \\ 5{,}99 \end{pmatrix}$
$\Sigma_{i,f}$	$\begin{pmatrix} 1 & 0{,}4 \\ 0{,}4 & 1 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 1{,}5 \end{pmatrix}$	$\begin{pmatrix} \mathbf{3} & \mathbf{2} \\ \mathbf{2} & \mathbf{3} \end{pmatrix}, \begin{pmatrix} \mathbf{3} & \mathbf{2} \\ \mathbf{2} & \mathbf{3} \end{pmatrix}$	$\begin{pmatrix} 1{,}00 & 0{,}39 \\ 0{,}39 & 0{,}98 \end{pmatrix}, \begin{pmatrix} 2{,}02 & 1{,}00 \\ 1{,}00 & 1{,}50 \end{pmatrix}$
$\mu_{\ell,g}$	1, 3, 5	0,8, 3,2, 4,9	1,00, 3,31, 5,62
$\sigma^2_{\ell,g}$	1, 4, 6,25	**3**, **3**, **3**	1,19, 3,50, 5,44

Kolmandal juhul (vt tabel 3) kasutasime alglähendite leidmiseks olemasolevaid andmeid. Rakendasime andmetele K-keskmiste meetodit, et leida saadud klasterduste põhjal kahe MRT seisundi ja kolme KT seisundi parameetrite hinnangute alglähendid. Ka ülejäänud parameetrite alglähendid hindasime andmete põhjal, kasutades leitud klastrite esinemise sagedusi andmestikus. Täpsemad arvutusmeetodid on näha lisas 2.

Tabel 3: Hinnatud parameetrid K-keskmiste meetodiga leitud alglähendite komplektiga

	tegelik parameeter	leitud alglähend	hinnatud parameeter
π_i	$\begin{pmatrix} 0,5 & 0,5 \end{pmatrix}$	$\begin{pmatrix} 0,42 & 0,58 \end{pmatrix}$	$\begin{pmatrix} 0,40 & 0,60 \end{pmatrix}$
p_{ij}	$\begin{pmatrix} 0,7 & 0,3 \\ 0,2 & 0,8 \end{pmatrix}$	$\begin{pmatrix} 0,67 & 0,33 \\ 0,28 & 0,72 \end{pmatrix}$	$\begin{pmatrix} 0,70 & 0,30 \\ 0,20 & 0,80 \end{pmatrix}$
$c_{i\ell}$	$\begin{pmatrix} 0,7 & 0,1 & 0,2 \\ 0,1 & 0,6 & 0,3 \end{pmatrix}$	$\begin{pmatrix} 0,63 & 0,25 & 0,12 \\ 0,33 & 0,45 & 0,22 \end{pmatrix}$	$\begin{pmatrix} 0,76 & 0,12 & 0,12 \\ 0,31 & 0,46 & 0,23 \end{pmatrix}$
$\mu_{i,f}$	$\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 \\ 6 \end{pmatrix}$	$\begin{pmatrix} 1,05 \\ 2,19 \end{pmatrix}, \begin{pmatrix} 4,19 \\ 6,16 \end{pmatrix}$	$\begin{pmatrix} 1,00 \\ 1,99 \end{pmatrix}, \begin{pmatrix} 3,99 \\ 6,00 \end{pmatrix}$
$\Sigma_{i,f}$	$\begin{pmatrix} 1 & 0,4 \\ 0,4 & 1 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 1,5 \end{pmatrix}$	$\begin{pmatrix} 0,98 & 0,42 \\ 0,42 & 1,28 \end{pmatrix}, \begin{pmatrix} 1,61 & 0,66 \\ 0,66 & 1,22 \end{pmatrix}$	$\begin{pmatrix} 1,00 & 0,39 \\ 0,39 & 0,99 \end{pmatrix}, \begin{pmatrix} 2,01 & 0,99 \\ 0,99 & 1,49 \end{pmatrix}$
$\mu_{\ell,g}$	1, 3, 5	0,76, 3,54, 6,91	0,97, 3,60, 6,43
$\sigma^2_{\ell,g}$	1, 4, 6,25	0,86, 0,78, 1,98	1,25, 1,70, 3,51

Näeme, et ka kolmandal juhul on enamik parameetritest üsna hästi hinnatud, kuid KT seisundite keskväärtuste ja dispersioonide hinnangutes on suuremaid erinevusi tegelike parameetrite väärtustega võrreldes. Esimese kahe juhuga võrreldes on ka kaalude $c_{i\ell}$ hinnangud ebatäpsemad. Lisaks paneme tähele, et algtõenäosused on kõigil kolmel juhul hinnatud samaks: määratud $(0,5 \quad 0,5)$ asemel $(0,4 \quad 0,6)$, mis vastab ka tegelikult realiseerunud algseisundite sagedustele simuleeritud 50 jada seas.

Katsetasime alglähendite leidmist andmete klasterdamise abil veel kahel juhul, kus kasutasime selleks vaid vastavalt 20 ja 40 jada. Tulemustes suuri erinevusi ei kajastunud.

Paneme tähele, et kohad, kus antud näites saime kehvemaid tulemusi, on valitud parameetrite komplekti keerukuse tõttu ka põhjendatavad. Vaadates näiteks KT komponentide keskväärtuseid ja dispersioone, näeme, et komponentide jaotused kattuvad omavahel suures osas ega olegi seetõttu lihtsasti eristatavad. Samas MRT seisundite jaotused on määratud parameetrite korral väga selgesti eristatavad, sellest ka nende ja üleminekutõenäosuste täpsed hinnangud. Kokkuvõttes näeme, et kui alglähendid on lähedased tegelikele parameetritele, siis algoritm hindab küllaltki hästi.

3 Mudeli rakendamine MRT ja KT andmetele

Rakendame eelmises peatükis kirjeldatud tingliku sõltumatuse mudelit, et prognoosida KT vaatluseid MRT vaatluste põhjal. Artiklis Kuljus *et al.* (2018) on samale probleemile lähenetud teisi varjatud tunnustega mudeleid kasutades. Kirjeldame tingliku sõltumatuse mudeli käitumist ning toome välja sarnasused ja suuremad erinevused artikli tulemustega võrreldes.

3.1 Andmed

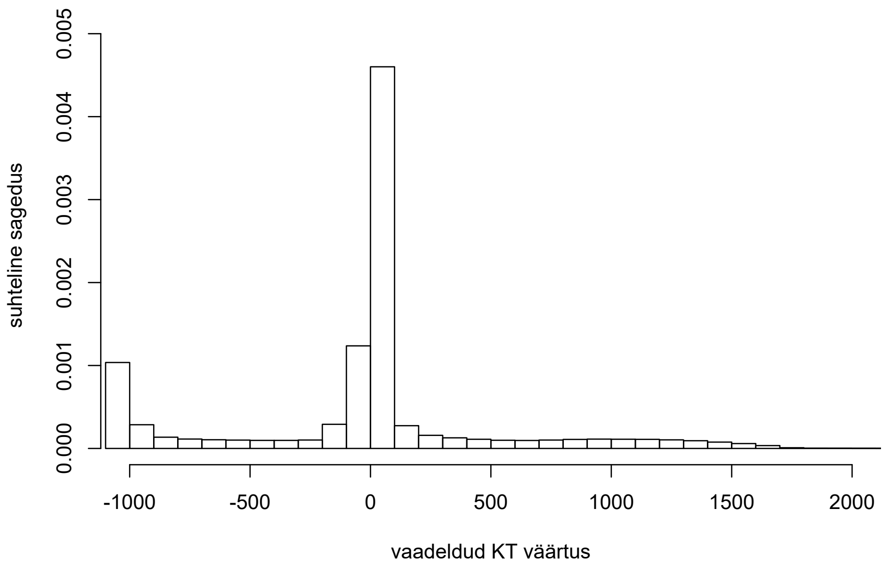
Alapeatükk on kirjutatud artikli Kuljus *et al.* (2018) põhjal.

Kasutame viie patsiendi pea andmeid, mis artiklis on nummerdatud kui pea 1, 2, 4, 8 ja 9. Iga pea kohta on olemas KT-pilt ning neli MRT-pilti, millest viimased on saadud erinevaid magnetvälja parameetrite komplekte kasutades (täpsem kirjeldus on leitav artiklis). MRT vaatluste väärtused varieeruvad vahemikus 0-800. Tabelis 4 on toodud mõned näited KT väärtuste kohta Hounsfieldi skaalal.

Tabel 4: KT väärtused tavalisemate kudede korral (Naidich *et al.*, 2012)

kude	KT väärtus
õhk	< -1000
rasv	$-100\ldots-20$
vesi	$-20\ldots20$
valgeollus	$20\ldots35$
hallollus	$30\ldots40$
lihaskude	$20\ldots40$
kaltsifikatsioon	> 150
luu	$800\ldots1200$

Eri allikatest võib aga erinevaid kudedele vastavaid KT väärtuste vahemikke leida. Näiteks meie andmestikus on mõned luule vastavad väärtused ka suurusjärgus kuni 3000, mis viitab tihedamale luule (Bibb *et al.*, 2015). KT väärtuste jaotus peade 2, 4, 8 ja 9 korral on näha joonisel 1.

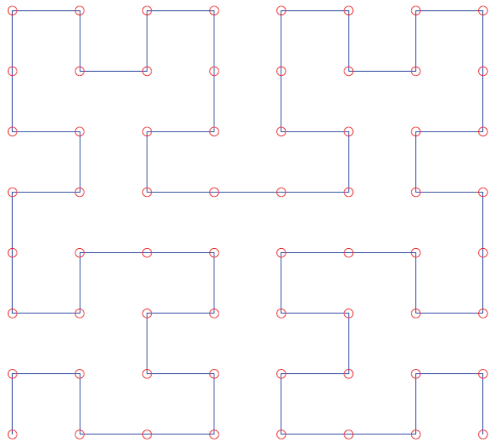


Joonis 1: KT väärtuste jaotus peade 2, 4, 8 ja 9 korral

Kõik pildid on viidud kujule, kus mõõdetud väärtused on olemas iga voksli jaoks kuubis suurusega $192\times192\times192$, seejuures piltidevaheline vokslite ühtivus on tagatud. Et eristada pea vaatluste voksleid ümbritseva õhu vokslitest, on välja arvutatud binaarne tunnus: 1 - pea, 0 - ümbritsev õhk.

Et saaksime 3-mõõtmelistele andmetele tingliku sõltumatuse mudelit rakendada, tuleb need esmalt viia jada kujule ehk järjestada. Selleks on tavaline kasutada Hilberti joont, mis püüab läbi ruumi liikudes arvesse võtta ja säilitada andmepunktide lokaalset struktuuri. Näide Hilberti joonest kahemõõtmelisel juhul on toodud joonisel 2. Pärast sellist andmepunktide järjestamist on igal vaatlusel kaks naabrit.

Kujutades ette pea horisontaalset või vertikaalset lõiku joonise 2 keskel, näeme, et joon võib vahel peast välja ümbritseva õhu alasse liikuda ja seejärel tagasi sisse pöörduda. Kuna meid huvitavad vaid pea vaatluste vokslid ning on kohti, kus joon väljub peast ühe voksli juures, kuid tagasi peasse siseneb sellise kaudu, mis ei ole viimatise naabriks, siis ühe katkematu jada asemel tekib meil mitu sõltumatut pea vaatluste jada.



Joonis 2: Näide Hilberti joonest tasandil

Pea 1 vaatlustest moodustub näiteks 12 239 sõltumatut jada, kus vaatluste voksleid kokku on 1 853 702. Keskmiseks jada pikkuseks on 152 ning kõige pikemas jadas on 108 205 vokslit. Kokku on pea 1 kohta ka 2299 sellist jada, mis koosnevad vaid ühest vaatlusest. Teiste peade korral on kõik suurusjärgud sarnased. Kuna andmeid on meil piisavalt ning ühe-elemendilised jadad näiteks üleminekutõenäosuste kohta mingit infot juurde ei anna, siis neid me mudeli hindamisel ei kasuta.

3.2 KT-piltide hindamine

Rakendame tingliku sõltumatuse mudelit, et leida MRT-piltide põhjal KT-pildi hinnang ehk nii-öelda substituut-KT-pilt. Teeme läbi näite, kus leiame pea 1 jaoks substituut-KT-pildi, kasutades parameetrite hindamiseks ülejäänud nelja pea andmeid. Seega moodustavad pead 2, 4, 8 ja 9 meie treeningandmestiku.

Mudeli parameetrite hinnangud leiame EM-algoritmiga. Peatükis 2 nägime, et kõikide parameetrite hinnangute arvutusvalemid on taandatavad edasi- ja tagasi-tõenäosuste kasutamisele. Vastavad funktsioonid ja EM-algoritm on implementeeritud statistikatarkvaraga R (vt lisa 1). Alglähendite valiku protsessi kirjeldame järgmises alapeatükis.

Olgu vaadeldava pea kohta mõõtmised T voksli jaoks. Pärast parameetrite hindamist saame substituut-KT-pildi arvutada voksli kaupa:

$$sKT_t = \sum_{i=1}^{K} \left(\sum_{\ell=1}^{M} \hat{c}_{i\ell} \hat{\mu}_{\ell,g} \right) P(Z_t = i | x^T, \hat{\theta}), \quad t = 1, \ldots, T. \tag{3}$$

Võrdluseks, tavalise HMM korral on arvutusvalemiks

$$s\widetilde{KT}_t = \sum_{i=1}^{K} \tilde{\mu}_i(x_t) P(Z_t = i | x^T, \hat{\psi}), \quad t = 1, \ldots, T,$$

kus $\tilde{\mu}_i(x_t) = E(Y_t | x_t, Z_t = i, \hat{\psi})$ on KT vaatluse tinglik keskväärtus ning $\hat{\psi}$ tähistab kõiki HMM parameetrite hinnanguid (Kuljus *et al.*, 2018). Paneme tähele, et kaalud $P(Z_t = i | x^T)$ on mõlemas arvutusvalemis samal kujul. Suurim erinevus seisneb aga selles, et $s\widetilde{KT}_t$ arvutamisel tuleb MRT vaatluse väärtus x_t otseselt $\tilde{\mu}_i(x_t)$ leidmiseks avaldisse sisse. Seevastu sKT_t avaldises esinevad MRT väärtused vaid kaudselt kaalude kaudu. Seega, teoreetiliselt peaks HMM hinnang olema selle poolest informatiivsem.

Artikli Kuljus *et al.* (2018) eeskujul hindame kolm erinevat mudelit, kus MRT seisundite arvuks määrame $K = 5$ ning KT seisundite arvuks kas $M = 5$, $M = 8$ või $M = 10$. Tähistame mudeleid vastavalt M5, M8 ja M10.

Mudelite tulemuste võrdlemiseks arvutame välja ka substituutpiltide keskmised absoluutvead (MAE). Olgu j-ndale vokslile vastav tegelik KT väärtus KT_j, siis

$$MAE = \frac{1}{T}\sum_{j=1}^{T}|KT_j - sKT_j|.$$

Piirdume vaid ühe pea substituutpildi leidmisega, kuna suure andmehulga tõttu on mudelite hindamine arvutuslikult töömahukas. Olenevalt seisundite arvust võib ühe mudeli hindamise kogu protsess võtta aega ligi 25 tundi. Ühtlasi leiame, et ühe peaga näite läbi tegemine annab piisavalt aimu mudeli käitumise kohta.

3.3 Mudeli parameetrite alglähendite valik

Kuna EM-algoritmi tulemus sõltub alglähendite valikust ja algoritmi rakendamine on meie andmete ning implementeeritud funktsioonide korral ajamahukas, siis mudeli hindamisel parima võimaliku tulemuse garanteerimiseks tahame leida kolm head algparameetrite komplekti. Informatiivsete alglähendite leidmiseks kasutame aja kokkuhoiu mõttes olemasolevate treeningandmete väiksemaid alamandmestikke.

Moodustame nelja pea andmetest väiksemad alamandmestikud kolmel erineval viisil:

1) valime igast peast juhuslikult umbes neljandiku andmejadasid,

2) kasutame ainult pea 8 andmeid,

3) valime igast peast juhuslikult umbes pooled andmejadad.

Iga alamandmestiku korral leiame nn väikse EM-algoritmi (peatame algoritmi pärast 5. iteratsiooni) abil ühe alglähendite komplekti. Väikse EM-algoritmi rakendamiseks alamandmestikele on meil samuti vaja leida alglähendid. Need leiame K-keskmiste meetodi abil, mis omakorda kasutab juhuslikke alglähendeid.

Iga alamandmestiku korral rakendame nii MRT kui ka KT väärtustele K-keskmiste meetodit, et MRT väärtuste põhjal hinnata K klastri keskväärtused ja kovariatsioonimaatriksid ning KT väärtuste põhjal M klastri keskväärtused ja dispersioonid. Selle tulemusena saame iga voksli liigitada ühte MRT klastrisse i, $i \in \{1, \ldots, K\}$ ja ühte KT klastrisse ℓ, $\ell \in \{1, \ldots, M\}$. Kaalude $c_{i\ell}$ alglähendite leidmiseks kasutame kahe klasterduse ühisjaotust. Näiteks, leides kõikide (MRT klasterduse põhjal) klastrisse 3 kuuluvate vokslite korral, kui suur osa nendest kuulub (KT klasterduse põhjal) klastrisse 5, annab see meile alglähendi c_{35} jaoks. Algtõenäosused määrame esialgu võrdsetena $1/K$ ning üleminekutõenäosuste puhul lähtume sellest, et peadiagonaalil oleks võrdsed suuremad tõenäosused ($> 0,5$) ning ülejäänud tõenäosusmass oleks jagatud igas reas võrdselt.

Kirjeldatud alglähenditega viimegi vastavatel alamandmestikel läbi väikse EM-algoritmi sammu. Saame kõikidele parameetritele hinnangud, mille sobivust andmetega iseloomustab log-tõepära.

Eelnevalt kirjeldatud protsessi ehk klasterdamist ja seejärel väikse EM-algoritmi rakendamist viime kõigi kolme alamandmestiku korral läbi kolm korda, misjärel valime iga alamandmestiku korral kolmest hinnatud mudelist log-tõepära põhjal parima. Kokkuvõttes saame niimoodi iga alamandmestiku kohta ühe alglähendite komplekti ehk kokku kolm komplekti.

3.4 Tulemused

Hindasime mudelid M5, M8 ja M10. Toome välja tähtsamad tulemused ning tähelepanekud. Märgime ära ka peamised sarnasused ja erinevused artikli Kuljus *et al.* (2018) tulemustega.

KT komponentide parameetrite hinnangud on iga mudeli jaoks välja toodud tabelites 5, 6 ja 7, kus $\hat{\mu}_{\ell,g}^{(k)}$ ja $\hat{\sigma}_{\ell,g}^{(k)}$ tähistavad k-nda alglähendite komplektiga saadud hinnanguid. Parameetrid, mille korral saavutasime suurima log-tõepära (võrreldes teiste alglähenditega leitutega), on märgitud paksus kirjas.

Tabel 5: Hinnangud viiele KT komponendi keskväärtusele ja standardhälbele kolme erineva alglähendite komplekti korral

ℓ	1	2	3	4	5
$\hat{\mu}_{\ell,g}^{(1)}$	-1021,7	-731,9	-4,4	33,7	848,6
$\hat{\mu}_{\ell,g}^{(2)}$	**-1021,2**	**-709,2**	**-2,9**	**33,7**	**849,9**
$\hat{\mu}_{\ell,g}^{(3)}$	-1021,4	-743,7	-4,2	32,7	885,7
$\hat{\sigma}_{\ell,g}^{(1)}$	3,5	250,1	87,5	9,1	469,3
$\hat{\sigma}_{\ell,g}^{(2)}$	**4,3**	**261,1**	**85,4**	**8,9**	**468,4**
$\hat{\sigma}_{\ell,g}^{(3)}$	4,0	235,2	103,7	10,7	453,8

Tabel 6: Hinnangud kaheksale KT komponendi keskväärtusele ja standardhälbele kolme erineva alglähendite komplekti korral

ℓ	1	2	3	4	5	6	7	8
$\hat{\mu}_{\ell,g}^{(1)}$	**-1021,2**	**-722,6**	**-40,3**	**32,5**	**72,7**	**584,2**	**1192,1**	**1549,8**
$\hat{\mu}_{\ell,g}^{(2)}$	-1021,1	-719,8	-38,9	32,6	78,2	594,7	1078,3	1191,8
$\hat{\mu}_{\ell,g}^{(3)}$	-1021,1	-719,4	-40,1	32,6	76,0	595,6	1197,4	1250,6
$\hat{\sigma}_{\ell,g}^{(1)}$	**4,2**	**249,9**	**57,0**	**11,0**	**138,7**	**221,6**	**259,6**	**975,3**
$\hat{\sigma}_{\ell,g}^{(2)}$	4,5	249,7	58,6	11,0	143,1	214,1	547,9	239,0
$\hat{\sigma}_{\ell,g}^{(3)}$	4,5	249,9	57,8	11,0	141,4	221,4	253,1	853,8

Küll aga toome välja, et suurest andmemahust tulenevalt oli log-tõepärade suurusjärk -190 miljonit ning kui vaadata nende protsentuaalseid erinevusi erinevate alglähendite korral, siis eri mudelite puhul jäid need vahemikku 0,0003-0,097%. Seega, kõigi kolme alglähendite komplektiga leitud hinnangud on log-tõepära suhtes sarnased ja stabiilsed. Suurim log-tõepära oli mudelil M10, milleks oli -187 812 223. M8 suurim log-tõepära oli sellest 0,64% väiksem ning M5 oma 0,81% väiksem.

Näeme, et üldiselt käituvad hinnangud erinevate alglähendite korral üsna stabiilselt. Suuremaid erinevusi on märgata eelkõige suurtele KT väärtustele vastavate komponentide osas, mis peamiselt sisaldavad infot luu kohta. Näiteks, M8 mudeli 7. ja 8. komponendi ning M10 mudeli komponentide 8-10 puhul on hinnangute varieeruvus erinevate alglähendite korral märgatavalt suurem kui teiste komponentide korral. Jooniselt 1 näeme, et piirkonnas 1500 ümber sarnaneb KT väärtuste jaotus pigem ühtlase jaotusega, seega normaaljaotuse sobitamine võibki anda siinkohal ebastabiilseid tulemusi. Väga selgelt on iga mudeli korral paigas aga näiteks esimene komponent, mille keskväärtus vastab õhule.

Tabel 7: Hinnangud kümnele KT komponendi keskväärtusele ja standardhälbele kolme erineva alglähendite komplekti korral

ℓ	1	2	3	4	5
$\hat{\mu}_{\ell,g}^{(1)}$	-1023,8	-1008,5	-596,8	-48,5	33,7
$\hat{\mu}_{\ell,g}^{(2)}$	**-1023,8**	**-1008,0**	**-594,8**	**-48,5**	**33,8**
$\hat{\mu}_{\ell,g}^{(3)}$	-1023,8	-1007,6	-593,1	-50,0	33,8
$\hat{\sigma}_{\ell,g}^{(1)}$	0,3	14,7	271,5	53,9	8,1
$\hat{\sigma}_{\ell,g}^{(2)}$	**0,3**	**15,0**	**270,8**	**54,1**	**8,0**
$\hat{\sigma}_{\ell,g}^{(3)}$	0,3	15,4	270,5	52,7	7,9

ℓ	6	7	8	9	10
$\hat{\mu}_{\ell,g}^{(1)}$	41,3	308,9	946,3	1410,2	2571,4
$\hat{\mu}_{\ell,g}^{(2)}$	**40,1**	**273,4**	**799,8**	**1280,3**	**1460,0**
$\hat{\mu}_{\ell,g}^{(3)}$	39,8	269,8	825,3	1306,2	2460,6
$\hat{\sigma}_{\ell,g}^{(1)}$	42,3	211,7	240,7	170,7	458,2
$\hat{\sigma}_{\ell,g}^{(2)}$	**38,2**	**182,5**	**197,5**	**221,0**	**1042,0**
$\hat{\sigma}_{\ell,g}^{(3)}$	37,9	192,0	209,4	212,2	558,1

Paneme tähele, et viielt komponendilt kaheksale liikudes tulevad välja uued suuremate väärtustega KT klassid, mida M5 mudel ei tuvastanud. Artikli Kuljus *et al.* (2018) HMM mudelite korral võib märgata, et seisundite arvu suurendamine tekitas mõnda juba väiksema mudeli poolt tuvastatud väärtuste piirkonda uusi seisundeid juurde, andmata seejuures uut informatsiooni suuremate KT väärtustega klasside kohta. Näiteks, üleminek viielt seisundilt kaheksale tekitas keskväärtusega -13 komponendi asemele komponendid keskväärtustega -47 ning -39 ja keskväärtusega 32 komponendi asemele komponendid keskväärtustega 29, 34 ja 43. Tingliku sõltumatuse mudeli korral on hinnatud KT komponentide keskväärtused aga ühtlasemalt kogu väärtuste piirkonna peale laiali jaotunud. Edasi kümne komponendi peale liikudes näeme, et näiteks mudelis M8 keskväärtusega 32,5 hinnatud komponent on nüüd jaotunud komponentideks keskväärtustega 33,8 ja 40,1, tänu millele on eristatud vastavalt valge- ja hallollus.

Leitud substituutpiltide keskmised absoluutvead on välja toodud tabelis 8. Arvutasime substituutpildid ka peadele 2, 4, 8 ja 9, kuid kuna samade peade andmeid kasutasime tegelikult ka parameetrite hindamisel, siis ainus artikliga Kuljus *et al.* (2018) kõrvutatav tulemus on pea 1 kohta käiv esimesel real. Artiklis on sama info teiste mudelite jaoks tabelis 1 esimesel real.

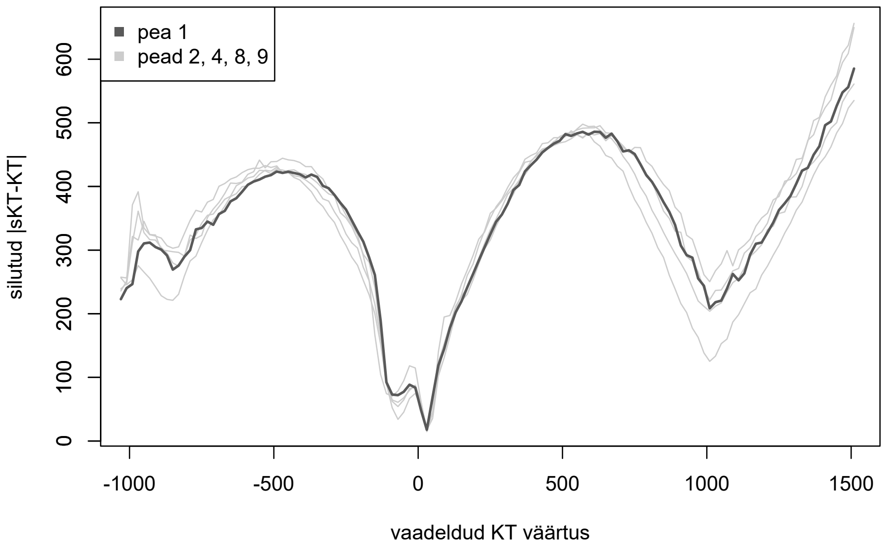
Tabel 8: Keskmised absoluutvead kolme mudeli korral

pea	M5	M8	M10
1	**174,44**	**156,38**	**158,66**
2	174,12	155,13	159,36
4	185,44	167,12	170,71
8	177,42	158,31	161,68
9	177,51	163,83	165,01

Näeme, et keskmine absoluutviga tuli madalaim mudeli M8 korral. HMM korral oli sama näitaja esimese pea jaoks 146,31 ehk veidi madalam, kuid suurusjärk on sellegipoolest sarnane. M10 jääb M8 mudelile MAE poolest napilt alla, sama on märgata ka HMM korral. Sisuliselt pole MAE väärtustel kahe viimase mudeli korral vahet, seega sellest aspektist võime öelda, et kümne komponendi hindamine kaheksa asemel ei anna meile olulist lisainformatsiooni.

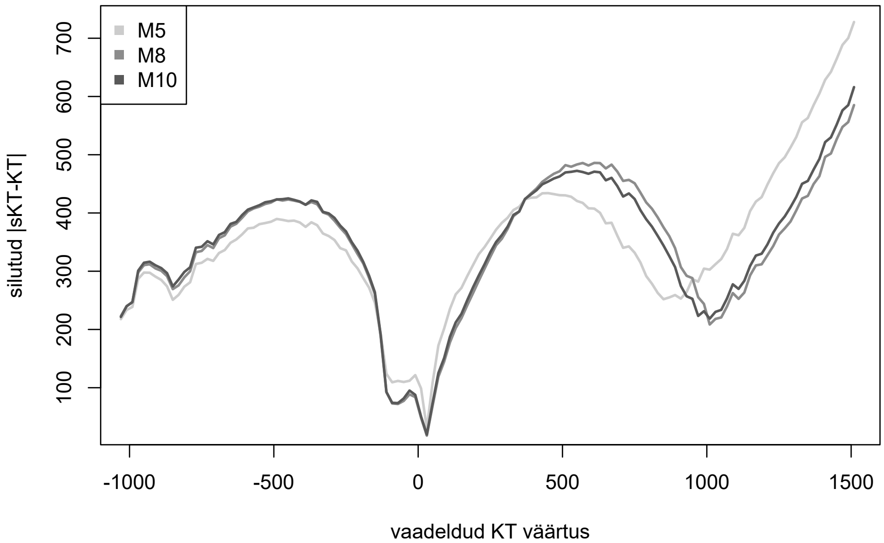
Nii M8 kui ka M10 suurimaks murekohaks on viimaste ehk luule vastavate komponentide hindamine. Nagu ka eelnevalt mainitud, siis KT väärtuste jaotus joonisel 1 selgitab mingil määral olukorda. Näeme, et lisaks väärtuste jaotuse sarnanemisele ühtlase jaotusega 1500 ümbruses on piirkonnas >1500 üleüldse väga vähe vaatlusi. Sellest tulenevalt on lisaks tulemuste ebastabiilsusele ka vastavate komponentide standardhälbed väga suured.

Probleemne luu piirkond tuleb selgelt välja ka joonistel 3 ja 4, kus on kujutatud substituutpildi ja tegeliku KT-pildi erinevuste absoluutväärtused. Joonistel toodud jäägid on silutud ehk täpsemini, iga (mittekattuva) 20-ühikulise väärtuste vahemiku kohta on leitud keskmine jääk ja kantud joonisele.



Joonis 3: Jääkide silutud absoluutväärtused M8 mudeli korral

Jooniselt 3 näeme, et erinevate peade korral käituvad jäägid üsna sarnaselt. Joonisel 4 on aga võrreldud esimese pea jaoks arvutatud substituut-KT-pildi jääke kolme erineva mudeli korral: suurima tõepära saavutanud M5, M8 ja M10. Näeme, et M5 tulemus erineb kahest ülejäänust ning on oluliselt kehvem väärtuste piirkonnas 1000-1500. M8 ja M10 seast ühte teisele eelistada on raske, kuna M10 käitub veidi paremini väärtuste vahemikus 500-1000, M8 aga vahemikus 1000-1500.

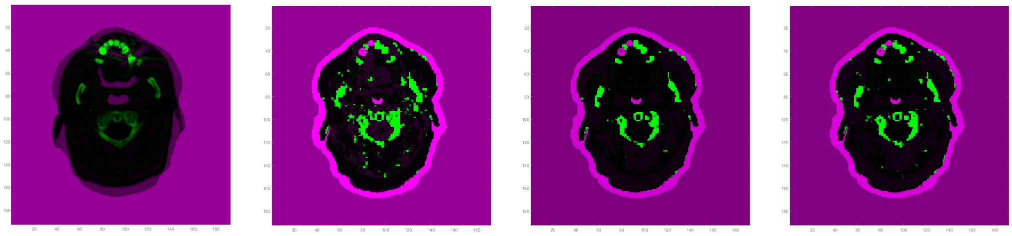


Joonis 4: Kolme mudeli jääkide võrdlus esimese pea korral

Jääkide joonised sarnanevad üldkuju poolest artikli Kuljus *et al.* (2018) omadega, kuid näiteks luu piirkonnas 800-1500 on tingliku sõltumatuse mudelite jäägid selgelt suuremad. HMM korral on väärtuse 1500 juures keskmine jääkide absoluutväärtus veidi üle 400, meie mudelite korral jääb sama suurus aga vahemikku 600-700. Ka õhule vastavate väärtuste juures (-1000 ümber) on HMM sooritus töös leitud mudelite omast parem. Jääkide absoluutväärtuste keskmine selles piirkonnas on HMM korral ligikaudu 100, tingliku sõltumatuse mudelite korral aga üle 200. Ülejäänud negatiivsete väärtuste juures on visuaalsel vaatlusel HMM ja meie mudelite sooritus jääkide poolest samaväärne. Väärtuste vahemikus 0-1000 on aga HMM korral jäägid jällegi väiksemad.

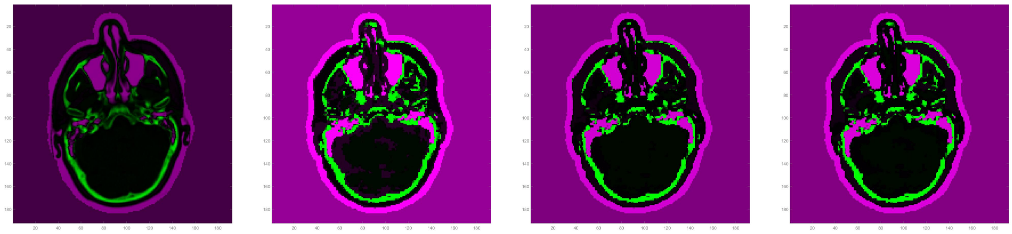
Teame, et MRT piltidele jäädvustatud info luu ja õhu kohta pole nende selgeks eristamiseks piisav ning see kajastub ka meie prognoositud substituutpiltides. Vaadates pea 1 KT- ja substituut-KT-pilte horisontaalsete lõikude haaval, on näha, et on pea piirkondi, kus prognoositud pildid näevad tegelikuga väga sarnased välja, kuid on ka piirkondi, kus luu ja õhu prognoosid on kas omavahel segamini aetud või lihtsalt valed.

Joonistel 5 ja 6 on pea horisontaalsetest lõikudest toodud kaks erinevat näidet. Roheline värvus vastab luule ning lilla õhule, taustavärvus ei kuulu piltide ega hinnangute juurde, vaid on selliselt seadistatud.



Joonis 5: Lõik 54, vasakult: KT-pilt, mudelitega M5, M8 ja M10 hinnatud substituut-KT-pildid

Joonisel 5 on üks paljudest lõikudest, mis läbib piirkonda, kus on suu ja hambad. Antud joonisel on hammaste näol hea näide sellest, kuidas mõnes kohas on luu ekslikult õhuks hinnatud. Lisaks on pea keskel olev õhuala suuresti tuvastamata jäänud. Mitmel järgmisel lõigul olid aga hambad korrektselt tuvastatud. Näeme, et substituutpiltidest on mudeli M5 pilt kõige mürasem ehk pehmete kudede piirkonnas võib näha luule vastavaid rohelisi täppe, M8 ja M10 omad on sellega võrreldes veidi selgemad. Joonisel 6 on näide nina piirkonda läbivast lõigust, kus substituutpildid vastavad tegelikule KT-pildile oluliselt paremini.



Joonis 6: Lõik 81, vasakult: KT-pilt, mudelitega M5, M8 ja M10 hinnatud substituut-KT-pildid

Substituutpiltide arvutamise paremaks mõistmiseks toome ära hinnangufunktsiooni (3) parima mudeli M8 korral. Kasutame KT segujaotuste keskväärtuste hinnangute tähistamiseks $\hat{\mu}_i^{KT}$, $i \in \{1, \ldots, K\}$, st

$$\hat{\mu}_i^{KT} = \sum_{\ell=1}^{K} \hat{c}_{i\ell} \hat{\mu}_{\ell,g},$$

siis KT-pildi hinnangufunktsioon avaldub kujul

$$sKT_t = \sum_{i=1}^{K} \hat{\mu}_i^{KT} P(Z_t = i | x^T, \hat{\theta}).$$

Kaalude hinnangud $\hat{c}_{i\ell}$ on ära toodud tabelis 9 ja $\hat{\mu}_{\ell,g}$ tabeli 6 reas $\hat{\mu}_{\ell,g}^{(1)}$. Nende põhjal arvutades saame:

$$(\hat{\mu}_1^{KT}, \ldots, \hat{\mu}_5^{KT}) \approx (-835, 1008, 33, -63, -14).$$

Näeme, et seisundile $Z_t = 1$ vastava $\hat{\mu}_1^{KT}$ saamiseks kaalutakse kokku suuresti $\hat{\mu}_{1,g} = -1021,2$ ja $\hat{\mu}_{2,g} = -722,6$. Seisundile $Z_t = 2$ vastav KT segujaotuse keskväärtus $\hat{\mu}_2^{KT}$ on saadud suuresti $\hat{\mu}_{6,g}$ ja $\hat{\mu}_{7,g}$ kokkukaalumisel. Klass $Z_t = 3$ vastab KT klassile keskväärtusega $\hat{\mu}_{4,g} = 32,5$. Tabelist 9 näeme, et KT klassile 8 keskväärtusega $\hat{\mu}_{8,g} = 1549,8$ vastavad kaalud $\hat{c}_{i,8}$ on väga väikesed.

Tabel 9: Kaalude $c_{i\ell}$ hinnangud parima mudeli M8 korral

$i \backslash \ell$	1	2	3	4	5	6	7	8
1	0,5238	0,4164	0,0388	0,0026	0,0161	0,0020	0,0000	0,0002
2	0,0001	0,0006	0,0000	0,0000	0,0073	0,2982	0,6745	0,0193
3	0,0000	0,0000	0,0000	0,9780	0,0220	0,0000	0,0000	0,0000
4	0,0058	0,0485	0,8420	0,0023	0,0921	0,0093	0,0000	0,0000
5	0,0161	0,1472	0,0345	0,1465	0,5514	0,0992	0,0004	0,0047

Kokkuvõttes saame öelda, et antud näites tuvastab tingliku sõltumatuse mudel KT seisundeid kogu väärtuste piirkonna ulatuses ühtlasemalt kui teised artiklis Kuljus *et al.* (2018) vaadeldud mudelid. Muus osas on mudeli sooritus üldjoontes kas samaväärne või veidi kehvem. Peamiseks probleemiks on luu ja õhu eristamine ning parameetrite hindamine luu piirkonnas ja nägime, et KT seisundite arvu suurendamine seda ei lahenda. Ilmselt saab siit järeldada, et mudelid, mis baseeruvad juhendamiseta õppel, ei ole antud ülesande lahendamiseks sobivad.

Järgmise sammuna tasuks proovida mudeleid, kus anname ise näiteks õhu ja luu kohta mudelile lisainformatsiooni ette, et sooritus nendes piirkondades saaks paraneda. Teine potentsiaalne võimalus mudelit parandada oleks kasutada mitme režiimiga varjatud Markovi mudeleid, kus erinevad režiimid võimaldaksid paremini arvesse võtta pea anatoomiat ja kudede erinevat jaotust erinevates piirkondades. Näiteks kudedevahelised üleminekutõenäosused võivad pea erinevates piirkondades olla väga erinevad.

Kokkuvõte

Magistritöö eesmärk oli KT vaatluste hindamine MRT vaatluste põhjal, kasutades selleks üht varjatud Markovi mudeli erijuhtu. Toodi ülevaade klassikalisest varjatud Markovi mudelist ja selle parameetrite hindamiseks kasutatavast EM-algoritmist, et seejärel kirjeldada töös kasutatavat tingliku sõltumatuse mudelit. Kasutati eeldust, et teades vaatlustele vastavate varjatud tunnuste väärtuseid ehk kudede klasse, on MRT ja KT vaatlused omavahel sõltumatud. Et võimaldada KT ja MRT vaatluste jaoks erinevat varjatud ahela seisundite arvu, modelleeriti KT vaatluste jaotust kui normaaljaotuste segu ning käsitleti selle komponente kui KT seisundeid.

Enne tingliku sõltumatuse mudeli rakendamist MRT ja KT andmetele viidi läbi lihtne näide simuleeritud andmetega, et illustreerida ja kontrollida vastava EM-algoritmi tööd. Selgus, et algoritm töötab väga hästi kudede klasside korral, mille jaotused on teineteisest hästi eraldatud. Jaotuste suurema omavahelise kattumise korral tulid parameetrite hinnangud veidi ebatäpsemad. Sellegipoolest oli näha, et kui algländendid on lähedased tegelikele parameetritele, siis algoritm töötab küllaltki hästi.

Töö viimases osas rakendati tingliku sõltumatuse mudelit viie pea andmetele, millest ühele leiti ülejäänud nelja pea põhjal hinnatud parameetritega substituut-KT-pilt. Selleks hinnati kolm erinevat mudelit, kus MRT seisundite arvuks määrati $K = 5$ ja KT seisundite arvuks kas $M = 5$, $M = 8$ või $M = 10$. Mudeleid nimetati vastavalt M5, M8 ja M10. Iga mudeli hindamiseks kasutati kolme erinevat alglähendite komplekti. KT parameetrite hinnangud toodi välja iga mudeli iga alglähendite komplekti puhul. Tulemusi võrreldi nii erinevate alglähendite komplektide vahel kui ka mudelitevaheliselt. Lisaks võrreldi neid artikli Kuljus *et al.* (2018) tulemustega, kus on samale probleemile lähenetud teisi varjatud tunnustega mudeleid kasutades.

Leitud hinnangud käitusid erinevate alglähendite komplektide korral üsna stabiilselt. Mudeleid omavahel võrreldes selgus, et M5 andis kõige kehvema tulemuse. M8 ja M10 seast üht teisele eelistada oli raske, kuid M8 näis keskmise absoluutvea põhjal andvat veidi parema tulemuse. Võrreldes artiklis Kuljus *et al.* (2018) kasutatud mudelite sooritusega oli tingliku sõltumatuse mudel üldjoontes kas samaväärne või veidi kehvem. Positiivsena paistis aga välja, et antud näites tuvastas tingliku sõltumatuse mudel KT seisundeid kogu väärtuste piirkonna ulatuses ühtlasemalt kui teised mudelid.

Mudeli peamiseks probleemiks oli luu ja õhu eristamine ning parameetrite hindamine luu piirkonnas, kus vaatluseid oli vähem ning nende jaotus sarnanes pigem ühtlase jaotusega kui neile sobitatud normaaljaotusega. Probleem kajastus selgelt mudelite jääkide joonistel ning oli visuaalselt näha ka mõne konkreetse substituut-KT-pildi horisontaalse lõigu korral. Kuna MRT-piltide põhjal üksi ongi raske luule vastavaid kudede klasse õhust eristada, siis ilmselt pole juhendamiseta õppel baseeruvad mudelid siinkohal piisavad. Tasuks proovida mudeleid, kus anname ise õhu ja luu kohta lisainformatsiooni ette.

Üldiselt oli sellegipoolest näha, et mudelil on juhendamisega õppe või mitme režiimiga varjatud Markovi mudeli korral potentsiaali saavutada häid substituut-KT-piltide hinnanguid. Sama mudelit oleks huvitav rakendada mõnele teisele anatoomiliselt lihtsamale keha piirkonnale, kus kudede klasse on lihtsam eristada. Ilmselt oleks mudeli sooritus sellises olukorras parem ning võib-olla annaks tulemusi ka reaalselt ära kasutada.

Kasutatud kirjandus

Bibb, R., Eggbeer, D. & Paterson, A. (2015). *Medical Modelling (Second Edition)*. Oxford: Woodhead Publishing.

Bilmes, J. (1998). *A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models.* TR-97-021. Berkeley: International Computer Science Institute.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* New York: Springer.

Izenman, A. J. (2008). *Modern multivariate statistical techniques: regression, classification, and manifold learning.* New York: Springer.

Kuljus, K. *et al.* (2018). Comparison of hidden Markov chain models and hidden Markov random field models in estimation of computed tomography images. *Communications in Statistics: Case Studies, Data Analysis and Applications.* **4**(1), 46-55.

Naidich, T. P. *et al.* (2012). *Imaging of the Brain: Expert Radiology Series.* Philadelphia: Elsevier Health Sciences.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE.* **77**(2), 257-286.

Lisad

Lisa 1. Edasi-tagasi-tõenäosuste ja EM-algoritmi programmikood

```
# funktsioon edasi−tõenäosuste arvutamiseks
forwsc = function(pm, pi0, emissions){
  K = nrow(pm)
  T = nrow(emissions)
  if (T==1) emissions = matrix(emissions, nrow=1)
  phi = matrix(pi0, nrow = 1)
  alpha = matrix(rep(NA, K * T), nrow = T)
  scale = rep(0,T)
  for (i in 1:T) {
    if (i > 1) phi = phi %*% pm
    phi = phi %*% diag(emissions[i,])
    sumphi = sum(phi)
    phi = phi/sumphi
    scale[i] = sumphi
    alpha[i, ] = phi
  }
  answer = list(alpha=alpha, scale=scale)
  return(answer)
}
# funktsioon tagasi−tõenäosuste arvutamiseks
backwsc = function(pm, emissions, lastrow, scale){
  K = nrow(pm)
  T = nrow(emissions)
  beta = matrix(rep(NA, K * T), nrow = T)
  beta[T, ] = lastrow
  phi = matrix(lastrow, ncol = 1)
  for (i in seq(T − 1, 1, −1)) {
```

```
    phi = pm %*% diag(emissions[i+1,]) %*% phi
    phi = phi/scale[i+1]
    beta[i,] = phi
  }
  return(beta)
}
# funktsioon parameetrite uuendamiseks ühe jada jaoks
sequenceEM =
    function(pi0,pm,cm,fmatr,gmatr,fgmatr,mri,ct,pikkus) {
  K = nrow(pm)
  M = ncol(cm)
  edasitn = forwsc(pm,pi0,fgmatr)
  alpha = edasitn$alpha; scalevec = edasitn$scale
  beta = backwsc(pm,fgmatr,rep(1,K),scalevec)
  gammaz = matrix(alpha*beta,nrow=pikkus)
  gammazr = vector("list",K)
  gammazrT = matrix(rep(0,K*M),nrow=K)
  for(i in 1:K) {
    gi = matrix(gammaz[,i],ncol=1); ci = matrix(cm[i,],nrow=1)
    nim = c(1/gmatr%*%t(cm)[,i])
    gammazr[[i]] = (gi%*%ci)*gmatr*nim
    gammazrT[i,] = colSums(gammazr[[i]])
  }
  gammar = Reduce('+',gammazr)

  ksilist = vector("list",pikkus-1)
  for (t in 1:(pikkus-1)) {
    fg = matrix(fgmatr[t+1,],ncol=1)
    ksilist[[t]] = (1/scalevec[t+1])*
                t((fg%*%alpha[t,])*c(beta[t+1,]))*pm
  }
  ksisum = Reduce('+',ksilist)
  if (pikkus==2) {
```

```
    gammazNpup = gammaz[1,]
  } else gammazNpup = colSums(gammaz[-pikkus,])

  ctsum = rep(0,M); ct2sum = rep(0,M)
  for (t in 1:pikkus) {
    ctsum = ctsum+ct[t]*gammar[t,];
    ct2sum = ct2sum+(ct[t])^2*gammar[t,]
  }
  mrisum = vector("list",K); mrisum2 = vector("list",K);
  for (k in 1:K) {
    mrisum[[k]] = rep(0,mriarv)
    mrisum2[[k]] = matrix(0,nrow=mriarv,ncol=mriarv)
    for (t in 1:pikkus) {
      mrisum[[k]] = mrisum[[k]]+gammaz[t,k]*mri[t,]
      mrisum2[[k]] =
          mrisum2[[k]]+gammaz[t,k]*(mri[t,]%*%t(mri[t,]))
    }
  }
  loglikup = sum(log(scalevec))
  answer = list(piNup=gammaz[1,],ksisum=ksisum,
                gammazNpup=gammazNpup,gammazrNup=gammazrT,
                gammazNup=colSums(gammaz),gammarNup=colSums(gammar),
                ctsum=ctsum,ct2sum=ct2sum,mrisum=mrisum,
                mrisum2=mrisum2,loglikup=loglikup)
  return(answer)
}
# funktsioon ting. sõlt. mudeli rakendamiseks N jada jaoks
CondIndEM = function(pi0,pm,cm,meanct,varct,meanmri,covmri,
                     obslist,mriarv,convcrit,maxiter) {
  N = length(obslist);
  iter = 0; logliknew = 1; goon = TRUE;
  K = nrow(pm); M = ncol(cm); loglikvec = rep(0,maxiter)
  while (goon==TRUE) {
```

```
piN = rep(0,K); ksiN = matrix(0,nrow=K,ncol=K)
gammazNp = rep(0,K); gammazrN = matrix(rep(0,K*M),nrow=K)
    gammazN = rep(0,K); gammarN = rep(0,M)
meanctN = rep(0,M); varctN = rep(0,M)
meanmriN = vector("list",K); covmriN = vector("list",K)

for (k in 1:K) {
  meanmriN[[k]] = rep(0,mriarv)
  covmriN[[k]] = matrix(0,nrow=mriarv,ncol=mriarv)
}
loglikN = 0
for (e in 1:N) {
  pikkus = nrow(obslist[[e]])
  mri = obslist[[e]][,2:(mriarv+1)]
  ct = obslist[[e]][,1]
  fmatr = matrix(0,nrow=pikkus,ncol=K)
  for (k in 1:K) {
    fmatr[,k] = dmnorm(mri,mean=meanmri[[k]],
                       varcov=covmri[[k]],log=FALSE)
  }
  gmatr = matrix(0,nrow=pikkus,ncol=M)
  for (l in 1:M) {
    gmatr[,l] = dnorm(ct,mean=meanct[l],
                      sd=sqrt(varct[l]),log=FALSE)
  }
  fgmatr = fmatr*(gmatr%*%t(cm))
  if (pikkus>1) {
    uplist = sequenceEM(pi0,pm,cm,fmatr,gmatr,
                        fgmatr,mri,ct,pikkus)
  }
  piN = piN+uplist$piNup; ksiN = ksiN+uplist$ksisum
  gammazNp = gammazNp+uplist$gammazNpup
  gammazrN = gammazrN+uplist$gammazrNup
```

```
      gammazN = gammazN+uplist$gammazNup
      gammarN = gammarN+uplist$gammarNup
      meanctN = meanctN+uplist$ctsum
      varctN = varctN+uplist$ct2sum
      for (k in 1:K) {
        meanmriN[[k]] = meanmriN[[k]]+ uplist$mrisum[[k]]
        covmriN[[k]] = covmriN[[k]]+ uplist$mrisum2[[k]]
          }
        loglikN = loglikN+uplist$loglikup
    }
    pi0 = piN/N;
    for (k in 1:K) pm[k,] = ksiN[k,]/gammazNp[k]
    for (k in 1:K) cm[k,] = gammazrN[k,]/gammazN[k]
    for (l in 1:M) {
      meanct[l] = meanctN[l]/gammarN[l]
      varct[l] = varctN[l]/gammarN[l]-(meanct[l])^2
    }
    for (k in 1:K){
      meanmri[[k]] = meanmriN[[k]]/gammazN[k]
      covmri[[k]] = covmriN[[k]]/gammazN[k]-
                              meanmri[[k]]%*%t(meanmri[[k]])
    }
    loglikold = logliknew; logliknew = loglikN
        iter = iter+1; loglikvec[iter] = loglikN
    relcrit = abs((logliknew-loglikold)/loglikold)
    goon = (relcrit>convcrit)&(iter<maxiter)
  }
  vastus = list(pi0=pi0,pm=pm,cm=cm,meanct=meanct,
                            varct=varct,meanmri=meanmri,
                            covmri=covmri,iter=iter,
                            loglikvec=loglikvec[1:iter])
  return(vastus)
}
```

Lisa 2. Näite programmikood

```
#library(markovchain)
#library(mnormt)

K = 2 # seisundite arv
M = 3 # segumudeli komponentide arv
mriarv = 2 # MRT piltide arv

# algtõenäosused
p0 = c(0.5,0.5)
# üleminekumaatriks
pm = matrix(c(0.7,0.3,0.2,0.8), byrow=TRUE, nrow=K)
mc = as(pm, "markovchain"); names(mc) = c("1","2")

# MRT keskväärtused kahe jada ja kahe seisundi jaoks
meanmri1 = c(1,2); meanmri2 = c(4,6)
meanmri = list(meanmri1,meanmri2)

# MRT kov maatriksid kahe jada ja kahe seisundi jaoks
covmri1 = matrix(c(1,0.4,0.4,1),nrow=2,byrow=TRUE)
covmri2 = matrix(c(2,1,1,1.5),nrow=2,byrow=TRUE)
covmri = list(covmri1,covmri2)

# KT keskväärtused ja standardhälbed kolme komponendi jaoks
meanct = c(1,3,5)
sdct = c(1,2,2.5)
# kaalud
c1 = c(0.7,0.1,0.2)
c2 = c(0.1,0.6,0.3)
c = list(c1,c2)

N = 50 # sõltumatute jadade arv
```

```
set.seed(0)
T = sample(500:1000,N)
X = matrix(rep(0,(mriarv+1)*sum(T)),ncol=mriarv+1)
Y = matrix(rep(0,2*sum(T)),ncol=2)
vaatlused = vector("list",N)
seisundid = vector("list",N)
for(n in 1:N){
  pikkus = T[n]
  t0 = sample(names(mc),1,prob=p0)
  # varjatud ahel
  Z = rmarkovchain(n = pikkus, object = mc, t0 = t0)
  seisundid[[n]] = Z
  if(n==1) i = 0 else i = sum(T[1:(n-1)])
  for (t in 1:pikkus){
    k = as.numeric(Z[t])
        # MRT vaatluste ahel
    X[i+t,] = c(rmnorm(1,mean = meanmri[[k]],
                            varcov = covmri[[k]]),n)
    l = sample(1:3,prob=c[[k]],size=1)
        # KT vaatluste ahel
    Y[i+t,] = c(rnorm(n=1,mean=meanct[l],sd=sdct[l]),n)
  }
  vaatlused[[n]] = cbind(Y[(i+1):(i+pikkus),1],
                         X[(i+1):(i+pikkus),1:mriarv])
}

#####################################################
# alglähendite leidmine
kx = kmeans(X[,1:mriarv],2,iter.max=50,nstart=10)
ky = kmeans(Y[,1],3,iter.max=50,nstart=10)

meanmri1_alg = kx$centers[1,]
meanmri2_alg = kx$centers[2,]
```

```
meanmri_alg = list(meanmri1_alg, meanmri2_alg)
covmri1_alg = cov(X[kx$cluster==1,1:mriarv])
covmri2_alg = cov(X[kx$cluster==2,1:mriarv])
covmri_alg = list(covmri1_alg, covmri2_alg)

jrk = c(2,3,1)
meanct_alg = ky$centers[jrk]; meanct_alg
varct_alg =  c(var(Y[ky$cluster==jrk[1],1]),
               var(Y[ky$cluster==jrk[2],1]),
               var(Y[ky$cluster==jrk[3],1]))
cm_alg = matrix(prop.table(table(kx$cluster,ky$cluster),1)
                                                  ,nrow=2)[,jrk]
üleminekud = table(kx$cluster[1:sum(T)-1]-kx$cluster[2:sum(T)])
p12 = üleminekud["-1"]/sum(kx$cluster[-sum(T)]==1)
p21 = üleminekud["1"]/sum(kx$cluster[-sum(T)]==2)
pm_alg = matrix(c(1-p12,p12,p21,1-p21), byrow=TRUE, ncol=K)
p0_alg = prop.table(table(kx$cluster[c(1,cumsum(T[1:(N-1)])+1)]))

convcrit = 0.0001
maxiter = 20
res = CondIndEM(pi0=p0_alg,pm=pm_alg,cm=cm_alg,meanct=meanct_alg,
                varct=varct_alg,meanmri=meanmri_alg,covmri=covmri_alg,
                obslist=vaatlused,mriarv=mriarv,convcrit,maxiter)
```

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Lisette Pajula,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Kompuutertomograafia piltide hindamine magnetresonantstomograafia piltide põhjal varjatud Markovi mudeli abil“, mille juhendaja on Kristi Kuljus, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commonsi litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.

3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.

4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Lisette Pajula

25.05.2021