

TARTU ÜLIKOOL
MATEMAATIKA-INFORMAATIKATEADUSKOND
MATEMAATILISE STATISTIKA INSTITUUT

Mark Gimbutas

Varjatud Markovi välja mudelid puudel

Magistritöö matemaatilise statistika erialal (30 EAP)

Juhendaja: Jüri Lember

Tartu 2015

Varjatud Markovi välja mudelid puudel

Varjatud Markovi välja mudel on statistiline mudel, millel on kaks komponenti: vaatlused ja vaatluseid genereeriv juhuslik protsess, mis on tegeliku huvi objekt, aga mis ise ei ole vaadeldav. Kui see varjatud juhuslik protsess on Markovi ahel, siis tulemus on varjatud Markovi mudel, mida on ohtralt uuritud ja rakendatud. Andmete struktuur on tihti keerulisem kui jada – näiteks pikslid moodustavad pildil võre. Puudel on aga eelis üldisemate graafide ees, sest puukujulise varjatud kihi korral on võimalik mitmeid vajalikke arvutusi teha kiiresti ja täpselt.

Käesoleva töö eesmärk on varjatud Markovi mudelite segmenteerimisteooria üldistamine teistele varjatud Markovi välja mudelitele, eelkõige puudel defineeritud mudelitele. Töö esimeses pooles on esitatud ülevaade Markovi väljadest ja defineeritud varjatud Markovi välja mudel. Töö teises pooles on esitatud varjatud Markovi mudelite hübriidriski, vastavaid parimaid joondusi arvutava algoritmi ja k -bloki riski üldistused puudele. Hübriidriski väärtus seisneb konkreetse headuse kriteeriumi paikaseadmises, mis võimaldab ühendada keskmiselt kõige vähem vigu tegev klassifitseerija ja kõige suurema tõepäraga tulemusi andev klassifitseerija ühtsesse, heade omadustega klassifitseerijate peresse, mille seast on võimalik valida sobiv kompromiss tõepära ja täpsuse vahel. Samuti on esitatud Baum-Welchi algoritmi vahetu üldistus puudele, millega saab hinnata mudeli parameetreid. Hübriidalgoritmi töö tulemusi ja nende täpsust on illustreeritud simulatsioonidega.

Märksõnad: *Markovi ahelad, varitunnuste analüüs*

Hidden Markov tree models

Hidden Markov random field model is a statistical model that has two components: observations and a random process of actual interest which generates observations, yet is unobservable. If the hidden random process is a Markov chain, then we have a hidden Markov model, which is extensively studied and applied. The structure of the data is often more complex than just a sequence – e.g., pixels of an image form a lattice. Trees have an advantage of computational kind over more general graphs, since many necessary calculations are fast and accurate, if the hidden layer has a tree structure.

Hereby we aim to generalise the theory of segmentation of hidden Markov models to other hidden Markov random field models, especially hidden Markov tree models. An overview of hidden Markov random fields is presented and hidden Markov random field model is defined in the first half of the thesis. In the second half, combined risks, an algorithm to implement corresponding decoders and k -block risks of hidden Markov models are generalised to trees. Combined risk sets a criterion of goodness that allows uniting the most accurate classifier and the classifier that returns results of highest likelihood to form a single family of hybrid decoders with nice properties, from which the desired compromise between accuracy and likelihood can be chosen. A straightforward generalisation of Baum-Welch algorithm is also presented to estimate model parameters. Hybrid decoders are illustrated by simulations.

Key words: *Markov chains, latent variable analysis*

Sisukord

Sissejuhatus	4
1 Markovi väli	5
1.1 Mõiste	5
1.2 Gibbsi jaotus	9
1.3 Puukujuline sõltuvusstruktuur	15
1.3.1 Tähistused	16
1.3.2 Tulemused	16
2 Varjatud Markovi välja mudel	22
3 Segmenteerimine ja risk	27
3.1 Segmenteerimisülesande tutvustus	27
3.1.1 Klassifitseerimisteooria põhimõisteid	29
3.2 Erinevaid kaofunktsioone ja riske varjatud Markovi välja mudelitele .	30
4 Hübriidrisk puudel	33
4.1 Hübriidalgoritm	33
4.2 k -bloki risk ja selle üldistused	36
4.2.1 k -blokk kui tipp koos järglastega	37
4.2.2 k -blokk kui tipp koos esivanematega	41
4.2.3 Hübriidriski tõlgendus	44
4.3 Simulatsioonid	46
5 Parameetrite hindamine	51
5.1 Baum-Welchi algoritm	51
5.2 Simulatsioonid	52
Viited	57
Lisad	59
Lisa 1. Hübriidalgoritm ja funktsioonid silumistõenäosuste arvutamiseks .	59
Lisa 2. Funktsioonid simuleerimiseks	61
Lisa 3. Funktsioonid parameetrite hindamiseks	63

Sissejuhatus

Varjatud Markovi välja mudel on statistiline mudel, millel on kaks komponenti: vaatlused ja vaatluseid genereeriv juhuslik protsess. Tegelik huvi objekt on see viimane, mis aga ise ei ole vaadeldav. Kui varjatud juhuslik protsess on Markovi *ahel*, siis tulemus on varjatud Markovi mudel. Varjatud Markovi mudeleid on ohtralt uuritud ja rakendatud, näiteks (automaatses) kõnetuvastuses heli põhjal sõnade ära arvamiseks ja molekulaarbioloogias DNA järjestuse põhjal erinevate funktsionaalsete piirkondade leidmiseks. Andmete struktuur võib aga olla keerulisem kui lihtne järjestus – näiteks fülogeneetilistes uuringutes paiknevad vaatlused fülogeneesipuu tippudes ja piltidega tegeledes moodustavad vaatlused (pikslid) lausa *võre*. Puud on siiski ahvatlevamad kui üldised graafid, sest puukujulise sõltuvusstruktuuriga andmete korral on võimalik mitmeid vajalikke arvutusi teha kiiresti ja täpselt. Seetõttu võib õnnestuda puudega edukalt lähendada ka keerulisematel graafidel antud mudeleid.

Peamised rakendused varjatud Markovi (välja) mudelitele seisnevad varjatud kihi oleku ära arvamises või mudeli parameetrite hindamises. Ülesande lihtsustamiseks eeldataksegi, et varjatud kiht on Markovi omadusega – piiri fikseerimisel on ühel pool piiri toimuv tegevus sõltumatu teisel pool piiri toimuvast. Selgub, et selle idee täpseks väljendamiseks on mitmeid veidi erinevaid võimalusi. Samuti on mitmeid erinevaid võimalusi varjatud kihi oleku prognoosimiseks. Näiteks kuuleme lauset poolvõõras keeles, ja üritame aru saada, mida öeldi. Üks äärmuslik meetod on iga üksiku hääliku juures mõistatada, et mis täht see võis olla, ja seejärel koostada lause nõnda välja pakutud tähtedest. Arvatavasti saame paljudele tähtedele pihta, kuid välja võib tulla ka vale ortograafiaga sõnu või hoopis selliseid, mida pole olemas. Teine äärmus on korruga meele hoida kogu kuulnud lauset ja lasta peas ükshaaval läbi kõiki võimalikke korrektseid lauseid selles keeles, proovides aru saada, mida kõige tõenäolisemalt võidi öelda. Tulemus on (meie keeleteadmiste piires) korrektne lause, kuid võime mõne haruldasema väljendi segi ajada klišeega. Nende kahe äärmuse vahel on kompromiss: proovida aru saada sõnadest ehk tähtede *blokkidest*. Analoogiline idee on varjatud Markovi mudelite jaoks formuleeritud, üldistame selle ka puudel antud varjatud Markovi välja mudelitele.

Käesoleva töö eesmärk on varjatud Markovi mudelite segmenteerimisteooria üldistamine teistele varjatud Markovi välja mudelitele, eelkõige puudel defineeritud mudelitele.

Esimeses peatükis antakse ülevaade Markovi väljadest ja teistest lähedastest struktuuridest ning asjassepuutuvatest omadustest. Põhitähelepanu on suunatud puudel defineeritud Markovi väljadele. Teises peatükis defineeritakse varjatud Markovi välja mudel nõnda, et see oleks varjatud Markovi mudeli üldistus. Kolmandas peatükis tuuakse sisse põhimõisteid tehisõppe valdkonnast ja defineeritakse üldiseid riskifunktsioone varjatud Markovi väljade segmenteerimiseks. Neljandas peatükis üldistatakse varjatud Markovi mudelite kontekstis välja töötatud hübriidklassifitseerijate pere puudele ja esitatakse vastavaid parimaid joonduseid arvutav algoritm. Kuigi varjatud Markovi mudelite korral on hübriidklassifitseerijate tõlgendamiseks üks võtmeidee, siis puude jaoks on see üldistuv mitmes suunas. Viimasel peatükis esitatakse varjatud Markovi mudeli parameetrite hindamiseks kasutatava Baum-Welchi algoritmi vahetu üldistus puudel antud mudelitele.

1 Markovi väli

1.1 Mõiste

Kui eesmärgiks on uurida juhuslikku vektorit, aga ei ole oskusi, tahtmist või viitsimist võtta arvesse kõiki võimalikke sõltuvusstruktuure vektori komponentide vahel, tuleb teha lihtsustusi. Võime ette kujutada, et vektori komponentideks olevad juhuslikud suurused paiknevad ruumis näiteks nagu pikslid ekraanil. Sel juhul saame ette kujutada iga piksliga külgnevaid pikseleid ehk tema naabreid. Lihtsustus, millega käesolevas peatükis tutvume, seisneb sisuliselt järgnevas: piksli värv sõltub vahetult vaid naaberpikslite värvist, mitte ülejäänud ekraani seisundist.

Täpsemalt, olgu I lõplik indeksite hulk, milles on n elementi. Kui pole teisiti öeldud, siis $I := \{1, 2, \dots, n\}$.

Definitsioon 1 (Bremaud, 1999 lk 254). Hulga I osahulkade süsteemi $N = \{\mathcal{N}_u\}_{u \in I}$ nimetame *naabruskondade süsteemiks* hulgal I , kui iga $u \in I$ korral kehtib

1. $u \notin \mathcal{N}_u$ ja
2. $u \in \mathcal{N}_v \Rightarrow v \in \mathcal{N}_u$.

Olukorras $u \in \mathcal{N}_v$ ütleme, et element u on naaber elemendile v .

Naabruskondade süsteemi hulgal I saame kujutada graafina, kus tippude $u, v \in I$ vahel on serv parajasti siis, kui need tipud on vastastikused naabrid (eelneva definitsiooni teise nõude tõttu on naabrussuhe vastastikune). Sellest inspireerituna suhtume paari (I, N) kui graafi tippude hulgaga I ja servade hulgaga, mille määrab naabruskondade süsteem N .

Olgu $\bar{\mathbf{S}} := (S_u)_{u \in I}$ juhuslik vektor, $S_u, u \in I$ olgu *diskreetsed* juhuslikud suurused, millel on *lõplik hulk* võimalikke väärtuseid, mis sisalduvad hulgas $K := \{1, 2, \dots, k\}$. Hulka K nimetame vahel *seisundite ruumiks* ja tema elemente *seisunditeks*. Iga mittetühja $J \subset I$ korral tähistame $\bar{\mathbf{S}}_J := (S_u)_{u \in J}$, juhusliku vektori $\bar{\mathbf{S}}_J$ mingit realisatsiooni tähistame $\bar{\mathbf{s}}_J$ ja nimetame (*seisundite*) *konfiguratsiooniks*.

Definitsioon 2 (Bremaud, 1999 lk 254 järgi). Juhuslikku vektorit $\bar{\mathbf{S}} = (S_u)_{u \in I}$ nimetatakse (*juhuslikuks*) *Markovi väljaks* naabruskondade süsteemi N suhtes või graafil (I, N) , kui iga $u \in I$ korral kehtib

$$\mathbf{P}(S_u = s_u \mid \bar{\mathbf{S}}_{I \setminus \{u\}} = \bar{\mathbf{s}}_{I \setminus \{u\}}) = \mathbf{P}(S_u = s_u \mid \bar{\mathbf{S}}_{\mathcal{N}_u} = \bar{\mathbf{s}}_{\mathcal{N}_u}),$$

kui vaid need tinglikud tõenäosused on defineeritud.

Anneme kohe ka alternatiivse definitsiooni läbi tingliku sõltumatuse mõiste.

Definitsioon 3 (Bremaud, 1999 lk 11 järgi). Olgu X, Y ja Z diskreetsed juhuslikud vektorid, mis võtavad väärtusi vastavalt hulkades \mathcal{X}, \mathcal{Y} ja \mathcal{Z} . Ütleme, et juhuslikud vektorid X ja Y on tinglikud sõltumatud, kui Z on fikseeritud, kirjutame $X \perp Y \mid Z$, kui iga $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ korral kehtib

$$\mathbf{P}(X = x, Y = y \mid Z = z) = \mathbf{P}(X = x \mid Z = z)\mathbf{P}(Y = y \mid Z = z),$$

kui vaid $\mathbf{P}(Z = z) > 0$.

Väide 1 (Bremaud, 1999 lk 12 järgi). Olgu X, Y ja Z diskreetsed juhuslikud vektorid, mis võtavad väärtuseid hulkades vastavalt \mathcal{X}, \mathcal{Y} ja \mathcal{Z} . Olgu $\mathcal{Y}^* := \{y \in \mathcal{Y} : \mathbf{P}(Y = y) > 0\}$. Kui leidub mingi funktsioon $g : \mathcal{X} \times \mathcal{Y}^* \rightarrow [0, 1]$, nii et iga $(x, y, z) \in \mathcal{X} \times \mathcal{Y}^* \times \mathcal{Z}$ korral

$$\mathbf{P}(X = x \mid Z = z, Y = y) = g(x, y),$$

kui vaid $\mathbf{P}(Z = z, Y = y) > 0$, siis iga $(x, y) \in \mathcal{X} \times \mathcal{Y}^*$ korral $g(x, y) = \mathbf{P}(X = x \mid Y = y)$ ja juhuslikud vektorid X ja Z on tinglikud sõltumatud, tingimusel et Y on fikseeritud.

Tõestus. Tõestuse leiab raamatust Bremaud (1999) lk 12. \square

Väide 2. Olgu $\bar{\mathbf{S}} = (S_u)_{u \in I}$ diskreetne juhuslik vektor, mis võtab väärtusi hulgas K^n . Juhuslik vektor $\bar{\mathbf{S}}$ on Markovi väli graafil (I, N) parajasti siis, kui iga $u \in I$ korral $S_u \perp \bar{\mathbf{S}}_{(I \setminus \mathcal{N}_u) \setminus u} \mid \bar{\mathbf{S}}_{\mathcal{N}_u}$.

Tõestus. Otsene järeldus väitest 1, kui iga u korral võtta $X = S_u, Y = \bar{\mathbf{S}}_{\mathcal{N}_u}$ ja $Z = \bar{\mathbf{S}}_{(I \setminus \mathcal{N}_u) \setminus u}$. \square

Näeme, et suvaline diskreetne juhuslik vektor on Markovi väli sellise naabruskondade süsteemi suhtes, mis määrab iga indeksi naabriteks kõik ülejäänud indeksid. Huvi pakuvad pigem väikeste naabruskondadega Markovi väljad.

Näide 1 (Markovi ahel kui Markovi väli). Olgu $\{X_i\}_{i \in I}$ diskreetse ajaga Markovi ahel ehk

$$\mathbf{P}(X_i = k_i \mid X_1 = k_1, X_2 = k_2, \dots, X_{i-1} = k_{i-1}) = \mathbf{P}(X_i = k_i \mid X_{i-1} = k_{i-1}), \quad i \in I \setminus \{1\}.$$

Definitsioon ütleb otsesõnu, et käesolev seisund sõltub kogu minevikust vaid eelneva seisundi kaudu. Midagi sarnast oleks vaja saada ka tuleviku kohta.

Paneme tähele, et kehtib

$$\mathbf{P}(X_{i+1} = k_{i+1} \mid X_i = k_i, X_{i-1} = k_{i-1}) = \mathbf{P}(X_{i+1} = k_{i+1} \mid X_i = k_i),$$

kui $i \in \{2, 3, \dots, n-1\}$ (formaalne tõestus allpool, väide 6). Järelikult

$$\begin{aligned} & \mathbf{P}(X_{i+1} = k_{i+1} \mid X_i = k_i) \mathbf{P}(X_i = k_i \mid X_{i-1} = k_{i-1}) \\ &= \mathbf{P}(X_{i+1} = k_{i+1} \mid X_i = k_i, X_{i-1} = k_{i-1}) \mathbf{P}(X_i = k_i \mid X_{i-1} = k_{i-1}) \\ &= \mathbf{P}(X_i = k_i, X_{i+1} = k_{i+1} \mid X_{i-1} = k_{i-1}). \end{aligned} \quad (1)$$

Lühiduse mõttes tähistame juhuslikku vektorit $(X_i, X_{i+1}, \dots, X_j) =: X_i^j$ ja seisundite vektorit $(k_i, k_{i+1}, \dots, k_j) =: k_i^j$, kui $i \leq j$. Paneme tähele, et ketireegli ja Markovi ahela definitsiooni tõttu kehtib iga $i \in I \setminus \{1\}$ korral

$$\begin{aligned} \mathbf{P}(X_i^n = k_i^n \mid X_1^{i-1} = k_1^{i-1}) &= \prod_{j=i}^n \mathbf{P}(X_j = k_j \mid X_1^{j-1} = k_1^{j-1}) \\ &= \prod_{j=i}^n \mathbf{P}(X_j = k_j \mid X_{j-1} = k_{j-1}). \end{aligned} \quad (2)$$

Hulgas I valime välja osahulgad $\mathcal{N}_1 := \{2\}, \mathcal{N}_n := \{n-1\}$ ja iga $i \in \{2, 3, \dots, n-1\}$ korral $\mathcal{N}_i := \{i-1, i+1\}$. Näha on, et $N := \{\mathcal{N}_i\}_{i \in I}$ moodustab naabruskondade süsteemi hulgal I .

Kontrollime nüüd Markovi välja definitsiooni 2 nõude täidetust.

Kui $i \in \{2, 3, \dots, n-1\}$, siis

$$\begin{aligned}
& \mathbf{P}(X_i = k_i \mid X_1^{i-1} = k_1^{i-1}, X_{i+1}^n = k_{i+1}^n) \\
&= \frac{\mathbf{P}(X_i = k_i, X_{i+1}^n = k_{i+1}^n \mid X_1^{i-1} = k_1^{i-1})}{\sum_l \mathbf{P}(X_i = l, X_{i+1}^n = k_{i+1}^n \mid X_1^{i-1} = k_1^{i-1})} \\
&= \frac{\prod_{j=i}^n \mathbf{P}(X_j = k_j \mid X_{j-1} = k_{j-1})}{\sum_l \mathbf{P}(X_i = l \mid X_{i-1} = k_{i-1}) \mathbf{P}(X_{i+1} = k_{i+1} \mid X_i = l) \prod_{j=i+2}^n \mathbf{P}(X_j = k_j \mid X_{j-1} = k_{j-1})} \\
&= \frac{\mathbf{P}(X_i = k_i \mid X_{i-1} = k_{i-1}) \mathbf{P}(X_{i+1} = k_{i+1} \mid X_i = k_i)}{\sum_l \mathbf{P}(X_i = l \mid X_{i-1} = k_{i-1}) \mathbf{P}(X_{i+1} = k_{i+1} \mid X_i = l)} \\
&= \frac{\mathbf{P}(X_{i+1} = k_{i+1}, X_i = k_i \mid X_{i-1} = k_{i-1})}{\sum_l \mathbf{P}(X_{i+1} = k_{i+1}, X_i = l \mid X_{i-1} = k_{i-1})} \\
&= \frac{\mathbf{P}(X_{i+1} = k_{i+1}, X_i = k_i \mid X_{i-1} = k_{i-1})}{\mathbf{P}(X_{i+1} = k_{i+1} \mid X_{i-1} = k_{i-1})} \\
&= \mathbf{P}(X_i = k_i \mid X_{i+1} = k_{i+1}, X_{i-1} = k_{i-1}).
\end{aligned}$$

Teine võrdus tuleneb võrdusest 2 ja neljas tuleneb võrdusest 1.

Kui $i = 1$, siis

$$\begin{aligned}
\mathbf{P}(X_1 = k_1 \mid X_2^n = k_2^n) &= \frac{\mathbf{P}(X_2^n = k_2^n \mid X_1 = k_1) \mathbf{P}(X_1 = k_1)}{\mathbf{P}(X_2^n = k_2^n)} \\
&= \frac{\prod_{j=2}^n \mathbf{P}(X_j = k_j \mid X_{j-1} = k_{j-1}) \mathbf{P}(X_1 = k_1)}{\prod_{j=3}^n \mathbf{P}(X_j = k_j \mid X_{j-1} = k_{j-1}) \mathbf{P}(X_2 = k_2)} \\
&= \frac{\mathbf{P}(X_2 = k_2 \mid X_1 = k_1) \mathbf{P}(X_1 = k_1)}{\mathbf{P}(X_2 = k_2)} \\
&= \mathbf{P}(X_1 = k_1 \mid X_2 = k_2).
\end{aligned}$$

Esimene ja viimane võrdus tulevad Bayesi valemist, teine võrdusest 2.

Kui $i = n$, siis Markovi välja definitsiooni nõue on täidetud otse Markovi ahela definitsiooni tõttu. Järelikult Markovi ahel on Markovi väli ahelakujulise naabruskondade süsteemi suhtes.

Üldiselt vastupidine ei kehti: ahelakujulisel graafil defineeritud Markovi väli ei tarvitse olla Markovi ahel.

Näide 2 (Ahelal defineeritud Markovi väli, mis pole Markovi ahel, Zachary, 1981 lk 35). Olgu $I = \{1, 2, 3, 4, 5\}$, $K = \{0, 1\}$, ja $\bar{\mathbf{S}} = (S_u)_{u \in I}$ juhuslik vektor, mis võtab väärtusi hulgas K^5 . Määrame naabriteks sellised hulga I elemendid, mille vahe absoluutväärtus on 1 – nii saamegi ahela 1–2–3–4–5. Olgu vektori $\bar{\mathbf{S}}$ jaotus selline, et $\mathbf{P}(\bar{\mathbf{S}} = (0, 0, 0, 0, 0)) = \mathbf{P}(\bar{\mathbf{S}} = (1, 1, 0, 1, 1)) = 0.5$ ja kõigi ülejäänud seisundite tõenäosused olgu nullid. Veendume, et $\bar{\mathbf{S}}$ on Markovi väli. Selleks rehkendame välja vajalikud tinglikud tõenäosused, kui vaid need on defineeritud.

$$\begin{aligned}
1 &= \mathbf{P}(S_1 = 0 \mid S_2 = 0, S_3 = 0, S_4 = 0, S_5 = 0) = \mathbf{P}(S_1 = 0 \mid S_2 = 0), \\
0 &= \mathbf{P}(S_1 = 0 \mid S_2 = 1, S_3 = 0, S_4 = 1, S_5 = 1) = \mathbf{P}(S_1 = 0 \mid S_2 = 1), \\
1 &= \mathbf{P}(S_2 = 0 \mid S_1 = 0, S_3 = 0, S_4 = 0, S_5 = 0) = \mathbf{P}(S_2 = 0 \mid S_1 = 0, S_3 = 0), \\
0 &= \mathbf{P}(S_2 = 0 \mid S_1 = 1, S_3 = 0, S_4 = 1, S_5 = 1) = \mathbf{P}(S_2 = 0 \mid S_1 = 1, S_3 = 0), \\
1 &= \mathbf{P}(S_3 = 0 \mid S_1 = 0, S_2 = 0, S_4 = 0, S_5 = 0) = \mathbf{P}(S_3 = 0 \mid S_2 = 0, S_4 = 0), \\
1 &= \mathbf{P}(S_3 = 0 \mid S_1 = 1, S_2 = 1, S_4 = 1, S_5 = 1) = \mathbf{P}(S_3 = 0 \mid S_2 = 1, S_4 = 1), \\
1 &= \mathbf{P}(S_4 = 0 \mid S_1 = 0, S_2 = 0, S_3 = 0, S_5 = 0) = \mathbf{P}(S_4 = 0 \mid S_3 = 0, S_5 = 0), \\
0 &= \mathbf{P}(S_4 = 0 \mid S_1 = 1, S_2 = 1, S_3 = 0, S_5 = 1) = \mathbf{P}(S_4 = 0 \mid S_3 = 1, S_5 = 1), \\
1 &= \mathbf{P}(S_5 = 0 \mid S_1 = 0, S_2 = 0, S_3 = 0, S_4 = 0) = \mathbf{P}(S_5 = 0 \mid S_4 = 0), \\
0 &= \mathbf{P}(S_5 = 0 \mid S_1 = 1, S_2 = 1, S_3 = 0, S_4 = 1) = \mathbf{P}(S_5 = 0 \mid S_4 = 1).
\end{aligned}$$

Ei ole tarvis vaadata sündmuste $\{S_u = 1\}$ tinglikke tõenäosusi, sest et need on määratud sündmuste $\{S_u = 0\}$ vastavate tinglike tõenäosustega.

Küll aga

$$\mathbf{P}(S_4 = 1 \mid S_3 = 0, S_2 = 0, S_1 = 0) = 0 \neq \mathbf{P}(S_4 = 1 \mid S_3 = 0, S_2 = 1, S_1 = 1) = 1,$$

järelikult ei ole tegu Markovi ahelaga.

Mõnevõrra keerukama kontranäite on teinud Cox (1977).

Nii käesolevas töös kui ka mõnel pool mujal (näiteks Bremaud, 1999) oleme Markovi välja defineerinud nii-öelda *lokaalse Markovi omaduse* kaudu, nõudes, et iga *üksik* komponent vektoris $\bar{\mathbf{S}}$ oleks ülejäänud vektorist tinglikult sõltumatu, tingimusel et selle üksiku komponendi naabrid on fikseeritud. Markovi välja on võimalik defineerida ka teisiti (näiteks Grimmett, 2010), nõudes eeltoodule midagi sarnast kõigi mittetühjade vektori $\bar{\mathbf{S}}$ osavektorite kohta.

Definitsioon 4 (globaalne Markovi omadus, Grimmett, 2010 lk 144 järgi). Olgu antud graaf (I, N) ja juhuslik vektor $\bar{\mathbf{S}} = (S_u)_{u \in I}$. Ütleme, et juhuslik vektor $\bar{\mathbf{S}}$ (või selle vektori jaotus) rahuldab *globaalset Markovi omadust*, kui iga mittetühja $J \subset I$ korral kehtib

$$\bar{\mathbf{S}}_J \perp \bar{\mathbf{S}}_{(I \setminus J) \setminus \partial J} \mid \bar{\mathbf{S}}_{\partial J},$$

kus hulka

$$\partial J := \left(\bigcup_{u \in J} \mathcal{N}_u \right) \setminus J$$

nimetame hulga J naabruskonnaks.

Globaalne Markovi omadus on tugevam kui lokaalne Markovi omadus. Sellele asjaolule tuginebki näidete 1 ja 2 võimalikkus. Nimelt on ahelate (ja, nagu hiljem näeme, ka puude) korral lokaalsel ja globaalsel Markovi omadusel ilus alternatiivne tõlgendus. Lokaalne Markovi omadus on sisuliselt *kahepoolne*, sest fikseerida tuleb nii eelnev kui ka järgnev naaber, et vaadeldav ahela komponent oleks sõltumatu ülejäänud minevikust ja tulevikust. Tuleb välja, et globaalne Markovi omadus on selles mõttes *ühepoolne* ehk samaväärne nõudega, et ahela iga komponent oleks tinglikult sõltumatu minevikust, tingimusel et vahetult eelnev naaber on fikseeritud (vaata väidet 5). Seega Markovi ahel rahuldab kindlasti globaalset Markovi omadust ja selles seisnebki Markovi ahela ja ahelal antud üldise Markovi välja erinevus. Põhjaliku arutelu ühepoolse (globaalse) ja kahepoolse (lokaalse) Markovi omaduse omavahelistest suhetest võib leida raamatu Georgii (2011) peatükkidest 10, 11 ja 12.

1.2 Gibbsi jaotus

Statistilisest füüsikast on pärit veel üks lihtsustav eeldus juhusliku vektori jaotuse kohta. Kui tegu on väga paljudest sõltuvatest üksikkomponentidest koosneva süsteemiga, siis jällegi tundub ahvatlev kirjeldada kogu süsteemi käitumist vaid lokaalsete mõjude kaudu. Füüsikas on võetud esmapilgul teistsugune lähenemine ja avaldatud huvipakkuva vektori $\bar{\mathbf{S}}$ tõenäosusfunktsioon korrutisena üle teatavate naabruskondade.

Definitsioon 5 (Bremaud, 1999 lk 256). Fikseerime graafi (I, N) . Mittetühja osahulka $C \subset I$ nimetame *klikiks*, kui kõik elemendid hulgas C on vastastikused naabrid. Lisaks nimetame klikkideks kõiki üheelemendilisi osahulki hulgas I .

Definitsioon 6 (Bremaud, 1999 lk 257 järgi). Olgu $\mathcal{X} \subset \mathbb{R}^d$ ja olgu antud funktsioonide kogum $V_C : \mathcal{X}^n \rightarrow \mathbb{R} \cup \{-\infty\}$, $C \subset I$. Nimetame seda kogumit *Gibbsi potentsiaaliks* hulgal \mathcal{X}^n graafi (I, N) suhtes, kui

1. V_C on konstantselt null, kui C pole klikk, ja
2. iga $\bar{\mathbf{s}}, \bar{\mathbf{z}} \in \mathcal{X}^n$ ja iga $C \subset I$ korral kehtib

$$(\bar{\mathbf{s}}_C = \bar{\mathbf{z}}_C) \Rightarrow (V_C(\bar{\mathbf{s}}) = V_C(\bar{\mathbf{z}})).$$

Eeltoodud tingimuste mõte on lasta Gibbsi potentsiaalil sõltuda ainult klikkidest, lubades neil funktsioonidel siiski formaalselt omada sama määramispiirkonda ja olla defineeritud kõigi I osahulkade jaoks. Siiski on iga $C \subset I$ korral selge ka kirja pildi $V_C(\bar{\mathbf{s}}_C)$ tähendus.

Definitsioon 7 (Bremaud, 1999 lk 257). Olgu $\bar{\mathbf{S}} = (S_u)_{u \in I}$ juhuslik vektor, mis võtab väärtusi hulgas \mathcal{X}^n . Olgu $V_C, C \subset I$, Gibbsi potentsiaal hulgal \mathcal{X}^n graafi (I, N) suhtes. Vektori $\bar{\mathbf{S}}$ jaotust nimetame *Gibbsi jaotuseks*, kui selle tihedus- või tõenäosusfunktsioon p avaldub kujul

$$p(\bar{\mathbf{s}}) \propto \exp(U(\bar{\mathbf{s}})), \tag{3}$$

kus

$$U(\bar{\mathbf{s}}) := \sum_{C \subset I} V_C(\bar{\mathbf{s}}),$$

iga $\bar{\mathbf{s}} \in \mathcal{X}^n$ korral. Kokkuleppeliselt $\exp(-\infty) := 0$.

Käesolevas peatükis pakub huvi otseloomulikult erijuht, kus $\bar{\mathbf{S}}$ on hulgas K^n väärtuseid võttev diskreetne juhuslik vektor, seega $\mathcal{X} = K$ ja $p(\bar{\mathbf{s}}) = \mathbf{P}(\bar{\mathbf{S}} = \bar{\mathbf{s}})$. Füüsikalisest sisust tingituna tuuakse funktsioonidest V_C summa ette välja miinusmärk ja positiivne konstant, mida tõlgendatakse kui temperatuuri T pöördväärtust. Summat $U(\bar{\mathbf{s}})$ nimetatakse sel juhul seisundite konfiguratsiooni $\bar{\mathbf{s}}$ energiaks. Avaldisest

$$\mathbf{P}(\bar{\mathbf{S}} = \bar{\mathbf{s}}) \propto \exp\left(-\frac{1}{T}U(\bar{\mathbf{s}})\right).$$

näeme, et mida kõrgem on konfiguratsiooni $\bar{\mathbf{s}}$ energia, seda madalam on selle esinemise tõenäosus. Samuti, mida kõrgem on temperatuur, seda vähem loeb erinevate konfiguratsioonide erinev energia, kuni piirjuhul $T \rightarrow \infty$ on kõik konfiguratsioonid võrdse tõenäosusega (sõltumata nende energiast).

Näide 3 (Gibbsi jaotuse kuju lihtsustatud põhjendus Grimmett, 2010 lk 145 järgi). Oletame, et igal seisundite konfiguratsioonil $\bar{\mathbf{s}}$ on energia $U(\bar{\mathbf{s}})$ ja süsteemi $\bar{\mathbf{S}}$ keskmine energia $E := \sum_{\bar{\mathbf{s}}} U(\bar{\mathbf{s}})\mathbf{P}(\bar{\mathbf{S}} = \bar{\mathbf{s}})$ on fikseeritud. Lühiduse mõttes tähistame siin $\mathbf{P}(\bar{\mathbf{S}} = \bar{\mathbf{s}}) =: \pi(\bar{\mathbf{s}})$. Otsime süsteemi $\bar{\mathbf{S}}$ konfiguratsioonide jaotust, mis maksimiseeriks entroopiat

$$H(\bar{\mathbf{S}}) = - \sum_{\bar{\mathbf{s}}} \pi(\bar{\mathbf{s}}) \ln \pi(\bar{\mathbf{s}}).$$

Leiame avaldise $\pi(\bar{\mathbf{s}})$ jaoks Lagrange'i määramata kordajate meetodil. Lagrange'i funktsioon on

$$L((\pi(\bar{\mathbf{s}}))_{\bar{\mathbf{s}}}, \beta) = H(\bar{\mathbf{S}}) - \beta \left(\sum_{\bar{\mathbf{s}}} U(\bar{\mathbf{s}})\pi(\bar{\mathbf{s}}) - E \right) - \lambda \left(\sum_{\bar{\mathbf{s}}} \pi(\bar{\mathbf{s}}) - 1 \right).$$

Edasi,

$$\begin{aligned} \frac{\partial}{\partial \pi(\bar{\mathbf{s}})} L((\pi(\bar{\mathbf{s}}))_{\bar{\mathbf{s}}}, \beta) &= -1 - \ln \pi(\bar{\mathbf{s}}) - \beta U(\bar{\mathbf{s}})\pi(\bar{\mathbf{s}}) - \lambda = 0 \\ \Rightarrow \ln \pi(\bar{\mathbf{s}}) &= -\beta U(\bar{\mathbf{s}}) - 1 - \lambda \\ \Rightarrow \pi(\bar{\mathbf{s}}) &\propto e^{-\beta U(\bar{\mathbf{s}})}. \end{aligned}$$

Tegu on maksimumiga, sest funktsioon H kui vektori $(\pi(\bar{\mathbf{s}}))_{\bar{\mathbf{s}}}$ funktsioon on nõgus. Järelikult Gibbsi jaotus on see jaotus, mis maksimiseerib konfiguratsioonide jaotuse entroopiat, tingimusel et keskmine energia on fikseeritud.

Osutub, et (teatud tüüpi) Gibbsi jaotuste ja Markovi väljade vahel on üks-ühene vastavus. Selline Markovi väli $\bar{\mathbf{S}}$, mille kõik konfiguratsioonid $\bar{\mathbf{s}} \in K^n$ on võimalikud, on ühtlasi ka Gibbsi jaotus. Kõigi konfiguratsioonide võimalikkuse nõuet annab formaalselt veidi leevendada nõudeks, mida nimetatakse positiivsuse tingimuseks.

Definitsioon 8 (Robert ja Casella, 1999 lk 294). Ütleme, et diskreetne juhuslik vektor $\bar{\mathbf{S}} = (S_u)_{u \in I}$, mis võtab väärtusi hulgas K^n , rahuldab *positiivsuse tingimust*, kui iga $\bar{\mathbf{s}} = (s_u)_{u \in I} \in K^n$ korral kehtib

$$(\mathbf{P}(S_u = s_u) > 0 \forall u \in I) \Rightarrow \mathbf{P}(\bar{\mathbf{S}} = \bar{\mathbf{s}}) > 0.$$

Sisuliselt nõuab positiivsuse tingimus, et ükski juhuslike suuruste vaheline “interaktsioon” ei tohi olla päris võimatu. Ainus võimalus selleks, et konfiguratsiooni $\bar{\mathbf{s}}$ tõenäosus oleks null, on see, kui tema mingi üksiku komponendi marginaalne tõenäosus on null. Alternatiivselt võime mõelda, et positiivsuse tingimus sisuliselt siiski nõuab kõigile seisunditele rangelt positiivset tõenäosust, kuid võimaldab formaalselt vaadelda ühte ja sedasama seisundite ruumi K kõigi $\bar{\mathbf{S}}$ komponentide jaoks. Ühelt poolt, olgu juhuslik vektor $\bar{\mathbf{S}} = (S_u)_{u \in I}$ selline, et iga $u \in I$ korral juhusliku suuruse S_u seisundite ruum on K_u ja iga $\bar{\mathbf{s}} \in K^* := K_1 \times K_2 \times \dots \times K_n$ korral $\mathbf{P}(\bar{\mathbf{S}} = \bar{\mathbf{s}}) > 0$. Võime kõigi juhuslike suuruste $S_u, u \in I$, ühise seisundite ruumi käsitleda hulka $K := \cup_{u \in I} K_u$ ja sel juhul juhuslik vektor $\bar{\mathbf{S}}$ konfiguratsioonide hulgaga K^n rahuldab positiivsuse tingimust: fikseerime $\bar{\mathbf{s}} \in K^n$, siis

$$\mathbf{P}(S_u = s_u) > 0 \forall u \in I \Rightarrow s_u \in K_u \forall u \in I \Rightarrow \mathbf{P}(\bar{\mathbf{S}} = \bar{\mathbf{s}}) > 0.$$

Teiselt poolt, olgu antud positiivsuse tingimust rahuldav vektor $\bar{\mathbf{S}} = (S_u)_{u \in I}$ konfiguratsioonide hulgaga K^n . Moodustame hulgad $K_u := \{j \in K : \mathbf{P}(S_u = j) > 0\}, u \in I$, ja $K^* := K_1 \times K_2 \times \dots \times K_n \subset K^n$. Paneme tähele, et

$$\bar{\mathbf{s}} \in K^* \Rightarrow \mathbf{P}(S_u = s_u) > 0 \forall u \in I \Rightarrow \mathbf{P}(\bar{\mathbf{S}} = \bar{\mathbf{s}}) > 0$$

ja

$$\bar{\mathbf{s}} \in K^n \setminus K^* \Rightarrow \exists u \in I : \mathbf{P}(S_u = s_u) = 0 \Rightarrow \mathbf{P}(\bar{\mathbf{S}} = \bar{\mathbf{s}}) = 0.$$

Väide 3 (Hammersley-Cliffordi teoreem Bremaud, 1999 lk 260 ja 262 järgi). Olgu $\bar{\mathbf{S}} = (S_u)_{u \in I}$ diskreetne juhuslik vektor, mis võtab väärtusi hulgas K^n . Kui vektor $\bar{\mathbf{S}}$ on Gibbsi jaotusega mingi hulgal K^n ja graafil (I, N) antud Gibbsi potentsiaali $\{V_C\}_{C \subset I}$ suhtes, siis vektor $\bar{\mathbf{S}}$ on Markovi väli graafil (I, N) , kusjuures iga $u \in I$ ja iga $\bar{\mathbf{s}} \in K^n$ korral kehtib

$$\mathbf{P}(S_u = s_u \mid \bar{\mathbf{S}}_{\mathcal{N}_u} = \bar{\mathbf{s}}_{\mathcal{N}_u}) = \frac{\exp\left(\sum_{C \subset I: u \in C} V_C(\bar{\mathbf{s}})\right)}{\sum_{j \in K} \exp\left(\sum_{C \subset I: u \in C} V_C(\bar{\mathbf{s}}_{I \setminus u}, s_u = j)\right)}.$$

Teistpidi, kui vektor $\bar{\mathbf{S}}$ on positiivsuse tingimust rahuldav Markovi väli graafil (I, N) , siis $\bar{\mathbf{S}}$ on Gibbsi jaotusega mingi graafil (I, N) ja hulgal K^n antud Gibbsi potentsiaali suhtes.

Positiivsuse tingimus on oluline, sest et vastasel korral ei pruugi olla võimalik isegi neljatipulisel graafil antud kahe-elementilise seisundite ruumiga Markovi välja tõenäosusfunktsiooni tegurdada.

Tabel 1: Vektori $\bar{\mathbf{S}}$ jaotus näites 4.

s_1	s_2	s_3	s_4	tõenäosus	s_1	s_2	s_3	s_4	tõenäosus
0	0	0	0	8^{-1}	0	1	0	0	0
1	0	0	0	8^{-1}	0	0	1	0	0
1	1	0	0	8^{-1}	1	0	1	0	0
1	1	1	0	8^{-1}	0	1	1	0	0
1	1	1	1	8^{-1}	0	1	0	1	0
0	1	1	1	8^{-1}	1	0	0	1	0
0	0	1	1	8^{-1}	1	0	1	1	0
0	0	0	1	8^{-1}	1	1	0	1	0

Näide 4 (Positiivsuse tingimuse olulisus, Grimmett, 2010 harjutus 7.2 lk 150). Olgu meil nelja tipuga ruudukujuline graaf $(I, N) = (\{1, 2, 3, 4\}, \{\{2, 4\}, \{1, 3\}, \{2, 4\}, \{1, 3\}\})$ ja juhuslik vektor $\bar{\mathbf{S}} = (S_u)_{u \in I}$, mis võtab väärtusi hulgas $\{0, 1\}^4$. Juhusliku vektori jaotus olgu antud tabeliga 1. Sarnaselt näitega 2 saab veenduda, et $\bar{\mathbf{S}}$ on Markovi väli. Näiteks tipu 1 jaoks

$$\begin{aligned}
 0.5 &= \mathbf{P}(S_1 = 1 \mid S_2 = 0, S_4 = 0) = \mathbf{P}(S_1 = 1 \mid S_2 = 0, S_3 = 0, S_4 = 0), \\
 0 &= \mathbf{P}(S_1 = 1 \mid S_2 = 0, S_4 = 1) = \mathbf{P}(S_1 = 1 \mid S_2 = 0, S_3 = 0, S_4 = 1) \\
 &= \mathbf{P}(S_1 = 1 \mid S_2 = 0, S_3 = 1, S_4 = 1), \\
 1 &= \mathbf{P}(S_1 = 1 \mid S_2 = 1, S_4 = 0) = \mathbf{P}(S_1 = 1 \mid S_2 = 1, S_3 = 0, S_4 = 0) \\
 &= \mathbf{P}(S_1 = 1 \mid S_2 = 1, S_3 = 1, S_4 = 0), \\
 0.5 &= \mathbf{P}(S_1 = 1 \mid S_2 = 1, S_4 = 1) = \mathbf{P}(S_1 = 1 \mid S_2 = 1, S_3 = 1, S_4 = 1),
 \end{aligned}$$

ülejäanud tippude analoogiliselt. Enamgi veel, vektor $\bar{\mathbf{S}}$ rahuldab ka globaalset Markovi omadust: hulga I ühe-elementiliste osahulkade jaoks “globaalne ja lokaalne Markovi omadus langevad kokku”, suuremate kui ühe-elementiliste osahulkade korral aga ammendab see osahulk koos oma naabruskonnaga kogu I .

Küll aga ei rahulda $\bar{\mathbf{S}}$ positiivsuse tingimust, sest näiteks $\mathbf{P}(S_1 = 0) = \mathbf{P}(S_2 = 1) = \mathbf{P}(S_3 = 0) = \mathbf{P}(S_4 = 0) = 0.5$, aga $\mathbf{P}(S_1 = 0, S_2 = 1, S_3 = 0, S_4 = 0) = 0$.

Paneme tähele, et graafi (I, N) klikkide hulk on

$$\mathcal{C} := \{\{1\}, \{2\}, \{3\}, \{4\}, \{1, 2\}, \{2, 3\}, \{3, 4\}, \{1, 4\}\}.$$

Veendume, et leidu funktsioone $f_C : \{0, 1\}^{|\mathcal{C}|} \rightarrow \mathbb{R}$, $C \in \mathcal{C}$, mis rahuldaksid iga $\bar{\mathbf{s}} \in \{0, 1\}^4$ korral

$$\prod_{C \in \mathcal{C}} f_C(\bar{\mathbf{s}}_C) = \mathbf{P}(\bar{\mathbf{S}} = \bar{\mathbf{s}}). \quad (4)$$

Oletame vastuväiteliselt, et selline funktsioonide komplekt $\{f_C\}_{C \in \mathcal{C}}$ leidub. Esmalt paneme tähele, et $\mathbf{P}(\bar{\mathbf{S}} = (0, 0, 0, 0)) > 0$, seega $f_{\{u\}}(0) > 0$ iga $u \in I$ korral, ja $\mathbf{P}(\bar{\mathbf{S}} = (1, 1, 1, 1)) > 0$, seega $f_{\{u\}}(1) > 0$ iga $u \in I$ korral.

Uurime tõenäosust

$$\mathbf{P}(\bar{\mathbf{S}} = (0, 1, 0, 0)) = \underbrace{f_{\{1\}}(0)f_{\{2\}}(1)f_{\{3\}}(0)f_{\{4\}}(0)}_{>0} f_{\{1,2\}}(0, 1)f_{\{2,3\}}(1, 0)f_{\{3,4\}}(0, 0)f_{\{1,4\}}(0, 0).$$

Tingimuse 4 kehtides peab see olema positiivne, sest

1. $f_{\{1,2\}}(0, 1) > 0$, sest $\mathbf{P}(\bar{\mathbf{S}} = (0, 1, 1, 1)) > 0$,
2. $f_{\{2,3\}}(1, 0) > 0$, sest $\mathbf{P}(\bar{\mathbf{S}} = (1, 1, 0, 0)) > 0$,
3. $f_{\{3,4\}}(0, 0) > 0$, sest $\mathbf{P}(\bar{\mathbf{S}} = (0, 0, 0, 0)) > 0$,
4. $f_{\{1,4\}}(0, 0) > 0$, sest $\mathbf{P}(\bar{\mathbf{S}} = (0, 0, 0, 0)) > 0$.

Ometi tabelist 1 näeme, et $\mathbf{P}(\bar{\mathbf{S}} = (0, 1, 0, 0)) = 0$. Vastuolu. Järelikult ei saa $\bar{\mathbf{S}}$ olla Gibbsi jaotusega.

Väide 4 (Grimmett, 2010 lk 146). Positiivsuse tingimuse kehtides on lokaalne ja globaalne Markovi omadus samaväärsed.

Tõestus. Olgu $\bar{\mathbf{S}} = (S_u)_{u \in I}$ positiivsuse tingimust rahuldav juhuslik vektor, mis võtab väärtusi hulgas K^n . Meenutame, et positiivsuse tingimuse kehtides võime iga $u \in I$ korral sisuliselt kitsendada komponendi S_u seisundite ruumi ja seejärel konstrueerida vektori $\bar{\mathbf{S}}$ jaoks uue võimalike väärtuste hulga K^* , nii et kõigi konfiguratsioonide $\bar{\mathbf{s}} \in K^*$ tõenäosused on rangelt positiivsed. Väite edasist tõestust vaata (Grimmett, 2010) lk 146. \square

Järgnevalt tutvume paari olulise Markovi väljaga, mida kasutatakse tihti simulatsioonides.

Näide 5 (Isingi mudel, Bregmaud, 1999 lk 258). Isingi mudel on mudel ferromagnetilise materjali (näiteks rauatüki) magnetiliste omaduste kvalitatiivseks kirjeldamiseks. Osakesed eeldatakse asuvat korrapärase võre sõlmpunktides, igal osakesel on kaks võimalikku seisundit. Lihtsuse mõttes vaatame kahemõõtmelist nelinurkset $m \times n$ -võret $\{1, 2, \dots, m\} \times \{1, 2, \dots, n\}$, mille näide on kujutatud joonisel 1. Formaalselt valime $I = \{1, 2, \dots, m\} \times \{1, 2, \dots, n\}$ ja punktile $(k, l) \in I$ defineerime naabruskonna $\mathcal{N}_{(k,l)} := \{(i, j) \in N_m \times N_n \mid (i - k)^2 + (j - l)^2 = 1\}$, naabruskonnad ühendame naabruskondade süsteemiks $N := \{\mathcal{N}_u\}_{u \in I}$. Tegu on tõepoolest naabruskondade süsteemiga, sest punkt ise ei kuulu oma naabruskonda ja naabrussuhe on vastastikune:

$$(i, j) \in \mathcal{N}_{(k,l)} \Leftrightarrow (i - k)^2 + (j - l)^2 = 1 \Leftrightarrow (k, l) \in \mathcal{N}_{(i,j)}.$$

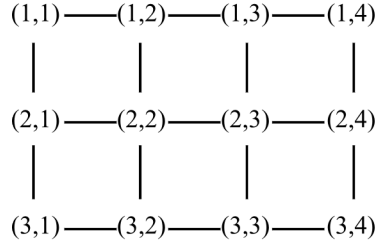
Punktis (k, l) asuva osakese seisund $S_{(k,l)}$ on juhuslik suurus võimalike väärtuste hulgaga $K = \{-1, 1\}$, kõigi osakeste seisundid koondame juhuslikku mn -vektorisse $\bar{\mathbf{S}}$. Mudelil on kolm parameetrit: temperatuur $T > 0$, naabritevahelise seose tugevus $J \in \mathbb{R}$ ja väline magnetväli $h \in \mathbb{R}$. Seisundi $\bar{\mathbf{s}} \in \{-1, 1\}^{mn}$ tõenäosus on

$$\frac{1}{Z} e^{-\frac{1}{T} H(\bar{\mathbf{s}})},$$

kus k on Boltzmanni konstant,

$$H(\bar{\mathbf{s}}) = -J \sum_{\{u,v\} \in E} s_u s_v - h \sum_{u \in I} s_u \text{ ja } Z = \sum_{\bar{\mathbf{z}} \in \{-1,1\}^{mn}} e^{-\frac{1}{T} H(\bar{\mathbf{z}})}.$$

Eelnevas E tähistab graafi (I, N) servade hulka ja s_u tähendab vektori $\bar{\mathbf{s}}$ elementi indeksiga u (punktis u asuva osakese seisundit).



Joonis 1: Kahemõõtmelise Isingi mudeli naabruskondade süsteemi näide.

Vektori $\bar{\mathbf{S}}$ jaotus on Gibbsi jaotus, sest graafi (I, N) klikid ongi parajasti üheelemendilised hulgad $\{u\}$, $u \in I$ ja naabrite paarid $\{u, v\} \in E$. Järelikult Isingi mudel on Gibbsi jaotus, seega Markovi väli.

Näide 6 (Potts'i mudel, Grimmett, 2010 lk 149 järgi). Potts'i mudel on (välise magnetväljata) Isingi mudeli üks võimalik üldistus lõplikule seisundite ruumile. Vaatame jällegi olukorda, kus (I, N) on kahemõõtmeline nelinurkne võre (nagu eelmises näiteski). Olgu $\bar{\mathbf{S}} = (S_u)_{u \in I}$ juhuslik vektor võimalike väärtuste hulgaga K^n , kus K on vähemalt kahe elemendiga lõplik hulk. Graafi (I, N) servade hulga tähistame jällegi E – need on ühtlasi ka selle graafi kahe-elemendilised klikid. Potts'i mudeli korral

$$\mathbf{P}(\bar{\mathbf{S}} = \bar{\mathbf{s}}) \propto \exp \left(\beta \sum_{\{u,v\} \in E} I_{s_u=s_v} \right), \quad (5)$$

kus $\beta \in \mathbb{R}$ on konstant. Kuna tõenäosus on proportsionaalne korrutisega üle (mõnedest) klikkidest sõltuvate funktsioonide, siis on tegu Gibbsi jaotusega, millele vastava Gibbsi potentsiaali ainsad nullist erinevad funktsioonid on

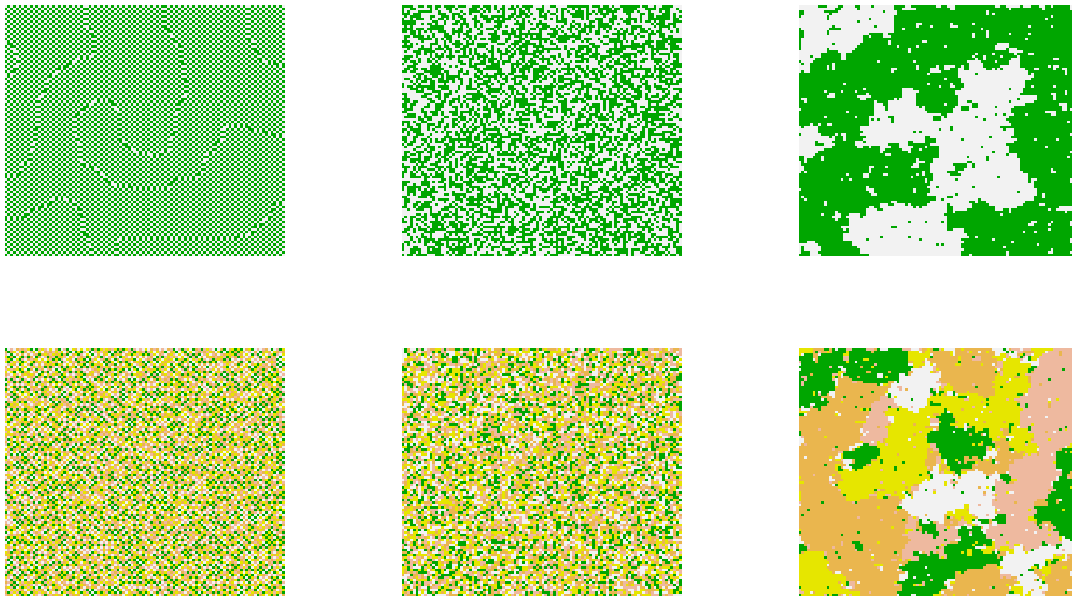
$$V_{\{u,v\}}(s_u, s_v) = \beta I_{s_u=s_v}, \{u, v\} \in E.$$

Vastavalt väitele 3 on tegu Markovi väljaga, kusjuures

$$\mathbf{P}(S_u = s_u \mid \bar{\mathbf{S}}_{\mathcal{N}_u} = \bar{\mathbf{s}}_{\mathcal{N}_u}) = \frac{\exp \left(\beta \sum_{v \in \mathcal{N}_u} I_{s_u=s_v} \right)}{\sum_{j \in K} \exp \left(\beta \sum_{v \in \mathcal{N}_u} I_{j=s_v} \right)}.$$

Paneme tähele, et kui $\beta = 0$, siis iga “pikslivärv” valitakse sõltumatult kõigist teistest pikslitest diskreetsest ühtlasest jaotusest. Kui $\beta < 0$, siis on naabriteks tõenäolisemalt erinevad seisundid, ja kui $\beta > 0$, siis moodustuvad pigem ühe ja sama seisundi klastrid. Mõne Potts'i mudeli realisatsiooni näited 100×100 võrel on joonisel 2. Kui Potts'i mudelil on kaks seisundit, siis võime need seisundid ümber tähistada arvudeks 1 ja -1. Sel juhul $I_{i=j} = 0.5(1 + ij)$, kui $i, j \in K$, ja

$$\mathbf{P}(\bar{\mathbf{S}} = \bar{\mathbf{s}}) \propto \exp \left(\beta \sum_{\{u,v\} \in E} I_{s_u=s_v} \right) \propto \exp \left(\frac{\beta}{2} \sum_{\{u,v\} \in E} s_u s_v \right)$$



Joonis 2: Potts'i mudelite realisatsioonide näiteid 100×100 võrel. Ülemises reas on võimalike seisundite arv $k = 2$, alumises reas $k = 5$. Vasakpoolses tulbas on $\beta = -100$, keskmises tulbas $\beta = 0$ ja parempoolses tulbas $\beta = 1.1 \ln(1 + \sqrt{k})$.

ehk tegu ongi Isingi mudeliga, kus $\frac{J}{T_k} = \frac{\beta}{2}$.

Kuna

$$\sum_{\{u,v\} \in E} I_{s_u=s_v} = \frac{1}{2} \sum_{u \in I} \sum_{v \in \mathcal{N}_u} I_{s_u=s_v} = \frac{1}{2} \sum_{u \in I} \left(|\mathcal{N}_u| - \sum_{v \in \mathcal{N}_u} I_{s_u \neq s_v} \right),$$

siis saame Potts'i mudeli kirjutada ka kujul

$$\mathbf{P}(\bar{\mathbf{S}} = \bar{\mathbf{s}}) \propto \exp \left(-\frac{\beta}{2} \sum_{u \in I} \sum_{v \in \mathcal{N}_u} I_{s_u \neq s_v} \right). \quad (6)$$

Sellel kujul Potts'i mudelit on lihtne veelgi üldistada, kui valemis 6 indikaator-funktsioon I asendada mingi muu funktsiooniga $\alpha : K \times K \rightarrow [0, \infty)$, nii et $\alpha(i, i) = 0 \forall i \in K$ ja $\alpha(i, j) = \alpha(j, i) > 0$, kui $i \neq j$ (Arnesen, 2010).

1.3 Puukujuline sõltuvusstruktuur

Nagu veendusime näites 1, siis levinud erijuht ülaldefineeritud Markovi väljast on (diskreetse ajaga) Markovi ahel, mispuhul (I, N) on lihtahel. Lihtahel on aga ka erijuht *puust* ehk sidusast tsükliteta graafist. Mil määral võiks ahelal defineeritud Markovi väljade omadused üle kanduda üldisematele, puudel defineeritud Markovi väljadele? Käesolevas alapeatükis laiendame mõningaid Markovi ahelatel kehtivaid võrduseid puude peal defineeritud Markovi väljadele.

1.3.1 Tähistused

Olgu (I, N) puu. Puu juureks määrame tipu indeksiga 1. Puu tippudel orienteerumise formaliseerimiseks defineerime funktsiooni $\rho : I \cup \{0\} \rightarrow I \cup \{0\}$, mis igale tipule $u > 1$ seab vastavusse tipu u vanema ehk selle naabertipu, mis on osa tippu u juurega ühendavast lihtahelast (iga kahte erinevat puu tippu ühendab täpselt üks lihtahel). Lisaks defineerime $\rho(1) := 0$ ja $\rho(0) := 0$. Tipu $u \in I$ lasteks nimetame neid tipu u naabertippe, mis ei kuulu tippu u juurega ühendavasse lihtahelasse (jäävad juurest kaugemale). Kui $J \subset I \cup \{0\}$, siis $\rho^{-1}(J)$ tähistab hulga J originaali hulgateooria mõttes. Erijuhul, kui hulgas $J = \{u\}$ on üks element, siis kõikjal tähistustes asendame $\{u\}$ lihtsalt u -ga. Näiteks $\rho^{-1}(\{u\}) =: \rho^{-1}(u)$ annab kätte tipu u laste hulga. Iga hulga $J \subset I \cup \{0\}$ ja naturaalarvu k korral tähistame

$$\rho^{-k}(J) := \underbrace{\rho^{-1}(\rho^{-1}(\dots\rho^{-1}(J)\dots))}_{k \text{ tükki}}, \quad \rho^0(J) := J$$

ja

$$R^-(J, k) := \bigcup_{j=0}^{k-1} \rho^{-j}(J).$$

Erijuhul $J = \{u\}$ kujutab hulk $R^-(u, k)$ endast tippu u koos laste, lastelaste, jne kuni $k - 1$ järku “alanevate sugulastega” – kokku k “põlvkonda” tippe.

Väga sageli tuleb järgnevas tekstis tegemist puu (I, N) alampuudega, mis koosnevad mingist tipust $u \in I$ koos kõigi tema “alanevate sugulastega”. Loomulikult nimetame sellise alampuu juureks tippu u ennast. Sellise alampuu tippude hulga jaoks võtame kasutusele uue tähistuse:

$$I_u := \bigcup_{k=1}^{\infty} R^-(u, k).$$

Sellise alampuu servade hulka määrava naabruskondade süsteemi tähistame N_u , mille võime kirjutada kui

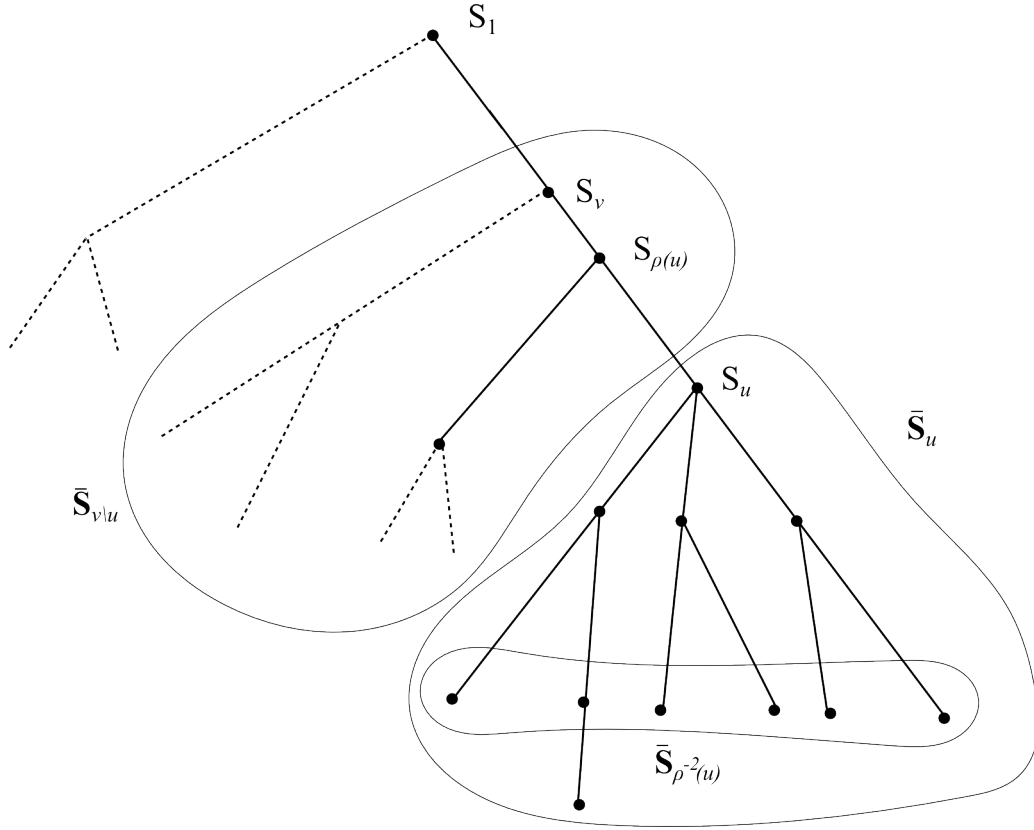
$$N_u = \{\mathcal{N}_v \cap I_u\}_{v \in I_u}.$$

Lühiduse mõttes tähistame veel iga $u \in I, v \in I_u$ korral $\bar{\mathbf{S}}_u := \bar{\mathbf{S}}_{I_u}$ ja $\bar{\mathbf{S}}_{u \setminus v} := \bar{\mathbf{S}}_{I_u \setminus I_v}$. Tähistusi illustreerib joonis 3, mis on tehtud Durand jt. (2004) eeskujul.

Mitmed järgnevad tõestused tuginevad induktsioonile puu või alampuu suuruse järgi. Täpsemalt, alampuu (I_u, N_u) sügavuseks nimetame tippude arvu (mitte servade arvu) pikimas graafile (I_u, N_u) kuuluvas lihtahelas, mis algab tipust u . Puu (I, N) sügavuseks nimetame loomulikult alampuu (I_1, N_1) sügavust. Tipu u sügavuseks nimetame alampuu (I_u, N_u) sügavust ja tähistame seda $d(u)$.

1.3.2 Tulemused

Tutvume puudel defineeritud Markovi väljade mõningate omadustega, põhitulemuseks on järgnev väide 5. Esmalt toome sisse nõndanimetatud ühepoolse Markovi omaduse ja ühe “reaalselt kasulikku” tegurdamisomaduse ning seome need seejärel globaalse Markovi omadusega. Kuna globaalsest Markovi omadusest järeldub igal juhul lokaalne Markovi omadus, siis igauks eeltoodud kolmest tingimusest on piisav (lokaalse)



Joonis 3: Tähistuste skeem.

Markovi omaduse kehtimiseks.

Olgu (I, N) puu ja $\bar{\mathbf{S}} = (S_u)_{u \in I}$ diskreetne juhuslik vektor, mis võtab väärtusi hulgas K^n . Olgu hulgas I ja K lõplikud.

Definitsioon 9 (Georgii, 2011 lk 190 järgi). Ütleme, et juhuslik vektor $\bar{\mathbf{S}}$ rahuldab *ühepoolset Markovi omadust*, kui iga $u > 1$ korral kehtib

$$\mathbf{P}(S_u = s_u \mid \bar{\mathbf{S}}_{1 \setminus u} = \bar{s}_{1 \setminus u}) = \mathbf{P}(S_u = s_u \mid S_{\rho(u)} = s_{\rho(u)}),$$

kui vaid $\bar{s}_{1 \setminus u} \in K^{n-|I_u|}$ on positiivse tõenäosusega.

Definitsioon 10. Ütleme, et juhuslik vektor $\bar{\mathbf{S}}$ rahuldab *tegurdamisomadust*, kui iga $u \in I$ ja iga $\bar{s}_u \in K^{|I_u|}$ korral kehtib

$$\mathbf{P}(\bar{\mathbf{S}}_u = \bar{s}_u) = \mathbf{P}(S_u = s_u) \prod_{v \in I_u \setminus u} \mathbf{P}(S_v = s_v \mid S_{\rho(v)} = s_{\rho(v)}),$$

kui vaid kõik tinglikud tõenäosused on defineeritud.

Märgime, et tegurdamisomaduse kehtides $\mathbf{P}(\bar{\mathbf{S}} = \bar{s}) = 0$ parajasti siis, kui mingi vanem-laps paari tõenäosus on null ehk leidub $u > 1$, et $\mathbf{P}(S_u = s_u, S_{\rho(u)} = s_{\rho(u)}) = 0$. Tõepoolest, kui $\mathbf{P}(S_u = s_u, S_{\rho(u)} = s_{\rho(u)}) = 0$, siis selgelt $\mathbf{P}(\bar{\mathbf{S}} = \bar{s}) = 0$. Teiselt poolt, kui $\mathbf{P}(\bar{\mathbf{S}} = \bar{s}) = 0$, siis on kaks võimalust. Kui kõik tinglikud tõenäosused $\mathbf{P}(S_u = s_u \mid S_{\rho(u)} = s_{\rho(u)})$ on defineeritud ehk $\mathbf{P}(S_{\rho(u)} = s_{\rho(u)}) > 0, u \in I$, siis tegurdamisomaduse tõttu

$$\mathbf{P}(\bar{\mathbf{S}} = \bar{s}) = \mathbf{P}(S_1 = s_1) \prod_{u > 1} \frac{\mathbf{P}(S_u = s_u, S_{\rho(u)} = s_{\rho(u)})}{\mathbf{P}(S_{\rho(u)} = s_{\rho(u)})}$$

ehk mingi $u \in I$ korral $\mathbf{P}(S_u = s_u, S_{\rho(u)} = s_{\rho(u)}) = 0$. Kui mingi tinglik tõenäosus pole defineeritud ehk mingi $u \in I$ korral $\mathbf{P}(S_{\rho(u)} = s_{\rho(u)}) = 0$, siis ka $\mathbf{P}(S_u = s_u, S_{\rho(u)} = s_{\rho(u)}) = 0$.

Väide 5. Ühepoolne Markovi omadus, tegurdamisomadus ja globaalne Markovi omadus on samaväärsed.

Paneme tähele, et kui $\bar{\mathbf{S}}$ rahuldab globaalset Markovi omadust, siis $\bar{\mathbf{S}}$ rahuldab ka lokaalset Markovi omadust ja seega on Markovi väli puul (I, N) . Teiselt poolt, kui on lihtsalt antud Markovi väli, siis vastavalt väitele 4 piisab näiteks positiivsuse tingimusest, et see Markovi väli rahuldaks globaalset Markovi omadust ja seega ka tegurdamisomadust.

Enne väite 5 tõest anname paar vajalikku tulemust. Kõigepealt näitame, kuidas saab mittevajalikke tingimusi ära visata või juurde võtta.

Väide 6. Olgu A, B, C ja $D_i, i \in J$, ülimalt loenduv hulk sündmuseid. Lisaks olgu $\mathbf{P}(B, C) > 0$ ja moodustagu sündmused $D_i, i \in J$, elementaarsündmuste ruumi tükelduse. Kui iga sellise $i \in J$ korral, et $\mathbf{P}(B, C, D_i) > 0$, kehtib $\mathbf{P}(A | B, C, D_i) = \mathbf{P}(A | B)$, siis $\mathbf{P}(A | B, C) = \mathbf{P}(A | B)$.

Tõestus. Kehtigu $\mathbf{P}(A | B, C, D_i) = \mathbf{P}(A | B)$ alati kui $\mathbf{P}(B, C, D_i) > 0$. Kuna $D_i, i \in J$ on sündmuste täissüsteem, siis

$$\mathbf{P}(A | B, C) = \sum_{i \in J} \mathbf{P}(A, D_i | B, C).$$

Paneme tähele, et kui $\mathbf{P}(B, C, D_i) = 0$, siis asjaolu $A \cap B \cap C \cap D_i \subset B \cap C \cap D_i$ tõttu kehtib $\mathbf{P}(A, B, C, D_i) \leq \mathbf{P}(B, C, D_i) = 0$ ja järelikult ka $\mathbf{P}(A, D_i | B, C) = 0$. Seega

$$\sum_{i \in J} \mathbf{P}(A, D_i | B, C) = \sum_{i: \mathbf{P}(B, C, D_i) > 0} \mathbf{P}(A, D_i | B, C).$$

Nüüd kasutame ketireeglit ja eeldust, saame

$$\begin{aligned} \mathbf{P}(A | B, C) &= \sum_{i: \mathbf{P}(B, C, D_i) > 0} \mathbf{P}(D_i | B, C) \mathbf{P}(A | B, C, D_i) \\ &= \sum_{i: \mathbf{P}(B, C, D_i) > 0} \mathbf{P}(D_i | B, C) \mathbf{P}(A | B) \\ &= \mathbf{P}(A | B) \sum_{i: \mathbf{P}(B, C, D_i) > 0} \mathbf{P}(D_i | B, C) \\ &= \mathbf{P}(A | B). \end{aligned}$$

□

Eelnevast tulemusest järeldub otseselt allolev väide, mis sisuliselt ütleb, et kui tuleviku prognoosimisel ei anna praegusele ajahetkele lisainformatsiooni kogu eelnev minevik, siis ei anna praegusele ajahetkele lisa ka mistahes osa kogu eelnevast minevikust.

Väide 7. Olgu (I, N) puu ja $\bar{\mathbf{S}} = (S_u)_{u \in I}$ diskreetne juhuslik vektor. Olgu $u > 1$ suvaline. Olgu $I^* \subset I_u$ mittetühi, $\bar{\mathbf{s}}_{I^*} \in K^{|I^*|}$ ja $A := \{\bar{\mathbf{S}}_{I^*} = \bar{\mathbf{s}}_{I^*}\}$. Võrdus

$$\mathbf{P}(A \mid \bar{\mathbf{S}}_{1 \setminus u} = \bar{\mathbf{s}}_{1 \setminus u}) = \mathbf{P}(A \mid S_{\rho(u)} = s_{\rho(u)})$$

kehtib iga positiivse tõenäosusega $\bar{\mathbf{s}}_{1 \setminus u} \in K^{n-|I_u|}$ korral parajasti siis, kui võrdus

$$\mathbf{P}(A \mid S_{\rho(u)} = s_{\rho(u)}, \bar{\mathbf{S}}_J = \bar{\mathbf{s}}_J) = \mathbf{P}(A \mid S_{\rho(u)} = s_{\rho(u)})$$

kehtib iga $J \subset I_{1 \setminus u} \setminus \rho(u)$ ja positiivse tõenäosusega $\bar{\mathbf{s}}_{\rho(u) \cup J} \in K^{1+|J|}$ korral.

Tõestus. \Leftarrow

Ilmne: vali $J = I_{1 \setminus u} \setminus \rho(u)$, tulemuseks on täpselt vajalik võrdus.

\Rightarrow Fikseerime $J \subset I_{1 \setminus u} \setminus \rho(u)$ ja sellise $\bar{\mathbf{s}}_{\rho(u) \cup J} \in K^{1+|J|}$, et $\mathbf{P}(\bar{\mathbf{S}}_{\rho(u) \cup J} = \bar{\mathbf{s}}_{\rho(u) \cup J}) > 0$.

Tähistame $J^c := (I_{1 \setminus u} \setminus \{\rho(u)\}) \setminus J$.

Kasutame väidet 6. Selleks valime väite 6 tähistustes $B = \{S_{\rho(u)} = s_{\rho(u)}\}$, $C = \{\bar{\mathbf{S}}_J = \bar{\mathbf{s}}_J\}$ ja ülimalt loenduvaks sündmuste täissüsteemiks $\{\bar{\mathbf{S}}_{J^c} = \bar{\mathbf{z}}_{J^c}\}$, $\bar{\mathbf{z}}_{J^c} \in K^{|J^c|}$. \square

Tõestus väitele 5. (ühepoolne Markovi omadus \Rightarrow tegurdamisomadus)

Eeldame, et iga $v > 1$ korral kehtib

$$\mathbf{P}(S_v = s_v \mid \bar{\mathbf{S}}_{1 \setminus v} = \bar{\mathbf{s}}_{1 \setminus v}) = \mathbf{P}(S_v = s_v \mid S_{\rho(v)} = s_{\rho(v)}),$$

kui vaid $\bar{\mathbf{s}}_{1 \setminus v} \in K^{n-|I_v|}$ on positiivse tõenäosusega. Vastavalt väitele 7 kehtib siis iga mittetühja $J \subset I \setminus I_v$ korral

$$\mathbf{P}(S_v = s_v \mid S_{\rho(v)} = s_{\rho(v)}, \bar{\mathbf{S}}_J = \bar{\mathbf{s}}_J) = \mathbf{P}(S_v = s_v \mid S_{\rho(v)} = s_{\rho(v)}),$$

kui vaid $\bar{\mathbf{s}}_{\rho(v) \cup J} \in K^{|J \cup \rho(v)|}$ on positiivse tõenäosusega. Väitest 1 järgeldub nüüd, et iga $v > 1$ korral on S_v ja $\bar{\mathbf{S}}_J$ tinglikult sõltumatud, kui $S_{\rho(v)}$ on fikseeritud, ehk

$$\begin{aligned} & \mathbf{P}(S_v = s_v, \bar{\mathbf{S}}_J = \bar{\mathbf{s}}_J \mid S_{\rho(v)} = s_{\rho(v)}) \\ &= \mathbf{P}(S_v = s_v \mid S_{\rho(v)} = s_{\rho(v)}) \mathbf{P}(\bar{\mathbf{S}}_J = \bar{\mathbf{s}}_J \mid S_{\rho(v)} = s_{\rho(v)}), \end{aligned} \quad (7)$$

kui vaid $\mathbf{P}(S_{\rho(v)} = s_{\rho(v)}) > 0$.

Olgu $u \in I$ suvaline. Olgu $\bar{\mathbf{S}} \in K^n$ selline, et $\mathbf{P}(S_v = s_v) > 0 \forall v \in \rho(I_u)$. Järjestame hulga I_u tipud kasvava sügavuse järgi jadasse v_1, v_2, \dots, v_m , siis iga $i \leq j$ korral $d(v_i) \leq d(v_j)$. Näha on, et see jada algab lehtedega ja lõpeb tipuga u . Paneme veel tähele, et iga $i < m$ korral $\rho(v_i) \in \{v_{i+1}, v_{i+2}, \dots, v_m\}$, sest et iga tipu vanema sügavus on rangelt suurem kui tipu enda sügavus.

Tähistame $A_i := \{S_{v_i} = s_{v_i}\}$ ja $A_{\rho(i)} := \{S_{\rho(v_i)} = s_{\rho(v_i)}\}$, kui $i \in \{1, 2, \dots, m\}$. Iga $i < m$ korral kehtib

$$\begin{aligned} & \mathbf{P}(A_i, A_{i+1}, \dots, A_m) \\ &= \mathbf{P}(A_{\rho(i)}) \mathbf{P}(A_i, A_{i+1}, \dots, A_m \mid A_{\rho(i)}) \\ &= \mathbf{P}(A_{\rho(i)}) \mathbf{P}(A_{i+1}, \dots, A_m \mid A_{\rho(i)}) \mathbf{P}(A_i \mid A_{\rho(i)}) \\ &= \mathbf{P}(A_{i+1}, \dots, A_m) \mathbf{P}(A_i \mid A_{\rho(i)}). \end{aligned}$$

Esimene ja kolmas võrdus tulevad ketireeglist, teine võrdus tuleb võrdusest 7. Teise ja kolmanda võrduse juures on kasutatud ka asjaolu, et $\rho(v_i) \in \{v_{i+1}, v_{i+2}, \dots, v_m\}$. Saadud tulemust korduvalt rakendades saame

$$\begin{aligned} \mathbf{P}(\bar{\mathbf{S}}_u = \bar{s}_u) &= \mathbf{P}(A_1, A_2, A_3, \dots, A_m) \\ &= \mathbf{P}(A_2, A_3, \dots, A_m) \mathbf{P}(A_1 \mid A_{\rho(1)}) \\ &= \mathbf{P}(A_3, \dots, A_m) \mathbf{P}(A_2 \mid A_{\rho(2)}) \mathbf{P}(A_1 \mid A_{\rho(1)}) \\ &\dots \\ &= \mathbf{P}(A_m) \prod_{i=1}^{m-1} \mathbf{P}(A_i \mid A_{\rho(i)}), \end{aligned}$$

nagu vaja.

(tegurdamisomadus \Rightarrow globaalne Markovi omadus)

Olgu $J \subset I$ mittetühi. Näitame, et tinglik tõenäosus $\mathbf{P}(\bar{\mathbf{S}}_J = \bar{s}_J \mid \bar{\mathbf{S}}_{I \setminus J} = \bar{s}_{I \setminus J})$ ei sõltu tegelikult vektori $\bar{s}_{I \setminus (J \cup \partial J)}$ väärtustest, ja järeldame vajaliku tulemuse väite 1 abil. Täpsemalt, olgu $\bar{s} \in K^n$ selline, et $\mathbf{P}(\bar{\mathbf{S}}_{I \setminus J} = \bar{s}_{I \setminus J}) > 0$. Jagame hulga I lõikumatuks osahulkadeks:

$$I = \underbrace{(\rho^{-1}(J) \setminus J)}_{\text{ise ei ole ja vanemad on } J\text{-s}} \cup \underbrace{(\rho^{-1}(J) \cap J)}_{\text{ise on ja vanemad on } J\text{-s}} \cup \underbrace{(J \setminus \rho^{-1}(J))}_{\text{ise on ja vanemad ei ole } J\text{-s}} \cup \underbrace{(I \setminus (J \cup \rho^{-1}(J)))}_{\text{ise ei ole ja vanemad ei ole } J\text{-s}}.$$

Iga $\bar{z}_J \in K^{|J|}$ korral, vastavalt eeltoodud tükeldusele ja eeldatud tegurdamisomadusele (tipu $u = 1$ jaoks),

$$\begin{aligned} &\mathbf{P}(\bar{\mathbf{S}}_J = \bar{z}_J, \bar{\mathbf{S}}_{I \setminus (J \cup \partial J)} = \bar{s}_{I \setminus (J \cup \partial J)}, \bar{\mathbf{S}}_{\partial J} = \bar{s}_{\partial J}) \\ &= \left(\prod_{u \in \rho^{-1}(J) \setminus J} \mathbf{P}(S_u = s_u \mid S_{\rho(u)} = z_{\rho(u)}) \right) \left(\prod_{u \in \rho^{-1}(J) \cap J} \mathbf{P}(S_u = z_u \mid S_{\rho(u)} = z_{\rho(u)}) \right) \cdot \\ &\quad \left(\prod_{\substack{u \in J \setminus \rho^{-1}(J) \\ u \neq 1}} \mathbf{P}(S_u = z_u \mid S_{\rho(u)} = s_{\rho(u)}) \right) \left(\prod_{\substack{u \in I \setminus (J \cup \rho^{-1}(J)) \\ u \neq 1}} \mathbf{P}(S_u = s_u \mid S_{\rho(u)} = s_{\rho(u)}) \right) \cdot \\ &\mathbf{P}(S_1 = s_1), \end{aligned} \tag{8}$$

kus $s_1 := z_1$, juhul kui $1 \in J$. Paneme tähele, et eelnevas korrutises ükski kolmest esimesest sulgudes olevast tegurist ei sõltu vektori $\bar{s}_{I \setminus (J \cup \partial J)}$ väärtustest.

Nüüd näeme, et ka

$$\mathbf{P}(\bar{\mathbf{S}}_J = \bar{s}_J \mid \bar{\mathbf{S}}_{I \setminus J} = \bar{s}_{I \setminus J}) = \frac{\mathbf{P}(\bar{\mathbf{S}}_J = \bar{s}_J, \bar{\mathbf{S}}_{I \setminus J} = \bar{s}_{I \setminus J})}{\sum_{\bar{z}_J \in K^{|J|}} \mathbf{P}(\bar{\mathbf{S}}_J = \bar{z}_J, \bar{\mathbf{S}}_{I \setminus J} = \bar{s}_{I \setminus J})} \tag{9}$$

ei sõltu vektori $\bar{s}_{I \setminus (J \cup \partial J)}$ väärtustest: kui viime nii lugeja kui nimetaja kujule 8, siis nimetajas saame selle kuju neljanda teguri (mis ei sõltu vektorist \bar{z}_J) tuua sulgude ette ja taandada. See aga oli ainus tegur, mis sõltus vektori $\bar{s}_{I \setminus (J \cup \partial J)}$ väärtusest. Samuti on avaldise 9 väärtus lõigus $[0, 1]$, sest lugejas olev korrutis on üks nimetaja liidetavatest. Vastavalt väitele 1 on vektorid $\bar{\mathbf{S}}_J$ ja $\bar{\mathbf{S}}_{I \setminus (J \cup \partial J)}$ tinglikult sõltumatud,

kui $\bar{\mathbf{S}}_{\partial J}$ on fikseeritud. Järelikult $\bar{\mathbf{S}}$ rahuldab globaalset Markovi omadust.

(*globaalne Markovi omadus \Rightarrow ühepoolne Markovi omadus*)

Eeldame, et $\bar{\mathbf{S}}$ rahuldab globaalset Markovi omadust. Fikseerime $u > 1$. Valime $J = I_u$. Paneme tähele, et $\partial I_u = \{\rho(u)\}$, seega iga $\bar{\mathbf{s}} \in K^n$ korral

$$\mathbf{P}(\bar{\mathbf{S}}_u = \bar{\mathbf{s}}_u \mid \bar{\mathbf{S}}_{1 \setminus u} = \bar{\mathbf{s}}_{1 \setminus u}) = \mathbf{P}(\bar{\mathbf{S}}_u = \bar{\mathbf{s}}_u \mid S_{\rho(u)} = s_{\rho(u)}),$$

kui vaid $\mathbf{P}(\bar{\mathbf{S}}_{1 \setminus u} = \bar{\mathbf{s}}_{1 \setminus u}) > 0$. Saame soovitud tulemuse, kui summeerime eelneva võrduse mõlemaid pooli üle $\bar{\mathbf{S}}_{I_u \setminus u}$ kõigi võimalike väärtuste. \square

2 Varjatud Markovi välja mudel

Võib juhtuda, et huvipakkuv protsess (näiteks Markovi väli) pole otseselt vaadeldav, vaid selle asemel näeme protsessi mingil määral “rikutud”, “ebapuhast” varianti. Näiteks võime ilusa, kontrastse must-valgete pikslitega pildi asemel näha müraga saastunud halli varjundeid. Sellist olukorda võime mudeldada justkui kahekihilise mudeli abil, kus nähtava vaatluste protsessi all eeldame olevat varjatud protsessi.

Täpsemalt, olgu $I := \{1, 2, \dots, n\}$ ja N temal antud naabruskondade süsteem. Vaatame kahte juhuslikku vektorit: diskreetne juhuslik vektor $\bar{\mathbf{S}} = (S_u)_{u \in I}$, mille väärtused on hulgas K^n (see saab olema *varjatud kiht*) ja (mitte tingimata diskreetne) juhuslik vektor $\bar{\mathbf{X}} = (X_u)_{u \in I}$, mille väärtused on hulgas \mathcal{X}^n , kus $\mathcal{X} \subset \mathbb{R}^d$ mingi $d \in \mathbb{N}$ korral (see saab olema *vaatluste* vektor).

Definitsioon 11 (Lember ja Koloydenko, 2014 eeskujul). Paari $(\bar{\mathbf{S}}, \bar{\mathbf{X}})$ nimetame *varjatud Markovi välja mudeliks (graafil (I, N))* (i.k. *hidden Markov random field model*), kui

1. $\bar{\mathbf{S}}$ on Markovi väli graafil (I, N) , mis rahuldab globaalset Markovi omadust,
2. vektori $\bar{\mathbf{X}}$ elemendid on omavahel tinglikult sõltumatud, kui $\bar{\mathbf{S}}$ on fikseeritud, ja
3. iga $u \in I$ korral X_u on tinglikult sõltumatu vektorist $\bar{\mathbf{S}}_{I \setminus u}$, kui S_u on fikseeritud.

Lihtsalt *varjatud Markovi mudeli* all mõeldakse tavaliselt varjatud Markovi välja mudelit, milles varjatud kiht on Markovi ahel. Juhime tähelepanu, et varjatud kihilt $\bar{\mathbf{S}}$ nõuame globaalset Markovi omadust, mitte lihtsalt lokaalset Markovi omadust. Ideeline põhjus seisneb selles, et soovime varjatud Markovi *välja* mudelitega üldistada varjatud Markovi mudeleid ahelatelt teistele graafidele (käesolevas töös eelkõige puudele). Näide 2 aga illustreerib, et kui Markovi väljalt mitte nõuda globaalset Markovi omadust, ei tarvitse Markovi väli ahelal üldse olla Markovi ahel. Praktiline põhjus seisneb selles, et edaspidises käsitluses on vaja tegurdamisomadust 10.

Toome vektori $(\bar{\mathbf{S}}, \bar{\mathbf{X}})$ jaotuse kirjeldamiseks sisse erinevate marginaal- ja tinglike tiheduste süsteemi Lember (2013) peatükk 7.2.1 eeskujul. Peamine põhjus tähistuste süsteemi muutmiseks seisneb selles, et vaatluste vektor $\bar{\mathbf{X}}$ võib olla ka pidev, mispuhul ei ole enam mõistlik rääkida näiteks sündmuse $\{\bar{\mathbf{X}} = \bar{\mathbf{x}}\}$ tõenäosusest. Mistahes mittetühjade $J_1, J_2 \subset I$ korral tähistame

$$p(J_1) := \mathbf{P}(\bar{\mathbf{S}}_{J_1} = \bar{\mathbf{s}}_{J_1}),$$

$$p(J_1 | J_2) := \mathbf{P}(\bar{\mathbf{S}}_{J_1} = \bar{\mathbf{s}}_{J_1} | \bar{\mathbf{S}}_{J_2} = \bar{\mathbf{s}}_{J_2}),$$

kui seisundite konfiguratsioon $\bar{\mathbf{s}}$ (või selle osa) on fikseeritud või muidu kontekstist selge. Kui $\bar{\mathbf{s}}$ (või selle osa) on aga vaja ilmutatult ära märkida, siis kirjutame selle p alaindeksisse. Näiteks $p_{\bar{\mathbf{s}}}(J_1 | J_2) := p_{\bar{\mathbf{s}}_{J_1} | \bar{\mathbf{s}}_{J_2}}(J_1 | J_2) := p(J_1 | J_2)$. Ka edaspidises tähistuses asendame võimaluse korral konfiguratsiooni osa $\bar{\mathbf{s}}_J$ lihtsalt vastava indeksite hulgaga J . Lühiduse huvides kasutame justnimelt sellist lähenemist, sest et edasises mõttearenduses kipub sageli just indeksite hulk olema väga muutlik,

varjatud seisundite konfiguratsioon \bar{s} ise aga kontekstist selge. Meenutame, et kui hulgas $J \subset I$ on üks element u , siis kirjutame $\{u\}$ asemel u . Seega tähistame näiteks $\mathbf{P}(S_u = j) =: p_j(u)$ ja $\mathbf{P}(S_u = j \mid S_v = i) =: p_{j|i}(u \mid v)$.

Iga mittetühja $J \subset I$ korral tähistagu $p(\bar{\mathbf{x}}_J \mid I)$ vaatluste vektori $\bar{\mathbf{X}}_J$ realisatsiooni $\bar{\mathbf{x}}_J$ tinglikku tihedust (või tõenäosust, kui $\bar{\mathbf{X}}$ on diskreetne), tingimusel et kõik seisundid $S_u, u \in I$, on fikseeritud. Iga $u \in I$ korral nimetame üksiku vaatluse X_u tinglikku jaotust, tingimusel et seisund $S_u = s_u$ on fikseeritud, *emissioonijaotuseks*. Emissioonijaotuse tihedusfunktsiooni (või tõenäosusfunktsiooni) tähistame f_{s_u} . Kui $(\bar{\mathbf{S}}, \bar{\mathbf{X}})$ on varjatud Markovi välja mudel, siis vastavalt definitsiooni 11 teisele ja kolmandale tingimusele iga mittetühja $J \subset I$ korral

$$p(\bar{\mathbf{x}}_J \mid I) = \prod_{u \in J} f_{s_u}(x_u).$$

Vektorite $\bar{\mathbf{X}}_J$ ja $\bar{\mathbf{S}}$ ühisjaotuse tiheduse kohal $(\bar{\mathbf{x}}_J, \bar{s})$ (defineerime ja) tähistame

$$p(\bar{\mathbf{x}}_J, I) := p(\bar{\mathbf{x}}_J \mid I)p(I),$$

kui $p(I) > 0$, ja $p(\bar{\mathbf{x}}_J, I) := 0$, kui $p(I) = 0$. Vektori $\bar{\mathbf{X}}_J$ marginaaljaotuse tihedus on nüüd

$$p(\bar{\mathbf{x}}_J) := \sum_{\bar{\mathbf{z}}} p_{\bar{\mathbf{z}}}(\bar{\mathbf{x}}_J, I)$$

ja mittetühjade $J_1, J_2 \subset I$ korral vektorite $\bar{\mathbf{X}}_{J_1}$ ja $\bar{\mathbf{S}}_{J_2}$ ühisjaotuse tihedus kohal $(\bar{\mathbf{x}}_{J_1}, \bar{s}_{J_2})$ on

$$p(\bar{\mathbf{x}}_{J_1}, J_2) := \sum_{\bar{\mathbf{z}}: \bar{\mathbf{z}}_{J_2} = \bar{s}_{J_2}} p_{\bar{\mathbf{z}}}(\bar{\mathbf{x}}_J, I).$$

Olgu $J_1, J_2 \subset I$ mittetühjad ja vaatlused $\bar{\mathbf{x}}$ ning varjatud konfiguratsioon \bar{s} fikseeritud. Ühistiheduse $p(\bar{\mathbf{x}}_{J_1}, J_2)$ ja marginaaltiheduse $p(\bar{\mathbf{x}}_{J_1})$ ning -tõenäosuse $p(J_2)$ abil defineerime

$$p(J_2 \mid \bar{\mathbf{x}}_{J_1}) := \frac{p(\bar{\mathbf{x}}_{J_1}, J_2)}{p(\bar{\mathbf{x}}_{J_1})},$$

$$p(\bar{\mathbf{x}}_{J_1} \mid J_2) := \frac{p(\bar{\mathbf{x}}_{J_1}, J_2)}{p(J_2)},$$

kui vaid $p(J_2) > 0$ ja $p(\bar{\mathbf{x}}_{J_1}) > 0$. Kui $J_1, J_2, J_3 \subset I$ on mittetühjad, siis sarnaselt defineerime

$$p(J_2 \mid \bar{\mathbf{x}}_{J_1}, J_3) := \frac{p(\bar{\mathbf{x}}_{J_1}, J_2 \cup J_3)}{p(\bar{\mathbf{x}}_{J_1}, J_3)},$$

$$p(\bar{\mathbf{x}}_{J_1}, J_2 \mid J_3) := \frac{p(\bar{\mathbf{x}}_{J_1}, J_2 \cup J_3)}{p(J_3)},$$

kui vaid $p(\bar{\mathbf{x}}_{J_1}, J_3) > 0$ ja $p(J_3) > 0$. Kokkuvõttes oleme loonud suuruste p süsteemi, mis käituvad rehendamisel üsna sarnaselt vastavate sündmuste tõenäosustele, kuid formaalselt ei ole tõenäosused (vaid tihedus- ja tõenäosusfunktsioonid).

Definitsiooni 11 teine ja kolmas tingimus nõuavad, et iga vaatlus võib sõltuda *ainult* "vahetult tema all" paiknevast varjatud seisundist. Kui juhuslik vektor $\bar{\mathbf{X}}$ on diskreetne, siis saame varjatud Markovi välja mudelit vaadelda (lihtsalt) Markovi

väljana, kui vektori $\bar{\mathbf{X}}$ indeksid I nummerdada ümber indeksiteks $I_+ = \{n+1, n+2, \dots, 2n\}$, nii et vanale indeksile u vastab uus $u+n$, valida indeksite hulga $I = \{1, 2, \dots, n\}$ asemel uueks indeksite hulgaks $I' := I \cup I_+$, naabruskondade süsteemiks $N' := \{\{\mathcal{N}_u \cup \{u+n\}\}_{u \in I}, \{u-n\}_{u=n+1}^{2n}\}$ (see uus naabruskondade süsteem jätab alles vanad naabruskonnad ja lisab iga vaatluse ainsaks naabriks vastava elemendi varjatud kihis) ja juhuslikuks vektoriks $\bar{\mathbf{S}}' := (\bar{\mathbf{S}}, \bar{\mathbf{X}})$ ühise seisundite ruumiga $K' := K \cup \mathcal{X}$. Muuhulgas võimaldab see vaadata varjatud Markovi mudelit Markovi väljana puul, millele on oma käsitluses tähelepanu pööranud Roch (2007).

Varjatud Markovi välja mudeli varjatud kihi tinglikustamine konkreetsetele vaatlustele ei riku ära varjatud kihi globaalset Markovi omadust, täpselt nagu tavaliste varjatud Markovi mudelite korralgi. Suhtudes varjatud seisunditesse kui vaatluste jaotuse parameetritesse, siis Bayesi statistika keeles võib mõelda, et globaalset Markovi omadust rahuldavad jaotused on kaasjaotusteks vaatluste jaotusele (kui vaid vaatluste jaotus rahuldab definitsiooni 11 tingimusi 2 ja 3): nii eeljaotus $p(I)$ kui ka järeljaotus $p(I | \bar{\mathbf{x}})$ kuuluvad globaalset Markovi omadust rahuldavate jaotuste perre.

Väide 8. Kui vektoril $\bar{\mathbf{S}}$ on globaalne Markovi omadus, siis ka vektori $\bar{\mathbf{S}}$ tinglikul jaotusel, tingimusel et vaatlused $\bar{\mathbf{X}}$ on fikseeritud, on globaalne Markovi omadus.

Tõestus. Olgu $J \subset I$ mittetühi ja olgu $\bar{\mathbf{x}} \in \mathcal{X}^n$ ja $\bar{\mathbf{s}} \in K^n$ sellised, et $p(\bar{\mathbf{x}}, I \setminus J) > 0$. Paneme tähele, et iga $\bar{\mathbf{z}} \in K^n$ korral

$$\begin{aligned} p_{\bar{\mathbf{z}}}(\bar{\mathbf{x}}, J | I \setminus J) &= p_{\bar{\mathbf{z}}}(J | I \setminus J) p_{\bar{\mathbf{z}}}(\bar{\mathbf{x}} | I) \\ &= p_{\bar{\mathbf{z}}}(J | \partial J) \prod_{u \in J} f_{z_u}(x_u) \prod_{u \in I \setminus J} f_{z_u}(x_u), \end{aligned} \quad (10)$$

kui vaid $p_{\bar{\mathbf{z}}}(I \setminus J) > 0$. Esimene võrdus tuleb suuruste $p_{\bar{\mathbf{z}}}$ definitsioonidest, teise võrduse juures kasutasime $\bar{\mathbf{S}}$ globaalset Markovi omadust ja asjaolu, et $(\bar{\mathbf{S}}, \bar{\mathbf{X}})$ on varjatud Markovi välja mudel. Analoogiliselt

$$p_{\bar{\mathbf{z}}}(\bar{\mathbf{x}}_J, J | \partial J) = p_{\bar{\mathbf{z}}}(J | \partial J) \prod_{u \in J} f_{z_u}(x_u). \quad (11)$$

Nüüd

$$\begin{aligned} p(J | \bar{\mathbf{x}}, I \setminus J) &= \frac{p(\bar{\mathbf{x}}, J | I \setminus J)}{\sum_{\bar{\mathbf{z}} \in K^n: \bar{\mathbf{z}}_{I \setminus J} = \bar{\mathbf{s}}_{I \setminus J}} p_{\bar{\mathbf{z}}}(\bar{\mathbf{x}}, J | I \setminus J)} \\ &= \frac{p(J | \partial J) \prod_{u \in J} f_{z_u}(x_u)}{\sum_{\bar{\mathbf{z}} \in K^n: \bar{\mathbf{z}}_{I \setminus J} = \bar{\mathbf{s}}_{I \setminus J}} p_{\bar{\mathbf{z}}}(J | \partial J) \prod_{u \in J} f_{z_u}(x_u)} \\ &= \frac{p(\bar{\mathbf{x}}_J, J | \partial J)}{\sum_{\bar{\mathbf{z}} \in K^n: \bar{\mathbf{z}}_{I \setminus J} = \bar{\mathbf{s}}_{I \setminus J}} p_{\bar{\mathbf{z}}}(\bar{\mathbf{x}}_J, J | \partial J)} \\ &= p(J | \bar{\mathbf{x}}, \partial J) \end{aligned}$$

Esimene võrdus tuleb suuruste p definitsioonidest. Teise võrduse juures kasutasime võrdust 10 ja taandasime nii lugejast kui nimetajast teguri $\prod_{u \in I \setminus J} f_{s_u}(x_u)$, arvestades et $u \in I \setminus J$ korral $s_u = z_u$. Kolmas võrdus tuleb võrdusest 11. Neljas võrdus tuleb jällegi suuruste p definitsioonidest. \square

Järgnevalt põhjendame analoogilise väite Gibbsi jaotuste jaoks.

Väide 9. Olgu $\bar{\mathbf{S}} = (S_u)_{u \in I}$ ja $\bar{\mathbf{X}} = (X_u)_{u \in I}$ sellised juhuslikud vektorid, et $\bar{\mathbf{S}}$ on diskreetne juhuslik vektor Gibbsi jaotusega graafil (I, N) ja hulgal K^n antud Gibbsi potentsiaali $V_C, C \subset I$, suhtes ning $\bar{\mathbf{X}}$ rahuldab definitsiooni 11 tingimusi 2 ja 3. Sel juhul vektori $\bar{\mathbf{S}}$ tinglik jaotus, tingimusel et $\bar{\mathbf{X}}$ on fikseeritud, on samuti Gibbsi jaotus, kusjuures

$$p_{\bar{\mathbf{S}}}(I \mid \bar{\mathbf{x}}) \propto \exp \left(U(\bar{\mathbf{s}}) + \sum_{u \in I} \ln f_{s_u}(x_u) \right),$$

kus

$$U(\bar{\mathbf{s}}) := \sum_{C \in \mathcal{C}} V_C(\bar{\mathbf{s}}).$$

Tõestus. Fikseeritud $\bar{\mathbf{x}}$ korral, vastavalt suuruste p definitsioonidele, kehtib

$$p(I \mid \bar{\mathbf{x}}) = \frac{p(\bar{\mathbf{x}}, I)}{p(\bar{\mathbf{x}})} \propto p(\bar{\mathbf{x}}, I) = p(I)p(\bar{\mathbf{x}} \mid I).$$

Järelikult

$$\begin{aligned} p(I \mid \bar{\mathbf{x}}) &\propto p(I)p(\bar{\mathbf{x}} \mid I) \\ &= p(I) \prod_{u \in I} f_{s_u}(x_u) \\ &\propto \exp(U(\bar{\mathbf{s}})) \exp \left(\ln \prod_{u \in I} f_{s_u}(x_u) \right), \end{aligned}$$

nagu vaja. Tegu on tõepoolest Gibbsi jaotusega: defineerime funktsioonide kogumi $V_C^*, C \subset I$, muutes vaid ühe-elementilistele klikkidele vastavaid funktsioone:

$$\begin{aligned} V_C^*(\bar{\mathbf{s}}) &:= V_C(\bar{\mathbf{s}}), & \text{kui } |C| \neq 1, \\ V_C^*(\bar{\mathbf{s}}) &:= V_C(\bar{\mathbf{s}}) + \ln f_{s_u}(x_u), & \text{kui } C = \{u\}, u \in I. \end{aligned}$$

On näha, et see funktsioonide kogum tõepoolest on Gibbsi potentsiaal samal graafil (I, N) ja samal hulgal K^n , kusjuures $p(I \mid \bar{\mathbf{x}})$ avaldub selle Gibbsi potentsiaali kaudu. Paneme tähele, et eeltoodud võrduses on vaatlused $\bar{\mathbf{x}}$ fikseeritud ja Gibbsi potentsiaal $V_C^*, C \subset I$, selles mõttes neist enam ei sõltu (kuigi x_u on avaldises kirjas). \square

Varjatud seisundite kui vaatluste parameetrite hindamise kontekstis nimetatakse suurust $-U(\bar{\mathbf{s}})$ (konfiguratsiooni $\bar{\mathbf{s}}$) eelenergiaks (i.k. *prior energy*) ja suurust $-U(\bar{\mathbf{s}}) - \sum_{u \in I} \ln f_{s_u}(x_u)$ järelenergiaks (i.k. *posterior energy*). Paneme tähele, et mida väiksem energia, seda suurem tõenäosus.

Lõpuks toome ka ära ühe puudel kehtiva kasuliku tegurdamisomaduse.

Väide 10. Kui $(\bar{\mathbf{S}}, \bar{\mathbf{X}})$ on varjatud Markovi välja mudel puul (I, N) , siis iga $\bar{\mathbf{s}} \in K^n$ ja iga $\bar{\mathbf{x}} \in \mathcal{X}^n$ korral kehtib

$$p(\bar{\mathbf{x}}, I) = p(1) \prod_{u>1} p(u \mid \rho(u)) \prod_{u \in I} f_{s_u}(x_u), \quad (12)$$

kui vaid need tinglikud tõenäosused on defineeritud.

Tõestus. Olgu $\bar{\mathbf{s}} \in K^n$ ja $\bar{\mathbf{x}} \in \mathcal{X}^n$ sellised, et võrduses 12 toodud tinglikud tõenäosused on defineeritud. Kui $\bar{\mathbf{s}}$ on selline, et $p(I) = 0$, siis ka $p(\bar{\mathbf{x}}, I) = 0$. Kui $\bar{\mathbf{s}}$ on selline, et $p(I) > 0$, siis on defineeritud kõik tinglikud tõenäosused $p(u \mid \rho(u))$ ja tinglik tihedus $p(\bar{\mathbf{x}} \mid I)$ ning kehtib $p(\bar{\mathbf{x}}, I) = p(I)p(\bar{\mathbf{x}} \mid I)$. Kuna $\bar{\mathbf{S}}$ rahuldab globaalset Markovi omadust, siis vastavalt väitele 5 kehtib

$$p(I) = p(1) \prod_{u>1} p(u \mid \rho(u)).$$

Kuna $(\bar{\mathbf{S}}, \bar{\mathbf{X}})$ on varjatud Markovi välja mudel puul (I, N) , siis

$$p(\bar{\mathbf{x}} \mid I) = \prod_{u \in I} f_{s_u}(x_u).$$

Kaks eelnevat võrdust kokku annavadki vajaliku tulemuse. □

Eelnevast väitest näeme, et puu peal antud varjatud Markovi välja mudeli määravad üheselt algjaotus $(p_j(1))_{j \in K}$, üleminekutõenäosused $(p_{j|i}(u \mid \rho(u)))_{j \in K, i \in K, u > 1}$ ja emissioonijaotuste tihedused $(f_{s_u})_{s_u \in K, u \in I}$.

3 Segmenteerimine ja risk

3.1 Segmenteerimisülesande tutvustus

Käesolevas peatükis tegeleme varjatud seisundite ära arvamiseiga antud vaatluste alusel. Varjatud Markovi (välja) mudelite kontekstis nimetatakse seda *segmenteerimiseks* – varjatud protsessi jagamiseks erinevas seisundis olevateks juppideks. Just segmenteerimisülesanne on tihti esmase huvi objektiks varjatud Markovi (välja) mudelite juures (Lember jt., 2011) ja just segmenteerimises avaldub tihti nende mudelite praktiline väärtus, näiteks erinevate DNA regioonide tuvastamine järjestusandmete põhjal (Peshkin ja Gelfand, 1999), eellasolekute rekonstrueerimine fülogeneetilistes uuringutes (Bykova jt., 2013) ja erinevate alade-objektide ära tundmine värvipiltidel (Xiang-yang jt., 2015). Järgnevad lõigud on kirjutatud Lember ja Koloydenko (2014) alusel.

Erinevaid lahendusi segmenteerimisülesandele on aga mitmeid. Üks levinud viis on leida selline seisundite konfiguratsioon $\bar{\mathbf{s}} \in K^n$ (nüüd ja edaspidi nimetame igasuguse segmenteerimise tulemust $\bar{\mathbf{s}} \in K^n$ *joonduseks*), et fikseeritud vaatluste $\bar{\mathbf{x}} \in \mathcal{X}^n$ korral vaatluste ja varjatud seisundite ühistihedus $p_{\bar{\mathbf{s}}}(\bar{\mathbf{x}}, I)$ oleks maksimaalne. Mõeldes varjatud seisunditele kui vaatluste $\bar{\mathbf{X}}$ jaotuse parameetritele, on see lähenemine analoogiline suurima tõepära hinnanguga. Varjatud Markovi mudelite kontekstis nimetatakse seda joondust *Viterbi joonduseks*. Nimi tuleb Andrew Viterbi nime kandvast dünaamilise planeerimise algoritmist, mida eelkirjeldatud joonduse leidmiseks kasutada saab. Bayesi statistika keeles aga võime mõelda fikseeritud mudelist $p(I)$ kui varjatud kihi $\bar{\mathbf{S}}$ eeljaotusest ja tinglikust jaotusest $p(I | \bar{\mathbf{x}}) \propto p(\bar{\mathbf{x}}, I)$ kui varjatud kihi järeljaotusest – sellest inspireerituna nimetatakse suuruse $p(\bar{\mathbf{x}}, I)$ maksimiseerimise tulemust ka *MAP joonduseks* (i.k. *maximum a posteriori*). Kui $\bar{\mathbf{S}}$ on juhtumisi Gibbsi jaotusega, siis väite 9 tõttu langeb MAP joonduse leidmine kokku järeleenergia minimiseerimisega, sest

$$p_{\bar{\mathbf{s}}}(I | \bar{\mathbf{x}}) \propto \exp \left(- \left(-U(\bar{\mathbf{s}}) - \sum_{u \in I} \ln f_{s_u}(x_u) \right) \right).$$

Näide 7 (Gaussi müraga rikitud Gibbsi jaotus, Li, 2009 lk 15). Olgu $\bar{\mathbf{S}}$ Gibbsi jaotusega graafil (I, N) antud Gibbsi potentsiaali $V_C, C \subset I$, suhtes. Juhuslik vektor $\bar{\mathbf{S}}$ võib olla pidev, seisundite ruumiga \mathbb{R} . Oletame, et vektori $\bar{\mathbf{S}}$ realisatsioon jääb aga varjatuks ja selle asemel näeme Gaussi müraga saastunud versiooni. Täpsemalt, iga $u \in I$ korral S_u asemel näeme hoopis arvu $S_u + \varepsilon_u$, kus $\varepsilon_u \sim N(0, \sigma^2)$, $u \in I$, on omavahel sõltumatud ja vektorist $\bar{\mathbf{S}}$ sõltumatud juhuslikud suurused. Soovime siiski prognoosida varjatud seisundeid $\bar{\mathbf{S}}$.

Tähistame $X_u := S_u + \varepsilon_u, u \in I$. Vastavalt väitele 9 on vektori $\bar{\mathbf{S}}$ tinglik jaotus, tingimusel et vaatlused $\bar{\mathbf{X}}$ on fikseeritud, samuti Gibbsi jaotus ja kehtib

$$p_{\bar{\mathbf{s}}}(I | \bar{\mathbf{x}}) \propto \exp \left(U(\bar{\mathbf{s}}) + \sum_{u \in I} \ln f_{s_u}(x_u) \right)$$

Antud juhul on emissioonijaotused $X_u | S_u = s_u \sim N(s_u, \sigma^2)$, seega

$$f_{s_u}(x_u) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(s_u - x_u)^2}{2\sigma^2} \right)$$

ja

$$p_{\bar{s}}(I | \bar{\mathbf{x}}) \propto \exp \left(U(\bar{s}) - \frac{1}{2\sigma^2} \sum_{u \in I} (s_u - x_u)^2 \right).$$

Nagu juba märgitud, saame nüüd leida MAP joonduse avaldise

$$-U(\bar{s}) + \frac{1}{2\sigma^2} \sum_{u \in I} (s_u - x_u)^2$$

minimiseerimise teel. Näeme, et ühelt poolt oleks kõige kasulikum pakkuda igaks varjatud seisundiks s_u konkreetset vaadeldud väärtust x_u : see minimiseeriks ruutude summat $\sum_{u \in I} (s_u - x_u)^2$. Teiselt poolt aga oleks siis ilmselgelt tegu mudeli ülesobitumisega. Selle vastu aitabki liige $U(\bar{s})$, millel on “oma nägemus” optimaalsest konfiguratsioonist ja mis võiks näiteks liiga keeruliste konfiguratsioonide tõenäosuse alla kaaluda.

Alternatiivselt võime otsida sellist joondust $\bar{s} \in K^n$, et iga $u \in I$ korral selle konkreetse varjatud kihi elemendi “tõepära” $p(u | \bar{\mathbf{x}})$ oleks maksimaalne. Nimetame seda joondust *PMAP joonduseks* (i.k. *pointwise maximum a posteriori*). Just see joondus tagab keskmiselt vähima vigade arvu (tõestus allpool). Küll aga on see joondus “pime” konteksti suhtes: kui varjatud kihi kohta on näiteks teada, et teatud konfiguratsioonid on põhimõtteliselt võimatud, siis PMAP meetod ei lase ennast sellest häirida ja võib joondusena tagastada võimatu konfiguratsiooni.

Näide 8 (MAP ja PMAP võrdlus). Olgu $\bar{\mathbf{S}}$ Markovi väli nagu näites 4. Olgu iga $u \in I$ korral emissioonijaotuseks Bernoulli jaotus $X_u | S_u = 1 \sim B(1, 0.6)$ ja $X_u | S_u = 0 \sim B(1, 0.1)$. Võime mõelda, et näeme rikitud versiooni vektorist $\bar{\mathbf{S}}$, kusjuures varjatud seisund 1 saab muudetud tõenäosusega 0.4 ja varjatud seisund 0 saab muudetud tõenäosusega 0.1. Meie käsutusse on sattunud vaatlused $\bar{\mathbf{x}} = (x_u)_{u \in I} := (1, 0, 0, 1)$. Eesmärgiks on ära arvata see varjatud kihi konfiguratsioon, mis need vaatlused “tekitas”.

Leiame esmalt MAP joonduse. Selleks võtame vaatluse alla kõik võimalikud varjatud konfiguratsioonid $\bar{s} \in \{0, 1\}^4$ ja valime neist välja sellise, mille puhul $p(I, \bar{\mathbf{x}})$ on suurim. Paneme tähele, et kui $p(I) = 0$, siis $p(\bar{\mathbf{x}}, I) = 0$. Kui $p(I) > 0$, siis vastavalt varjatud Markovi välja mudeli eeldusele vaatluste tingliku sõltumatuse kohta

$$p(\bar{\mathbf{x}}, I) = p(I)p(\bar{\mathbf{x}} | I) = p(I) \prod_{u=1}^4 f_{s_u}(x_u).$$

Tabelist 2 näeme, et suurima ühistihedusega on konfiguratsioon $(1, 1, 1, 1)$, järelikult see ongi MAP joondus.

Leiame nüüd PMAP joonduse. Selleks vaatame iga kohta $u \in I$ eraldi ja valime igas kohas seisunditest $s_u \in \{0, 1\}$ välja selle, mille puhul $p(u | \bar{\mathbf{x}})$ on suurem. Fikseeritud vaatluste korral $p(u | \bar{\mathbf{x}}) \propto p(\bar{\mathbf{x}}, u)$ ehk need kaks suurust saavutavad maksimumi sama seisundi s_u korral. Kuna iga \bar{s} korral

$$p_{\bar{s}}(\bar{\mathbf{x}}, u) = \sum_{\bar{z} \in \{0,1\}^4: z_u = s_u} p_{\bar{z}}(\bar{\mathbf{x}}, I),$$

Tabel 2: Varjatud kihi (eel)jaotus ja vaatluste ning varjatud konfiguratsioonide ühistõenäosused näites 8.

s_1	s_2	s_3	s_4	$p_{\bar{s}}(I)$	$p_{\bar{s}}(\bar{\mathbf{x}}, I)$
0	0	0	0	8^{-1}	0.0010125
1	0	0	0	8^{-1}	0.006075
1	1	0	0	8^{-1}	0.0027
1	1	1	0	8^{-1}	0.0012
1	1	1	1	8^{-1}	0.0072
0	1	1	1	8^{-1}	0.0012
0	0	1	1	8^{-1}	0.0027
0	0	0	1	8^{-1}	0.006075
0	1	0	0	0	0
0	0	1	0	0	0
1	0	1	0	0	0
0	1	1	0	0	0
0	1	0	1	0	0
1	0	0	1	0	0
1	0	1	1	0	0
1	1	0	1	0	0

Tabel 3: Suurused $p(\bar{\mathbf{x}}, u)$ näites 8.

	$u = 1$	$u = 2$	$u = 3$	$u = 4$
$s_u = 0$	0.0109875	0.0158625	0.0158625	0.0109875
$s_u = 1$	0.017175	0.0123	0.0123	0.017175

siis saame kasutada tabelit 2. Summeerides sobivaid arve selle tabeli parempoolses veerus, leiame iga $u \in I$ ja $s_u \in \{0, 1\}$ korral $p_{\bar{s}}(\bar{\mathbf{x}}, u)$ ja kirjutame need arvud tabelisse 3. Sealt näeme, et PMAP joondus on $(1, 0, 0, 1)$. Tabelist 2 aga näeme, et vektor $\bar{\mathbf{S}}$ ei saa olla konfiguratsioonis $(1, 0, 0, 1)$, selle tõenäosus on null.

Näeme, et PMAP probleemiks on joondused, mille tõenäosus on null. Sellistel konfiguratsioonidel on ka oma nimi: kui vaatlused $\bar{\mathbf{x}}$ on fikseeritud ja $p(\bar{\mathbf{x}}) > 0$, siis konfiguratsiooni $\bar{\mathbf{s}} \in K^n$ nimetame *lubamatuks* (i.k. *inadmissible*), kui $p_{\bar{s}}(I | \bar{\mathbf{x}}) = 0$ (Lember ja Koloydenko, 2014 järgi). On segmenteerimismeetodeid, mis sisuliselt üritavad võimalikult sarnaneda PMAP joondusele, välistades siiski nullilise tõenäosusega joonduse. Ometi võivad ka need anda väga madala tõenäosusega (kuid siiski lubatavaid) joondusi. Samuti on võimalik MAP ja PMAP joondusi mitmel moel kombineerida (Lember ja Koloydenko, 2014).

3.1.1 Klassifitseerimisteooria põhimõisteid

Võtame hetkeks veidi üldisema vaatekoha, et tuua sisse tehisõppe valdkonnast pärit ja edasises mõttearenduses vajalikke mõisteid. Alapeatükk on kirjutatud Lember (2013) järgi.

Olgu $\bar{\mathbf{S}}$ ja $\bar{\mathbf{X}}$ samal tõenäosusruumil defineeritud juhuslikud vektorid. Täpsemalt, olgu $\bar{\mathbf{S}} = (S_u)_{u \in I}$ diskreetne juhuslik vektor, mille väärtused on lõplikus hulgas K^n , ja $\bar{\mathbf{X}} = (X_u)_{u \in I}$ juhuslik vektor, mille võimalike väärtuste hulk on \mathcal{X}^n , kus

$\mathcal{X} \subset \mathbb{R}^d$. Vektori $\bar{\mathbf{S}}$ realisatsiooni tõlgendame kui *klassi*, mis antud juhul on mitmemõõtmeline, ja vektori $\bar{\mathbf{X}}$ realisatsiooni tõlgendame kui *vaatlust*, mis praegu on ka mitmemõõtmeline.

Definitsioon 12. *Klassifitseerijaks* nimetame funktsiooni

$$g : \mathcal{X}^n \rightarrow K^n.$$

Vaatluse $\bar{\mathbf{x}} \in \mathcal{X}^n$ põhjal klassi $\bar{\mathbf{z}} \in K^n$ prognoosimist nimetatakse *klassifitseerimiseks*. Erijuhul, kui $(\bar{\mathbf{S}}, \bar{\mathbf{X}})$ moodustab varjatud Markovi (välja) mudeli, siis klassifitseerimist nimetatakse tihti *segmenteerimiseks*. Iga $\bar{\mathbf{x}} \in \mathcal{X}^n$ korral vektori $g(\bar{\mathbf{x}})$ komponenti indeksiga u tähistame $g(\bar{\mathbf{x}})_u$.

Definitsioon 13. *Kaofunktsiooniks* nimetame funktsiooni

$$L : K^n \times K^n \rightarrow [0, \infty].$$

Arvu $L(\bar{\mathbf{s}}, \bar{\mathbf{z}})$ tõlgendame kui kahju (või kadu), mis tekib, kui klassi $\bar{\mathbf{s}}$ pidada klassiks $\bar{\mathbf{z}}$. Loomulik on kasutada selliseid kaofunktsioone, mille korral $L(\bar{\mathbf{s}}, \bar{\mathbf{s}}) = 0$ iga $\bar{\mathbf{s}}$ jaoks.

Definitsioon 14. Klassifitseerija g *riskiks* nimetame keskväärtust

$$R(g) := EL(\bar{\mathbf{S}}, g(\bar{\mathbf{X}})).$$

Iga vaatluse $\bar{\mathbf{x}}$ korral defineerime klassifitseerija g *tingliku riski* kui arvu

$$R(g | \bar{\mathbf{x}}) := \sum_{\bar{\mathbf{z}} \in K^n} L(\bar{\mathbf{z}}, g(\bar{\mathbf{x}})) p_{\bar{\mathbf{z}}}(I | \bar{\mathbf{x}}).$$

Klassifitseerijat, mis iga vaatluse $\bar{\mathbf{x}}$ korral minimiseerib tinglikku riski $R(g | \bar{\mathbf{x}})$ üle kõigi klassifitseerijate g , nimetatakse *Bayesi klassifitseerijaks*.

Kuna tinglike riskidega tegeledes on vaatlused $\bar{\mathbf{x}}$ niikuinii tähistuses näha, siis võime samastada klassifitseerija g tema antava joondusega $\bar{\mathbf{s}} = g(\bar{\mathbf{x}})$. Samal põhjusel võib rääkida klassifitseerija tingliku riski asemel joonduse (või konfiguratsiooni) tinglikust riskist.

Väide 11. Bayesi klassifitseerija risk on väiksem või võrdne mistahes muu klassifitseerija riskiga.

Tõestus. Tõestuse võib leida näiteks Lember (2013) lk 10. □

3.2 Erinevaid kaofunktsioone ja riske varjatud Markovi välja mudelitele

Tutvume segmenteerimisülesande püstitamise mitmekesiste võimalustega. Otsime küll vaatlustele kõige paremini sobivat konfiguratsiooni, kuid *mis mõttes* kõige paremini sobivat? Sellele küsimusele aitavadki ühtses raamistikus täpselt vastata eelnevalt tutvustatud kaofunktsioonid ja (tinglikud) riskid, mis saavad konkreetseteks headuse kriteeriumiteks. Käesolevas alapeatükis tutvume levinud ja vähem levinud kaofunktsioonide ja vastavate (tinglike) riskidega, sealhulgas näitame, kuidas läbi kaofunktsioonide jõuda MAP ja PMAP joondusteni. Järgnevas alapeatükis

konstrueerime klassifitseerijad, mis puude peal antud varjatud Markovi välja mude-
lite korral minimiseerivad neid riske. Käesolev alapeatükk on kirjutatud Lember ja
Koloydenko (2014) eeskujul.

Tavalisim kaofunktsioon on *sümmeetriline kaofunktsioon*, mille käesolevas tähistame
 L_∞ , ja mis on defineeritud kui

$$L_\infty(\bar{\mathbf{s}}, \bar{\mathbf{z}}) := I_{\bar{\mathbf{s}} \neq \bar{\mathbf{z}}},$$

kus I on indikaatorfunktsioon. See karistab kõiki valesid konfiguratsioone, kuitahes
tõelähedasi või tõekaugeid, ühepalju. Mistahes klassifitseerija g tinglikuks riskiks
mistahes vaatluse $\bar{\mathbf{x}}$ korral on sel juhul

$$\begin{aligned} R_\infty(g | \bar{\mathbf{x}}) &:= \sum_{\bar{\mathbf{z}} \in K^n} L_\infty(\bar{\mathbf{z}}, g(\bar{\mathbf{x}})) p_{\bar{\mathbf{z}}}(I | \bar{\mathbf{x}}) \\ &= \sum_{\bar{\mathbf{z}} \in K^n: \bar{\mathbf{z}} \neq g(\bar{\mathbf{x}})} 1 \cdot p_{\bar{\mathbf{z}}}(I | \bar{\mathbf{x}}) \\ &= 1 - p_{g(\bar{\mathbf{x}})}(I | \bar{\mathbf{x}}). \end{aligned}$$

Näha on, et seda riski minimiseerib klassifitseerija, mis maksimiseerib suurust $p_{g(\bar{\mathbf{x}})}(I | \bar{\mathbf{x}}) \propto p_{g(\bar{\mathbf{x}})}(\bar{\mathbf{x}}, I)$. Järelikult R_∞ -riskile vastav Bayesi klassifitseerija on selline, mis
annab MAP joondusi. Paneme tähele, et MAP joondused minimiseerivad ka sellist
tinglikku riski:

$$\bar{R}_\infty(g | \bar{\mathbf{x}}) := -\frac{1}{n} \ln p_{g(\bar{\mathbf{x}})}(I | \bar{\mathbf{x}}).$$

Võime valida ka sellise kaofunktsiooni, mis võrdleb tegelikkust ja prognoosi punkt-
haaval:

$$L_1(\bar{\mathbf{s}}, \bar{\mathbf{z}}) := \frac{1}{n} \sum_{u \in I} I_{s_u \neq z_u}.$$

Mistahes klassifitseerija g tinglikuks riskiks mistahes vaatluse $\bar{\mathbf{x}}$ puhul on siis

$$\begin{aligned} R_1(g | \bar{\mathbf{x}}) &:= \sum_{\bar{\mathbf{z}} \in K^n} L_1(\bar{\mathbf{z}}, g(\bar{\mathbf{x}})) p_{\bar{\mathbf{z}}}(I | \bar{\mathbf{x}}) \\ &= \sum_{\bar{\mathbf{z}} \in K^n} \left(\frac{1}{n} \sum_{u \in I} I_{z_u \neq g(\bar{\mathbf{x}})_u} \right) p_{\bar{\mathbf{z}}}(I | \bar{\mathbf{x}}) \\ &= \frac{1}{n} \sum_{u \in I} \sum_{z_u \in K} I_{z_u \neq g(\bar{\mathbf{x}})_u} \sum_{\bar{\mathbf{z}}_{I \setminus u} \in K^{n-1}} p_{\bar{\mathbf{z}}}(I | \bar{\mathbf{x}}) \\ &= \frac{1}{n} \sum_{u \in I} \sum_{z_u \in K} I_{z_u \neq g(\bar{\mathbf{x}})_u} p_{z_u}(u | \bar{\mathbf{x}}) \\ &= \frac{1}{n} \sum_{u \in I} (1 - p_{g(\bar{\mathbf{x}})}(u | \bar{\mathbf{x}})) \\ &= 1 - \frac{1}{n} \sum_{u \in I} p_{g(\bar{\mathbf{x}})}(u | \bar{\mathbf{x}}). \end{aligned}$$

Näeme, et eeltoodud tingliku riski minimiseerimine on samaväärne suuruse $p_{g(\bar{\mathbf{x}})}(u | \bar{\mathbf{x}})$
maksimiseerimisega iga $u \in I$ korral – järelikult R_1 -riski minimiseerib PMAP joon-
dus. Paneme tähele, et kaofunktsioon L_1 kujutab endast valesti klassifitseeritud

komponentide S_u osakaalu kogu vektori $\bar{\mathbf{S}}$ seas. Arvestades nüüd riski definitsiooni $R(g) = EL(\bar{\mathbf{S}}, g(\bar{\mathbf{X}}))$ ja teadmist, et riski minimiseerib see klassifitseerija, mis minimiseerib tinglikku riski iga vaatluse korral (antud juhul siis PMAP), on näha, et PMAP joondus teeb kõigist võimalikest segmenteerimismeetoditest keskmiselt kõige vähem vigu.

Summa $\sum_{u \in I} p_{\bar{\mathbf{s}}}(u | \bar{\mathbf{x}})$ maksimiseerimine üle kõigi $\bar{\mathbf{s}} \in K^n$ on samaväärne iga liidetava $p_{s_u}(u | \bar{\mathbf{x}})$, $u \in I$, maksimiseerimisega üle $s_u \in K$. Kuna igal u , $\bar{\mathbf{x}}$ ja s_u korral $p_{s_u}(u | \bar{\mathbf{x}}) \geq 0$, siis eelnev on omakorda samaväärne korrutise $\prod_{u \in I} p_{s_u}(u | \bar{\mathbf{x}})$ maksimiseerimisega üle kõigi $\bar{\mathbf{s}} \in K^n$. Järelikult PMAP joondus minimiseerib ka järgnevat tinglikku riski:

$$\bar{R}_1(g | \bar{\mathbf{x}}) := -\frac{1}{n} \sum_{u \in I} \ln p_{g(\bar{\mathbf{x}})}(u | \bar{\mathbf{x}}).$$

Eelnevad riskid saame kokku võtta esmapilgul võib-olla mõneti kunstlikuna tunduva üldistuse abil, defineerides Lember ja Koloydenko (2014) eeskujul *kombineeritud riski* ehk *hübriidriski*

$$\bar{R}_{C_1, C_2}(g | \bar{\mathbf{x}}) := C_1 \bar{R}_1(g | \bar{\mathbf{x}}) + C_2 \bar{R}_\infty(g | \bar{\mathbf{x}}), \quad (13)$$

kus $C_1, C_2 \geq 0$ ja vähemalt üks neist on nullist erinev. Paneme tähele, et erijuhul $C_1 = 0$ minimiseerib kombineeritud riski MAP joondus ja erijuhul $C_2 = 0$ minimiseerib kombineeritud riski PMAP joondus. Järgnevas peatükis näeme, et seda riski minimiseerivat joondust on teatud juhul võimalik väga hästi arvutada ja see risk osutub loomulikuks üldistuseks, mis ühendab MAP ja PMAP klassifitseerijad ühtse klassifitseerijate pere äärmuslikeks liikmeteks.

4 Hübriidrisk puudel

4.1 Hübriidalgorithm

Valemiga 13 defineeritud hübriidrisk on kindlaks optimaalsuskriteeriumiks, mille järgi valida parim klassifitseerija. Kui $(\bar{\mathbf{S}}, \bar{\mathbf{X}})$ on varjatud Markovi välja mudel puul (I, N) , siis on võimalik seda Bayesi klassifitseerijat ka reaalselt rakendada, täpsemalt siis vastavat joondust puu tippude arvu n suhtes lineaarses ajas välja arvutada. See toimub teatud dünaamilise planeerimise algoritmi abil, mis on väga sarnane artikli Lember ja Koloydenko (2014) teoreemis 4 kirjeldatule.

Väide 12. Olgu $(\bar{\mathbf{S}}, \bar{\mathbf{X}}) = ((S_u)_{u \in I}, (X_u)_{u \in I})$ varjatud Markovi välja mudel puul (I, N) , mis on määratud teadaoleva algjaotuse $\pi_j := p_j(1), j \in K$, üleminekutõenäosuste $p_{uij} := p_{j|i}(u | \rho(u)), u > 1, i, j \in K$, ja emissioonijaotuste tiheduste $f_{s_u}, u \in I, s_u \in K$, abil. Olgu eelnevalt välja arvatud tõenäosused $p_j(u | \bar{\mathbf{x}})$ iga $u \in I$ ja $j \in K$ korral. Joondus $\bar{\mathbf{s}} = (s_u)_{u \in I} \in K^n$, mis antud vaatluste $\bar{\mathbf{x}}$ korral minimizeerib hübriidriski 13, on leitav järgmise algoritmi abil:

1. iga $u \in I$ ja $j \in K$ jaoks arvuta

$$\gamma_u(j) := C_1 \ln p_j(u | \bar{\mathbf{x}}) + C_2 \ln f_j(x_u),$$

2. iga lehe u ja iga $j \in K$ jaoks arvuta

$$\delta_u(j) := \gamma_u(j),$$

3. iga mitte-lehe u ja $j \in K$ jaoks arvuta

$$\delta_u(j) := \gamma_u(j) + \sum_{v \in \rho^{-1}(u)} \max_{i \in K} (C_2 \ln p_{vji} + \delta_v(i)),$$

4. vali

$$s_1 := \arg \max_{j \in K} (C_2 \ln \pi_j + \delta_1(j))$$

ja

$$s_u := \arg \max_{j \in K} (\delta_u(j) + C_2 \ln p_{u, s_{\rho(u)}, j}).$$

Lisaks täpselt teada olevale mudelile on selle algoritmi rakendamiseks vaja eelnevalt välja arvutada ka suurused $p_j(u | \bar{\mathbf{x}})$ iga $u \in I$ ja $j \in K$ jaoks. Neid arve nimetatakse *silumistõenäosusteks* (i.k. *smoothing probabilities*). Tegu on just nendesamade suurustega, mille järgi koostatakse PMAP joondus (valides iga $u \in I$ jaoks välja selle $j \in K$, mis maksimiseerib silumistõenäosust $p_j(u | \bar{\mathbf{x}})$). Võime mõelda, et tegu on konkreetses kohas $u \in I$ seisundi j nägemise marginaalsete järeltõenäosustega, tingimusel et vaatlused on sellised nagu need parajasti on. Silumistõenäosuseid on võimalik (vaatluseid ja mudeli parameetreid teades) arvutada teatud rekursiivsete algoritmidega, käesolevas töös on kasutatud Durand jt. (2004) lähenemist.

Tõestus väitele 12. Tähistame $\rho(1) := 0$, $s_0 := 0$ ja $p_{1,0,s_1} := \pi_{s_1}$. Iga konfiguratsiooni $\bar{s} \in K^n$ ja iga tipu $u \in I$ jaoks defineerime

$$U(\bar{s}_u, s_{\rho(u)}) := \sum_{v \in I_u} (\gamma_v(s_v) + C_2 \ln p_{v s_{\rho(v)} s_v}).$$

Leiame esialgse minimiseerimisülesandega ekvivalentse, kuid mugavama ülesande:

$$\begin{aligned} & \min_{\bar{s}} (C_1 \bar{R}_1(\bar{s} | \bar{x}) + C_2 \bar{R}_\infty(\bar{s} | \bar{x})) \\ &= \min_{\bar{s}} \left(-\frac{C_1}{n} \sum_{u \in I} \ln p(u | \bar{x}) - \frac{C_2}{n} \ln p(I | \bar{x}) \right) \\ &\Leftrightarrow \max_{\bar{s}} \left(C_1 \sum_{u \in I} \ln p(u | \bar{x}) + C_2 \ln p(\bar{x}, I) \right) \\ &= \max_{\bar{s}} \left(C_1 \sum_{u \in I} \ln p(u | \bar{x}) + C_2 \sum_{u \in I} \left(\ln f_{s_u}(x_u) + \ln p_{u s_{\rho(u)} s_u} \right) \right) \\ &= \max_{\bar{s}} \sum_{u \in I} \left(C_1 \ln p(u | \bar{x}) + C_2 \ln f_{s_u}(x_u) + C_2 \ln p_{u s_{\rho(u)} s_u} \right) \\ &= \max_{\bar{s}} \left(\sum_{u \in \rho^{-1}(1)} U(\bar{s}_u, s_1) + \gamma_1(s_1) + C_2 \ln p_{1,0,s_1} \right), \end{aligned} \quad (14)$$

pidades meeles tähistust $p_{1,0,s_1} := \pi_{s_1}$. Teine võrdus tuleb väitest 10.

Olgu $\bar{s} \in K^n$. Induktsiooniga tipu u sügavuse järgi näitame, et iga $j \in K$ korral

$$\delta_u(j) = \max_{\bar{s}_u: s_u=j} \left(\sum_{v \in \rho^{-1}(u)} U(\bar{s}_v, j) \right) + \gamma_u(j). \quad (15)$$

Kui u on leht ehk alampuu I_u sügavus on 1, siis $\rho^{-1}(u) = \emptyset$ (lehel pole lapsi) ja võrdus 15 kehtib tänu $\delta_u(j)$ definitsioonile: summa on null ja maksimiseerida pole midagi.

Eeldame, et võrdus 15 kehtib kõigi selliste tippude v jaoks, mille sügavus on väiksem kui $d > 1$. Olgu u mistahes tipp, mille sügavus on d – siis tippude $v \in \rho^{-1}(u)$ sügavused on tõepoolest väiksemad kui d . Nüüd, $\delta_u(j)$ definitsiooni tõttu, iga $j \in K$

korral kehtib

$$\begin{aligned}
\delta_u(j) &= \sum_{v \in \rho^{-1}(u)} \max_k (C_2 \ln p_{vjk} + \delta_v(k)) + \gamma_u(j) \\
&= \sum_{v \in \rho^{-1}(u)} \max_k \left(C_2 \ln p_{vjk} + \max_{\bar{s}_v: s_v=k} \left(\sum_{w \in \rho^{-1}(v)} U(\bar{s}_w, k) + \gamma_v(k) \right) \right) + \gamma_u(j) \\
&= \sum_{v \in \rho^{-1}(u)} \max_k \max_{\bar{s}_v: s_v=k} \left(C_2 \ln p_{vjk} + \sum_{w \in \rho^{-1}(v)} U(\bar{s}_w, s_v) + \gamma_v(s_v) \right) + \gamma_u(j) \\
&\quad \text{Paneme tähele, et } p_{vjk} = p_{v, s_{\rho(v)}, s_v}, \text{ kui } s_u = j, \text{ ja kasutame } U(\bar{s}_v, s_{\rho(v)}) \text{ definitsiooni.} \\
&= \sum_{v \in \rho^{-1}(u)} \max_k \max_{\bar{s}_v: s_v=k} U(\bar{s}_v, j) + \gamma_u(j) \\
&= \sum_{v \in \rho^{-1}(u)} \max_{\bar{s}_v} U(\bar{s}_v, j) + \gamma_u(j) \\
&= \max_{\bar{s}_u: s_u=j} \left(\sum_{v \in \rho^{-1}(u)} U(\bar{s}_v, s_u) \right) + \gamma_u(j),
\end{aligned}$$

nagu vaja.

Võrduste 14 ja 15 abil näeme nüüd, et

$$\begin{aligned}
\max_j (\delta_1(j) + C_2 \ln \pi_j) &= \max_j \left(\max_{\bar{s}: s_1=j} \left(\sum_{v \in \rho^{-1}(1)} U(\bar{s}_v, s_1) \right) + \gamma_1(j) + C_2 \ln p_{1,0,j} \right) \\
&= \max_j \max_{\bar{s}: s_1=j} \left(\sum_{v \in \rho^{-1}(1)} U(\bar{s}_v, s_1) + \gamma_1(j) + C_2 \ln p_{1,0,j} \right) \\
&\Leftrightarrow \min_{\bar{s}} (C_1 \bar{R}_1(\bar{s} \mid \bar{\mathbf{x}}) + C_2 \bar{R}_\infty(\bar{s} \mid \bar{\mathbf{x}})).
\end{aligned}$$

Seega tõepoolest saame vähima riskiga joonduse \bar{s} , kui valime

$$s_1 = \arg \max_j (\delta_1(j) + C_2 \ln \pi_j)$$

ja iga $u > 1$ korral

$$s_u = \arg \max_j \left(\delta_u(j) + C_2 \ln p_{u, s_{\rho(u)}, j} \right),$$

kus $s_{\rho(u)}$ on eelnevalt valitud sellesama protseduuri kohaselt. \square

Üldiselt ei tarvitse olla nii selge, missugune näeb välja mingi riski suhtes *parim võimalik* klassifitseerija ja kuidas seda klassifitseerijat rakendada. Kuna hübriidriski *minimiseerivaid* joonduseid on eelkirjeldatud algoritmi abil võimalik leida, siis hübriidrisk sisuliselt määrab endale vastava klassifitseerija ja seetõttu edaspidi nimetamegi hübriidklassifitseerijaks (või -joonduseks) seda klassifitseerijat (või -joondust), mis minimiseerib vastavat hübriidriski.

4.2 k -bloki risk ja selle üldistused

Anneme nüüd võrdusega 13 määratud hübriidriskile konstruktiivse selgituse, tuginedes artiklile Lember ja Koloydenko (2014). Oleme näinud, et PMAP joondus maksimiseerib “ühepunktilisi” tinglikke “tõenäosuseid” $p(u | \bar{\mathbf{x}})$. See tagab klassifitseerija, mis keskmiselt klassifitseerib kõige rohkem üksikuid varjatud kohti S_u õigesti. MAP joondus aga maksimiseerib sisuliselt “ n -punktulist” tinglikku “tõenäosust” $p(I | \bar{\mathbf{x}})$. Vastavat kaofunktsiooni vaadates võime (väikse kujutlusvõime hüppe abil) mõelda, et MAP klassifitseerib õigesti “kõige rohkem n -elemendilisi alamvektoreid”. Mõtlemes esialgu vaid varjatud Markovi mudelitele, kus varjatud kiht on Markovi ahel. Kas saaks äkki defineerida ka mingeid PMAP ja MAP meetodite *vahepealseid* variante, maksimiseerides näiteks järjestikustest paaridest sõltuvaid suuruseid $p(u \cup \rho(u) | \bar{\mathbf{x}})$, mis võiks lõppeda õigesti klassifitseeritud punktipaaride $(S_u, S_{\rho(u)})$ oodatava arvu maksimiseerimises? Või prooviks näiteks maksimiseerida õigesti klassifitseeritud järjestikuste kolmikute oodatavat arvu? Seda ideed mainis põgusalt juba Rabiner (1989). Täpsemalt, olgu (\mathbf{S}, \mathbf{X}) varjatud Markovi mudel. Markovi ahela k -elemendilist (sidusat) alamahelat nimetame k -blokkiks. Kasutades näite 1 tähistust, defineerime kaofunktsiooni

$$L_k(s_1^n, z_1^n) := \frac{1}{n - k + 1} \sum_{t=1}^{n-k+1} I_{s_t^{t+k-1} \neq z_t^{t+k-1}}.$$

Vaadates selle kaofunktsiooni definitsiooni on näha, et klassifitseerija, mis minimiseerib oodatavat L_k -kadu, minimiseeribki valesti klassifitseeritud k -blokkide keskmist arvu. Vastavaks tinglikuks riskiks, on

$$\begin{aligned} R_k(g | x_1^n) &= \sum_{s_1^n \in K^n} \frac{1}{n - k + 1} \sum_{t=1}^{n-k+1} I_{s_t^{t+k-1} \neq g(x_1^n)^{t+k-1}} p_{s_1^n}(\{1, 2, \dots, n\} | \bar{\mathbf{x}}) \\ &= 1 - \frac{1}{n + k - 1} \sum_{t=1}^{n+k-1} p_{g(x_1^n)}(\{t, \dots, t + k - 1\} | \bar{\mathbf{x}}). \end{aligned}$$

Kahjuks ei garanteeri aga R_k -riski minimiseeriv klassifitseerija lubatavat joondust, kui $k > 1$. Sellest ja veel mõnest R_k -riski probleemist saab üle, kui defineerida Lember ja Koloydenko (2014) järgi

$$\bar{R}_k(g | \bar{\mathbf{x}}) := -\frac{1}{n} \sum_{t=1-k}^{n-1} \ln p_{g(\bar{\mathbf{x}})}(\{\max(1, t + 1), \dots, \min(n, t + k)\} | \bar{\mathbf{x}}).$$

Eeltoodud riski definitsioonis esineva veidi krüptilise summa liikmeid illustreerib tabel 4.

Kui (I, N) on aga puu, mitte tingimata ahel, siis k -blokil on mitu erinevat üldistust. Ühelt poolt võib mõelda k -blokkist kui tipust koos kõigi oma kuni $k - 1$ -järku alanevate sugulastega. Puul moodustab selline “ k -blokk” üldiselt hargneva alampuu, justkui kolmnurga. Teiselt poolt võib mõelda k -blokkist kui tipust koos kõigi oma kuni $k - 1$ järku esivanematega. Puul on sel juhul tegu samuti (mittehargneva) ahelaga. Vastavalt nendele tõlgendustele võib \bar{R}_k -riskile samuti defineerida erinevaid üldistusi, kantuna õigesti klassifitseeritud k -blokkide arvu maksimiseerimise vaimust. Siiski osutub, et need kõik on võrdusega 13 defineeritud hübriidriski erijuhud ja seega on vastavad parimad joondused arvutatavad väites 12 kirjeldatud algoritmi abil.

Tabel 4: Näide \bar{R}_k -riski poolt arvestatavatest blokkidest olukorras, kus ahelas on $n = 10$ tippu ja bloki suurus $k = 3$.

	tipp									
t=-2	1	2	3	4	5	6	7	8	9	10
t=-1	1	2	3	4	5	6	7	8	9	10
t=0	1	2	3	4	5	6	7	8	9	10
t=1	1	2	3	4	5	6	7	8	9	10
t=2	1	2	3	4	5	6	7	8	9	10
t=3	1	2	3	4	5	6	7	8	9	10
t=4	1	2	3	4	5	6	7	8	9	10
t=5	1	2	3	4	5	6	7	8	9	10
t=6	1	2	3	4	5	6	7	8	9	10
t=7	1	2	3	4	5	6	7	8	9	10
t=8	1	2	3	4	5	6	7	8	9	10
t=9	1	2	3	4	5	6	7	8	9	10

4.2.1 k -blokk kui tipp koos järglastega

Alustame varjatud Markovi mudelite jaoks defineeritud \bar{R}_k -riski vahest kõige otsesema üldistusega, defineerides iga $k \in \mathbb{N}$ jaoks sama tähistusega tingliku riski

$$\bar{U}_k(\bar{\mathbf{s}} \mid \bar{\mathbf{x}}) := \prod_{u \in I} p(R^-(u, k) \mid \bar{\mathbf{x}}) \prod_{j=1}^{k-1} p(R^-(1, j) \mid \bar{\mathbf{x}}),$$

$$\bar{R}_k(g \mid \bar{\mathbf{x}}) := -\frac{1}{n} \ln \bar{U}_k(g(\bar{\mathbf{x}}) \mid \bar{\mathbf{x}}).$$

Ahela järjestikuste tippude k -bloki analoogiks puul on hulk $R^-(u, k)$ (defineeritud peatükis 1.3.1), mis koosneb tipust u koos kõigi oma kuni $k-1$ -järku alanevate sugulastega. Kui (I, N) on näiteks kümne tipuga ahel ja $k = 3$, siis iga $u \in \{1, 2, \dots, 10\}$ korral hulk $R^-(u, 3)$ on võrdne tabelis 4 indeksile $t = u - 1$ vastava blokiga. Algu-sest jääb aga kaks blokki puudu ($t = -2$ ja $t = -1$), neile vastavad tegurid on vaja juurde korrutada.

Fikseerime klassifitseerija g , vaatluse $\bar{\mathbf{x}}$ ja tähistame $\bar{\mathbf{s}} := g(\bar{\mathbf{x}})$. Nüüd defineerime $\bar{R}_k(g \mid \bar{\mathbf{x}})$ -ga väga analoogilise riski:

$$\bar{U}_k(\bar{\mathbf{s}}) := \prod_{u \in I_1} p(R^-(u, k)) \prod_{j=1}^{k-1} p(R^-(1, j)),$$

$$\bar{R}_k(\bar{\mathbf{s}}) := -\frac{1}{n} \ln \bar{U}_k(\bar{\mathbf{s}}).$$

Niimoodi defineeritud risk võib pakkuda ka iseseisvat huvi (vaata näiteks Lember ja Koloydenko (2014)), hetkel on see küll vaid vahend järgneva väite tõestuses esineva tähistuse lühiduse huvides.

Väide 13 (Lember ja Koloydenko, 2014 teoreem 6 eeskujul). Fikseerime sellise k , et $1 < k \leq d(1)$. Kehtib võrdus

$$\bar{R}_k(\bar{\mathbf{s}}) = \bar{R}_\infty(\bar{\mathbf{s}}) + \bar{R}_{k-1}(\bar{\mathbf{s}}).$$

Tõestus. Olgu $\bar{s} \in K^n$ ja $u \in I$. Kõigepealt märgime, et kui $p(I) = 0$, siis (vastavalt märkusele tegurdamisomaduse definitsiooni 10 järel) mingi $u \in I$ korral ka $p(\{u, \rho(u)\}) = 0$ ja seega $p(R^-(\rho(u), k)) = 0$, mistõttu tõestatava võrduse mõlema poole väärtus on ∞ . Edaspidi eeldame, et $p(I) > 0$.

Kui $d(u) < k$, siis $R^-(u, k) = R^-(u, k - 1)$, seega

$$p(R^-(u, k)) = p(R^-(u, k - 1)).$$

Kui $d(u) \geq k$, siis aga kehtib

$$\begin{aligned} & p(R^-(u, k)) \\ &= p(R^-(u, k - 1) \cup \rho^{-(k-1)}(u)) \\ &= p(R^-(u, k - 1))p(\rho^{-k+1}(u) \mid R^-(u, k - 1)) \\ &= p(R^-(u, k - 1))p(\rho^{-k+1}(u) \mid \rho^{-k+2}(u)) \\ &= p(R^-(u, k - 1)) \prod_{v \in \rho^{-k+1}(u)} p(v \mid \rho(v)). \end{aligned}$$

Esimene võrdus tuleb hulga $R^-(u, k)$ definitsioonist, teine ketireeglist, kolmas ühepoolsest Markovi omadusest ja neljas ketireeglist ning ühepoolsest Markovi omadusest. Seega kehtib

$$\begin{aligned} \prod_{u \in I} p(R^-(u, k)) &= \prod_{u \in I: d(u) < k} p(R^-(u, k)) \prod_{u \in I: d(u) \geq k} p(R^-(u, k)) \\ &= \prod_{u \in I: d(u) < k} p(R^-(u, k - 1)) \cdot \\ & \quad \prod_{u \in I: d(u) \geq k} \left(p(R^-(u, k - 1)) \prod_{v \in \rho^{-k+1}(u)} p(v \mid \rho(v)) \right) \\ &= \prod_{u \in I} p(R^-(u, k - 1)) \prod_{u \in I: d(u) \geq k} \prod_{v \in \rho^{-k+1}(u)} p(v \mid \rho(v)). \end{aligned}$$

Paneme tähele, et

$$I \setminus R^-(1, k - 1) = \bigcup_{u \in I: d(u) \geq k} \rho^{-k+1}(u). \quad (16)$$

Põhjendame selle võrduse. Kui $u \in R^-(1, k - 1)$, siis hulga R^- definitsiooni põhjal on tipp u puu juure ülimalt $k - 2$ järku alanev sugulane:

$$u \in R^-(1, k - 1) \Rightarrow \exists j \in \{0, 1, \dots, k - 2\} : u \in \rho^{-j}(1).$$

Seega tipul u sisuliselt ei ole olemas $k - 1$ järku esivanemat (formaalselt on selleks esivanemaks 0), sest juba $k - 2$ järku või veelgi hiljutisem esivanem on tipp 1. Järelikult ei ole tipp u ühegi tipu $v \in I$ jaoks $k - 1$ järku alanev sugulane ehk $u \notin \rho^{-(k-1)}(v) \forall v \in I$. Teiselt poolt, kui $u \notin R^-(1, k - 1)$, siis leidub tipul u (nullist erinev) $k - 1$ -järku esivanem. Järelikult leidub mingi tipp $v \in I$, nii et $u \in \rho^{-(k-1)}(v)$. Sealjuures peab tipu v sügavus olema vähemalt k , sest vastasel juhul oleks hulk

$\rho^{-(k-1)}(v)$ tühi.

Nüüd näeme, et kehtib

$$\begin{aligned} p(I) &= p(R^-(1, k-1))p(I \setminus R^-(1, k-1) \mid R^-(1, k-1)) \\ &= p(R^-(1, k-1))p(I \setminus R^-(1, k-1) \mid \rho^{k-2}(1)) \\ &= p(R^-(1, k-1)) \prod_{u \in I: d(u) \geq k} \prod_{v \in \rho^{-k+1}(u)} p(v \mid \rho(v)). \end{aligned}$$

Esimene võrdus tuleb ketireeglist, teine globaalsest Markovi omadusest ja kolmas võrdusest 16 ning tegurdamisomadusest ühepoolsest Markovi omadusest.

Lõpuks

$$\begin{aligned} \bar{U}_k(\bar{\mathbf{s}}) &= \left(\prod_{u \in I} p(R^-(u, k)) \right) \left(\prod_{j=1}^{k-1} p(R^-(1, j)) \right) \\ &= \left(\prod_{u \in I} p(R^-(u, k-1)) \prod_{u \in I: d(u) \geq k} \prod_{v \in \rho^{-k}(u)} p(v \mid \rho(v)) \right) \cdot \\ &\quad \left(p(R^-(1, k-1)) \prod_{j=1}^{k-2} p(R^-(1, j)) \right) \\ &= \bar{U}_{k-1}(\bar{\mathbf{s}}_1) p(R^-(1, k-1)) \prod_{u \in I: d(u) \geq k} \prod_{v \in \rho^{-k+1}(u)} p(v \mid \rho(v)) \\ &= \bar{U}_{k-1}(\bar{\mathbf{s}}_1) p(I). \end{aligned}$$

Kui võtta mõlemast võrduse poolest logaritmi ja jagada $-n$ -ga, saame soovitud tulemuse. \square

Eeltoodud väide, mille tõestasime sisuliselt riski $\bar{R}_k(\bar{\mathbf{s}})$ näitel, kehtib ka riski $\bar{R}_k(g \mid \bar{\mathbf{x}})$ jaoks.

Väide 14. Olgu $(\bar{\mathbf{S}}, \bar{\mathbf{X}})$ varjatud Markovi välja mudel, fikseerime klassifitseerija g . Kõigi positiivse tihedusega vaatluste $\bar{\mathbf{x}}$ korral kehtib võrdus

$$\bar{R}_k(g \mid \bar{\mathbf{x}}) = \bar{R}_\infty(g \mid \bar{\mathbf{x}}) + \bar{R}_{k-1}(g \mid \bar{\mathbf{x}}).$$

Tõestus. Väite 8 tõttu on ka vektori $\bar{\mathbf{S}}$ tinglik jaotus globaalse Markovi omadusega (tingimusel, et kõik vaatlused on fikseeritud), mis lubab vektori $\bar{\mathbf{S}}$ tinglike tõenäosusfunktsioonidega (näiteks $p(I \mid \bar{\mathbf{x}})$) manipuleerida täpselt nagu marginaalsete tõenäosusfunktsioonidega (näiteks $p(I)$). Selle teadmise valguses saame täpselt korrata väite 13 tõestust, kirjutades igale poole lisaks tingimuse $\bar{\mathbf{X}} = \bar{\mathbf{x}}$. \square

Riskide perel $\bar{R}_k(g \mid \bar{\mathbf{x}})$, $k \in \mathbb{N}$, on veel üks ilus omadus. Fikseerime vaatlused $\bar{\mathbf{x}}$ ja iga $k \in \mathbb{N}$ korral tähistame tinglikku riski $\bar{R}_k(g \mid \bar{\mathbf{x}})$ minimiseeriva klassifitseerija

$$g_k := \arg \min_g \bar{R}_k(g \mid \bar{\mathbf{x}}).$$

Tuleb välja, et joonduse $g_k(\bar{\mathbf{x}})$ tinglik tõenäosusfunktsioon $p_{g_k(\bar{\mathbf{x}})}(I \mid \bar{\mathbf{x}})$ kasvab monootoonselt koos bloki pikkusega $k \in \mathbb{N}$. Kuna tinglik tõenäosusfunktsioon $p_{g_k(\bar{\mathbf{x}})}(I \mid \bar{\mathbf{x}})$

on aga võrdeline riski $\bar{R}_\infty(g_k | \bar{\mathbf{x}})$ vastandarvuga, siis k kasvades väheneb joonduse $g_k(\bar{\mathbf{x}})$ \bar{R}_∞ -risk monotoonselt. Teisalt, joonduse $g_k(\bar{\mathbf{x}})$ \bar{R}_1 -risk väheneb monotoonselt koos bloki pikkuse k vähenemisega. Joondus $g_k(\bar{\mathbf{x}})$ seega (teatud mõttes) liigub “sujuvalt” MAP ja PMAP joonduse vahel.

Riskide pere $\bar{R}_k(g | \bar{\mathbf{x}})$, $k \in \mathbb{N}$, head omadused võtab kokku järgnev

Väide 15 (Lember ja Koloydenko, 2014 järeldus 8 järgi). Iga positiivse tihedusega vaatluse $\bar{\mathbf{x}}$ ja iga klassifitseerija g korral kehtivad

$$\bar{R}_k(g | \bar{\mathbf{x}}) = (k - 1)\bar{R}_\infty(g | \bar{\mathbf{x}}) + \bar{R}_1(g | \bar{\mathbf{x}}), \quad \text{kui } 1 \leq k \leq d(1); \quad (17)$$

$$g_k(\bar{\mathbf{x}}) \text{ on lubatav} \quad \text{kui } k > 1; \quad (18)$$

$$\bar{R}_\infty(g_k | \bar{\mathbf{x}}) \leq \bar{R}_\infty(g_{k-1} | \bar{\mathbf{x}}), \quad \text{kui } 1 < k \leq d(1); \quad (19)$$

$$\bar{R}_1(\hat{g}_k | \bar{\mathbf{x}}) \geq \bar{R}_1(g_{k-1} | \bar{\mathbf{x}}), \quad \text{kui } 1 < k \leq d(1). \quad (20)$$

Kahe viimase võrratuse tõestuseks on vaja üht abitulemust.

Väide 16 (Lember ja Koloydenko, 2014 lemma 16). Olgu A hulk, $F, G : A \rightarrow \mathbb{R} \cup \{-\infty, \infty\}$ funktsioonid ja $\alpha_1, \alpha_2 \in [0, 1]$, $\alpha_1 \leq \alpha_2$, arvud. Kui $a_1, a_2 \in A$ on sellised, et

$$\alpha_i F(a_i) + (1 - \alpha_i)G(a_i) \leq \alpha_i F(x) + (1 - \alpha_i)G(x), \quad i \in \{1, 2\}$$

iga $x \in A$ korral, siis

$$F(a_1) \geq F(a_2) \text{ ja } G(a_1) \leq G(a_2).$$

Tõestus väitele 15. Tõestus võrdusele 17. Tõestame induktsiooniga. Kui $k = 1$, siis kehtib selgelt. Kehtigu võrdus 17 mingi naturaalarvu k korral. Siis vastavalt väitele 14

$$\begin{aligned} \bar{R}_{k+1}(\bar{\mathbf{s}} | \bar{\mathbf{x}}) &= \bar{R}_\infty(\bar{\mathbf{s}} | \bar{\mathbf{x}}) + \bar{R}_k(\bar{\mathbf{s}} | \bar{\mathbf{x}}) \\ &= \bar{R}_\infty(\bar{\mathbf{s}} | \bar{\mathbf{x}}) + (k - 1)\bar{R}_\infty(\bar{\mathbf{s}} | \bar{\mathbf{x}}) + \bar{R}_1(\bar{\mathbf{s}} | \bar{\mathbf{x}}) \\ &= k\bar{R}_\infty(\bar{\mathbf{s}} | \bar{\mathbf{x}}) + \bar{R}_1(\bar{\mathbf{s}} | \bar{\mathbf{x}}), \end{aligned}$$

nagu vaja.

Tõestus omadusele 18. Kõigepealt veendume, et riski $\bar{R}_k(g | \bar{\mathbf{x}})$ miinimum üle klassifitseerijate g on lõplik. Mingi konfiguratsiooni $\bar{\mathbf{s}} \in K^n$ korral peab kehtima $p_{\bar{\mathbf{s}}}(I | \bar{\mathbf{x}}) > 0$ ja järelikult siis $p_{\bar{\mathbf{s}}}(u | \bar{\mathbf{x}}) > 0$ iga $u \in I$ korral. Defineerime sellise klassifitseerija g , mis rahuldaks $g(\bar{\mathbf{x}}) = \bar{\mathbf{s}}$. Sellise g jaoks kehtib $\bar{R}_\infty(g | \bar{\mathbf{x}}) < \infty$ esimese võrratuse tõttu ja $\bar{R}_1(g | \bar{\mathbf{x}}) < \infty$ teise võrratuse tõttu. Seega $\bar{R}_k(g | \bar{\mathbf{x}}) = (k - 1)\bar{R}_\infty(\bar{\mathbf{s}} | \bar{\mathbf{x}}) + \bar{R}_1(\bar{\mathbf{s}} | \bar{\mathbf{x}}) < \infty$, nagu vaja.

Kui $k > 1$, siis iga klassifitseerija g korral $\bar{R}_k(g | \bar{\mathbf{x}}) = \bar{R}_\infty(g | \bar{\mathbf{x}}) + \bar{R}_{k-1}(g | \bar{\mathbf{x}}) \geq \bar{R}_\infty(g | \bar{\mathbf{x}})$. Oletame, et joondus $g(\bar{\mathbf{x}})$ on lubamatu ehk $P(\bar{\mathbf{S}} = g(\bar{\mathbf{x}}) | \bar{\mathbf{X}} = \bar{\mathbf{x}}) = 0$. Sel juhul aga $\infty = \bar{R}_\infty(g | \bar{\mathbf{x}}) \leq \bar{R}_k(g | \bar{\mathbf{x}})$ ja tegu ei saa olla miinimumiga.

Tõestus võrratustele 19 ja 20. Fikseerime $k : 1 < k \leq d(1)$ ja vaatlused $\bar{\mathbf{x}}$. Kuna g_k on riskile $\bar{R}_k(g | \bar{\mathbf{x}})$ vastav Bayesi klassifitseerija, siis võrduse 17 tõttu

$$\begin{aligned} (k - 1)\bar{R}_\infty(g_k | \bar{\mathbf{x}}) + \bar{R}_1(g_k | \bar{\mathbf{x}}) &= \bar{R}_k(g_k | \bar{\mathbf{x}}) \leq \bar{R}_k(g | \bar{\mathbf{x}}) = (k - 1)\bar{R}_\infty(g | \bar{\mathbf{x}}) + \bar{R}_1(g | \bar{\mathbf{x}}) \\ \Leftrightarrow \frac{1}{k}\bar{R}_1(g_k | \bar{\mathbf{x}}) + \left(1 - \frac{1}{k}\right)\bar{R}_\infty(g_k | \bar{\mathbf{x}}) &\leq \frac{1}{k}\bar{R}_1(g | \bar{\mathbf{x}}) + \left(1 - \frac{1}{k}\right)\bar{R}_\infty(g | \bar{\mathbf{x}}) \end{aligned}$$

mistahes klassifitseerija g korral. Analoogiliselt

$$\frac{1}{k-1}\bar{R}_1(g_{k-1} | \bar{\mathbf{x}}) + \left(1 - \frac{1}{k-1}\right) \bar{R}_\infty(g_{k-1} | \bar{\mathbf{x}}) \leq \frac{1}{k-1}\bar{R}_1(g | \bar{\mathbf{x}}) + \left(1 - \frac{1}{k-1}\right) \bar{R}_\infty(g | \bar{\mathbf{x}})$$

mistahes klassifitseerija g korral.

Nüüd kasutame väidet 16. Valime selle väite tähistustes $F := \bar{R}_1(\cdot | \bar{\mathbf{x}})$, $G := \bar{R}_\infty(\cdot | \bar{\mathbf{x}})$, $\alpha_1 := \frac{1}{k}$, $\alpha_2 := \frac{1}{k-1}$, $a_1 := g_k$, $a_2 := g_{k-1}$ ja x olgu suvaline klassifitseerija kõigi klassifitseerijate hulgast A . Sel juhul

$$\bar{R}_1(g_k | \bar{\mathbf{x}}) \geq \bar{R}_1(g_{k-1} | \bar{\mathbf{x}}) \text{ ja } \bar{R}_\infty(g_k | \bar{\mathbf{x}}) \leq \bar{R}_\infty(g_{k-1} | \bar{\mathbf{x}}),$$

nagu vaja. □

4.2.2 k -blokk kui tipp koos esivanematega

Tõlgendame nüüd ahelal antud k -blokki kui tippu koos oma kõigi esivanematega kuni $k-1$ põlvkonnani minevikus. Kui eelnevalt vaatasime tippu koos oma kõigi $k-1$ -järku alanevate sugulastega, mis moodustavad puul “kolmnurkasid”, siis praegu hakkame vaatama sisuliselt “sirgeid juppe”. Motiveerivaks ideeks on endiselt õigesti klassifitseeritud k -blokkide oodatava arvu maksimiseerimine. Esialgu käsitleme erijuhtu $k=2$.

Vaatame korrutist

$$p(1) \prod_{u>1} p(\{u, \rho(u)\}).$$

Võiksime seada eesmärgiks maksimiseerida seda, kuid eeltoodud korrutises ei ole kõik elemendid esindatud võrdselt: rohkete lastega tipp esineb selles korrutises sagedamini (paari $\{u, \rho(u)\}$ vanema liikmena $\rho(u)$), kuid miks peaks tema seetõttu kaalukam olema? Seetõttu maksimiseerime hoopis korrutist, milles kõik tipud on esindatud (peaaegu) võrdselt:

$$p(1) \prod_{u>1} p(\{u, \rho(u)\}) \prod_{u \in I} p(u)^{c_* - c(u)},$$

kus $c(u) := |\rho^{-1}(u)|$ on tipu u laste arv ja $c_* := \max_{u \in I} c(u)$ on maksimaalne laste arv. Iga tippu on võetud nii mitu korda lisaks, kui palju jääb sel tipul puudu maksimaalsest laste arvust.

Eeldame, et $\bar{\mathbf{s}} \in K^n$ on selline, et $p(\rho(u)) > 0$ iga u korral (see garanteerib, et järnevad tinglikud tõenäosused on defineeritud). Tegurdamisomaduse tõttu

$$\begin{aligned} p(I) &= p(1) \prod_{u>1} p(u | \rho(u)) = p(1) \prod_{u>1} \frac{p(\{u, \rho(u)\})}{p(\rho(u))} \\ &= p(1) \prod_{u>1} p(\{u, \rho(u)\}) \prod_{u>1} p(\rho(u))^{-1} \\ &= p(1) \prod_{u>1} p(\{u, \rho(u)\}) \prod_{u \in I} p(u)^{-c(u)} \end{aligned}$$

Viimane võrdus tuleb asjaolust, et korrutises $\prod_{u>1} p(\rho(u))$ esineb tippu $\rho(u)$ nii mitu korda, kui palju on tal lapsi. Paneme tähele, et ülaltoodud võrdus kehtib

(triviaalselt) ka siis, kui $p(\rho(u)) = 0$ mingi u korral. Nüüd aga korrutame eelneva võrduse mõlemad pooled läbi suurusega $\prod_{u \in I} p(u)^{c_*}$, tulemuseks võrdus

$$p(1) \prod_{u > 1} p(\{u, \rho(u)\}) \prod_{u \in I} p(u)^{c_* - c(u)} = p(I) \prod_{u \in I} p(u)^{c_*}.$$

Kui tähistame selle võrduse vasaku poole \bar{U}_2^* ja defineerime $\bar{R}_2^*(\bar{s}) := -\frac{1}{n} \ln \bar{U}_2^*$, siis näemegi, et

$$\bar{R}_2^*(\bar{s}) = \bar{R}_\infty(\bar{s}) + c_* \bar{R}_1(\bar{s})$$

Üldiselt, kui $k \geq 2$, defineerime (iga $\bar{s} \in K^n$ jaoks)

$$\begin{aligned} \bar{U}_k^*(\bar{s}) &:= \prod_{u \in I} p_{\bar{s}} \left(\bigcup_{i=0}^{k-1} \rho^i(u) \setminus 0 \right) \prod_{j=0}^{k-2} p_{\bar{s}} \left(\bigcup_{i=0}^j \rho^i(u) \setminus 0 \right)^{c_*^{k-2+j}(c_* - c(u))} \\ \bar{R}_k^*(\bar{s}) &:= -\frac{1}{n} \ln \bar{U}_k^*(\bar{s}), \end{aligned}$$

meenutades kokkuleppeid $\rho(1) := 0$, $\rho(0) := 0$ ja ühe-elementiliste hulkade korral $\{u\} := u$. Defineerime veel $\bar{U}_k^*(\bar{s} \mid \bar{x})$ sarnaselt suurusele $\bar{U}_k^*(\bar{s})$, lisades ainult tõenäosusfunktsioonidele p tingimuse $\bar{X} = \bar{x}$. Seejärel loomulikult

$$\bar{R}_k^*(\bar{s} \mid \bar{x}) := -\frac{1}{n} \ln \bar{U}_k^*(\bar{s} \mid \bar{x}).$$

Korrutise \bar{U}_k^* sisu selgitamiseks vaatame näiteks erijuhtu $k = 3$, mispuhul

$$\bar{U}_k^*(\bar{s}) = \prod_{u \in I} p(\{u, \rho(u), \rho^2(u)\} \setminus 0) p(\{u, \rho(u)\} \setminus 0)^{c_* - c(u)} p(u)^{c_*(c_* - c(u))}.$$

Endiselt figureerivad tipp-vanem-vanavanem kolmikud $\{u, \rho(u), \rho^2(u)\}$, kuid lisan-deid nende kõrval on rikkalikult.

Nagu lubatud, siis kehtib

Väide 17. Iga $k \geq 2$ ja $\bar{s} \in K^n$ korral

$$\bar{R}_k^*(\bar{s} \mid \bar{x}) = \left(\sum_{i=0}^{k-2} c_*^i \right) \bar{R}_\infty(\bar{s} \mid \bar{x}) + c_*^{k-1} \bar{R}_1(\bar{s} \mid \bar{x}).$$

Enne tõestust esitame abistava väite.

Väide 18. Iga $k \geq 2$ ja $\bar{s} \in K^n$ korral kehtib

$$\prod_{u \in I} p \left(\bigcup_{i=0}^{k-1} \rho^i(u) \setminus 0 \right) = p(I) \prod_{u \in I} p \left(\bigcup_{i=0}^{k-2} \rho^i(u) \setminus 0 \right)^{c(u)}.$$

Tõestus. Olgu $k \geq 2$ suvaline. Kõigepealt märgime, et kui \bar{s} on selline, et $p(I) = 0$, siis võrdus kehtib (vaata märkust tegurdamisomaduse definitsiooni 10 järel). Eeldame nüüd, et $p(I) > 0$. Kehtib

$$\prod_{u \in I} p \left(\bigcup_{i=1}^{k-1} \rho^i(u) \setminus 0 \right) = \prod_{u \in I} p \left(\bigcup_{i=0}^{k-2} \rho^i(u) \setminus 0 \right)^{c(u)}, \quad (21)$$

sest võrduse vasakul poolel asuvas korrutises esineb tegur indeksiga u kokku nii mitu korda, kui palju on tipul $\rho(u)$ lapsi. Nüüd

$$\begin{aligned} \prod_{u \in I} p \left(\bigcup_{i=0}^{k-1} \rho^i(u) \setminus 0 \right) &= \prod_{u \in I} p \left(u \mid \bigcup_{i=1}^{k-1} \rho^i(u) \setminus 0 \right) p \left(\bigcup_{i=1}^{k-1} \rho^i(u) \setminus 0 \right) \\ &= \prod_{u \in I} p(u \mid \rho(u) \setminus 0) \prod_{u \in I} p \left(\bigcup_{i=1}^{k-1} \rho^i(u) \setminus 0 \right) \\ &= p(I) \prod_{u \in I} p \left(\bigcup_{i=0}^{k-2} \rho^i(u) \setminus 0 \right)^{c(u)}, \end{aligned}$$

nagu vaja. Esimene võrdus tuleb ketireeglist, teine ühepoolsest Markovi omadusest ja kolmas tegurdamisomadusest ning võrdusest 21. \square

Tõestus väitele 17. Käitume analoogiliselt väidete 13 ja 14 tõestusele. See tähendab, et näitame vajalikud seosed ära $\bar{R}_k^*(\bar{\mathbf{s}})$ jaoks ja siis järeldame soovitud tulemuse asjaolust, et vektori $\bar{\mathbf{S}}$ tinglikul jaotusel, tingimusel et vaatlused $\bar{\mathbf{X}}$ on fikseeritud, on ka globaalne Markovi omadus – seega võib järgnevat tõestust täpselt korrata, asendades lihtsalt tõenäosusfunktsioonid vastavate tinglike tõenäosusfunktsioonidega.

Oleme näinud, et erijuhul $k = 2$ soovitud tulemus kehtib. Olgu nüüd $k \geq 3$. Rehkendame:

$$\begin{aligned} &\bar{U}_k^*(\bar{\mathbf{s}}) \\ &= \prod_{u \in I} p \left(\bigcup_{i=0}^{k-1} \rho^i(u) \setminus 0 \right) \prod_{j=0}^{k-2} p \left(\bigcup_{i=0}^j \rho^i(u) \setminus 0 \right)^{c_*^{k-2+j}(c_*-c(u))} \\ &= p(I) \prod_{u \in I} p \left(\bigcup_{i=0}^{k-2} \rho^i(u) \setminus 0 \right)^{c(u)} p \left(\bigcup_{i=0}^{k-2} \rho^i(u) \setminus 0 \right)^{c_*-c(u)} \prod_{j=0}^{k-3} p \left(\bigcup_{i=0}^j \rho^i(u) \setminus 0 \right)^{c_*^{k-2+j}(c_*-c(u))} \\ &= p(I) \left(\prod_{u \in I} p \left(\bigcup_{i=0}^{k-2} \rho^i(u) \setminus 0 \right) \prod_{j=0}^{k-3} p \left(\bigcup_{i=0}^j \rho^i(u) \setminus 0 \right)^{c_*^{k-3+j}(c_*-c(u))} \right)^{c_*} \\ &= p(I) (\bar{U}_k^*(\bar{\mathbf{s}}))^{c_*}. \end{aligned}$$

Esimene ja viimane võrdus tulevad $\bar{U}_k^*(\bar{\mathbf{s}})$ definitsioonist ja teine võrdus väitest 18. Induktsiooniga saab nüüd veenduda, et

$$\bar{U}_k^*(\bar{\mathbf{s}}) = p(I)^{a(k)} \prod_{u \in I} p(u)^{c_*^{k-1}},$$

kus $a(k) = \sum_{i=0}^{k-2} c_*^i$. Võrduse mõlemast poolest naturaalaritmi võtmine ja $-n$ -ga läbi jagamine lõpetab tõestuse. \square

4.2.3 Hübriidriski tõlgendus

Meenutame hübriidriski definitsiooni ja väiteid 17 ning 15:

$$\begin{aligned}\bar{R}_{C_1, C_2}(g | \bar{\mathbf{x}}) &:= C_1 \bar{R}_1(g | \bar{\mathbf{x}}) + C_2 \bar{R}_\infty(g | \bar{\mathbf{x}}), \\ \bar{R}_k(g | \bar{\mathbf{x}}) &= \bar{R}_1(g | \bar{\mathbf{x}}) + (k-1) \bar{R}_\infty(g | \bar{\mathbf{x}}), \\ \bar{R}_k^*(g | \bar{\mathbf{x}}) &= c_*^{k-1} \bar{R}_1(g | \bar{\mathbf{x}}) + \left(\sum_{i=0}^{k-2} c_*^i \right) \bar{R}_\infty(g | \bar{\mathbf{x}}).\end{aligned}$$

Näeme, et nii \bar{R}_k kui ka \bar{R}_k^* on hübriidriski \bar{R}_{C_1, C_2} erijuhud. Risk \bar{R}_k võtab arvesse blokke, mille moodustab tipp koos kõigi oma kuni $k-1$ järku alanevate sugulastega, ja risk \bar{R}_k^* võtab arvesse blokke, mille moodustab tipp koos oma esivanemate jadaga kuni $k-1$ järku esivanemani. Paneme tähele, et kui varjatud kiht on Markovi ahel, siis maksimaalseks laste arvuks on 1 ja $\bar{R}_k(g | \bar{\mathbf{x}}) = \bar{R}_k^*(g | \bar{\mathbf{x}}) -$ tegu tõepoolest on erinevate üldistustega, mis ahelal langevad siiski kokku.

Konstantidelt C_1 ja C_2 võime nõuda, et $C_1 + C_2 = 1$, näiteks valides uued konstandid $C'_1 := \frac{C_1}{C_1 + C_2}$ ja $C'_2 := 1 - C'_1$. See toiming küll arvuliselt muudab riski \bar{R}_{C_1, C_2} , kuid ei muuda joondust (klassifitseerijat), mis minimiseerib seda riski – ja ainult see ongi ju oluline. Antud tähelepanekule tuginedes tähistame edaspidi hübriidriski lihtsalt \bar{R}_{C_1} , kus $C_1 \in [0, 1]$, pidades selle all silmas riski $\bar{R}_{C_1, 1-C_1}$ ja teades, et see tähistus ei kitsenda üldisust.

Vaatame kõigepealt riski \bar{R}_k , mille seos riskidega \bar{R}_1 ja \bar{R}_∞ on täpselt samasugune nii ahelatel kui ka mistahes teistel puudel. Valik $C_1 = \frac{1}{k}$ loob seose riski \bar{R}_k ja hübriidriski \bar{R}_{C_1} vahel (selles mõttes, et riski \bar{R}_k ja hübriidriski $\bar{R}_{C_1 = \frac{1}{k}}$ minimiseerib sama klassifitseerija). Kui bloki pikkus k on suurem kui 2, siis $\frac{1}{k} \leq 0.5$ ja seega \bar{R}_k -riskid hõlmavad pigem hübriidriskide pere “alumise poole”, kus $C_1 \leq 0.5$. Teiselt poolt võime mõelda, et seos $C_1 = \frac{1}{k}$ annab mõtte ka murdarvulistele bloki pikkustele – näiteks valik $C_1 = 0.4$ justkui vastaks bloki pikkusele 2.5 ehk tipule koos laste ja “poolte” lastelastega.

Erinevalt riskist \bar{R}_k sõltub risk \bar{R}_k^* aga puu struktuurist, täpsemalt siis maksimaalsest laste arvust. Seetõttu võivad erinevad bloki pikkused k vastata erinevatele hübriidriskidele, sõltuvalt konkreetsest puust (vaata joonist 4). Kuna aga

$$c_*^{k-1} \geq \sum_{i=0}^{k-2} c_*^i,$$

kui vaid $c_* \geq 2$ ja $k \geq 2$ (näidata saab seda näiteks induktsiooniga k järgi), siis

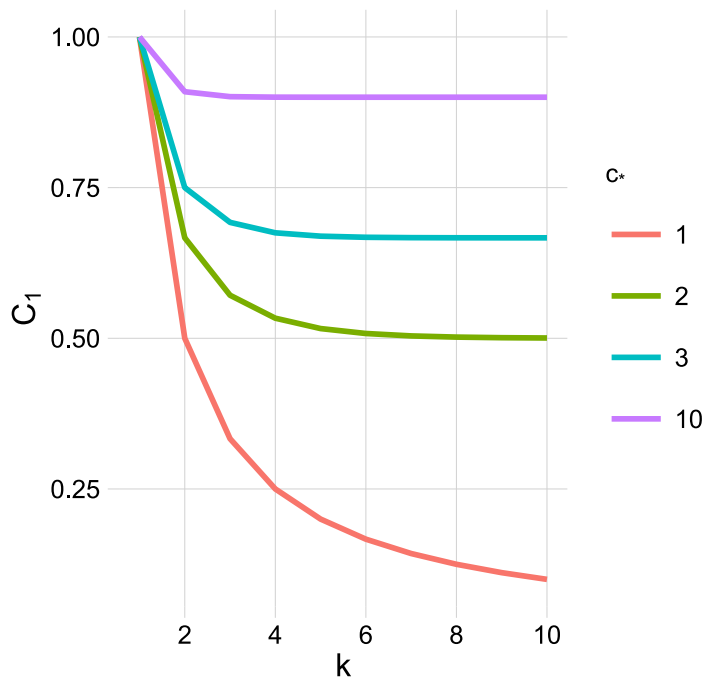
$$C_1 = \frac{c_*^{k-1}}{\sum_{i=0}^{k-1} c_*^i} \geq 0.5$$

ja riskid \bar{R}_k^* paiknevad pigem hübriidklassifitseerijate pere “ülemises pooles” (pane me tähele, et eeltoodud C_1 valik loob seose riski \bar{R}_k^* ja hübriidriski \bar{R}_{C_1} vahel).

Näeme, et hübriidriskide pere \bar{R}_{C_1} , $C_1 \in [0, 1]$, seobki PMAP klassifitseerija ($C_1 = 1$) ühte MAP klassifitseerijaga ($C_1 = 0$), tuues nende vahele terve suguvõsa klassifitseerijaid \bar{R}_k , mis minimiseerivad k -bloki riski mingit üldistust. Mida suurem on C_1 , seda väiksemat konteksti võetakse arvesse parima joonduse leidmisel. Tabelis 5 on erinevate C_1 valikute näiteid ja selgitusi, mis konkreetse C_1 valiku korral juhtub.

Tabel 5: Näiteid erinevate hübriidriskide tõlgendustest.

C_1	risk	tõlgendus
0	\bar{R}_∞	Tulemuseks MAP klassifitseerija – leitakse joondus \bar{s} , mille “tõepära” $p_{\bar{s}}(\bar{x}, I)$ on suurim (fikseeritud vaatluste \bar{x} korral). Üksikuid kohti S_u eraldi ega ka mingisuguseid blokke ei vaadata üldse.
$\frac{1}{3}$	\bar{R}_2	Võetakse arvesse blokke, mis koosnevad tipust koos kõigi laste ja lastelastega.
0.5	\bar{R}_2	Võetakse arvesse blokke, mis koosnevad tipust ja kõigist tema lastest.
$\frac{2}{3}$	\bar{R}_2^* kahend- puul	Võetakse arvesse kahendpuu tipp-vanem paare. Vastab murdarvulise bloki pikkuse $k = 1.5$ (tipp “koos poolte lastega”) kasutamisele riski \bar{R}_k kontekstis.
1	\bar{R}_1	Tulemuseks PMAP klassifitseerija – üksikutest seisunditest s_u panakse kokku selline joondus \bar{s} , et igasse konkreetseesse kohta u just see seisund s_u sobib kõige paremini (marginaalse järeltõenäosuse ehk silumistõenäosuse poolest). Ei arvestata absoluutselt mitte mingit konteksti ehk seda, missugusse seisundisse on prognoositud naabrid. Niimoodi saame joonduse, mis teeb keskmiselt kõige vähem eksimusi, kuid mille tõenäosus võib olla null.



Joonis 4: Konstandid C_1 , mis vastavad riskide \bar{R}_k^* ja \bar{R}_k hübriidriski \bar{R}_{C_1} kujul esitusele erinevate bloki pikkuste k ja maksimaalsete laste arvude c_* korral. Pane me tähele, et riski \bar{R}_k käitumist iseloomustab seesama joon, mis kirjeldab riski \bar{R}_k^* käitumist ahelal (kui $c_* = 1$).

4.3 Simulatsioonid

Oleme tutvunud hübriidriskiga ja veendunud, et seda riski minimeerivat joondust on puude peal võimalik küllalt kergesti leida. Illustreerime nüüd vastavaid joondusi ja uurime nende täpsust.

Kirjeldame üldjoontes simulatsioonikatset, millel põhinevad käesoleva alapeatüki järgnevad näited. Vajalikud detailid on märgitud konkreetsetes näidetes. Kasutatud on kuut tüüpi puid: ahel, “vähe hargnev puu” (umbes 1 kahekshargnemine kahekümne tipu kohta), “keskmiselt hargnev puu” (umbes 1 kahekshargnemine kümne tipu kohta), kahendpuu, kolmendpuu ja kümnendpuu. Erinevaid varjatud Markovi välja mudelid on kasutatud samuti 6, kõik on diskreetse emissioonijaotusega ja homogeensed selles mõttes, et üleminekutõenäosused ja emissioonijaotused on kõigi tippude jaoks samad. Mudelid on algjaotuse, üleminekutõenäosuste ja emissioonijaotuse kaudu kirjeldatud tabelis 6. Selle alapeatüki näidetes on klassifitseerijad saanud kasutada vastavate mudelite tegelikke, täpseid parameetreid. Simulatsioonid on teostatud Python versioon 2.7 abil, kasutatud on lisamooduleid ETE (Huerta-Cepas jt., 2010) ja NumPy (Van Der Walt jt., 2011). Töö lisas on esitatud simulatsioonide tegemisel kasutatud põhilised funktsioonid. Enamus jooniseid on tehtud R paketi ggplot2 abil (Wickham, 2009; R Core Team, 2014).

Meenutame hübriidriski definitsiooni:

$$\bar{R}_{C_1, C_2}(g \mid \bar{\mathbf{x}}) := C_1 \bar{R}_1(g \mid \bar{\mathbf{x}}) + C_2 \bar{R}_\infty(g \mid \bar{\mathbf{x}}).$$

Vastavalt alapeatükli 4.2.3 tehtud märkusele võime üldisust kitsendamata nõuda, et $C_1 + C_2 = 1$, mis lubab hübriidriski \bar{R}_{C_1, C_2} määrata vaid ühe parameetriga. Seda teed järgnevalt lähemegi, täpsustades vaid parameetri $C_1 \in [0, 1]$ ja viidates selle abil hübriidriskile $\bar{R}_{C_1, 1-C_1}$.

Näide 9 (Hübriidklassifitseerijate erinevused). Tutvume erinevate joonduste väljanägemisega joonise 5 abil.

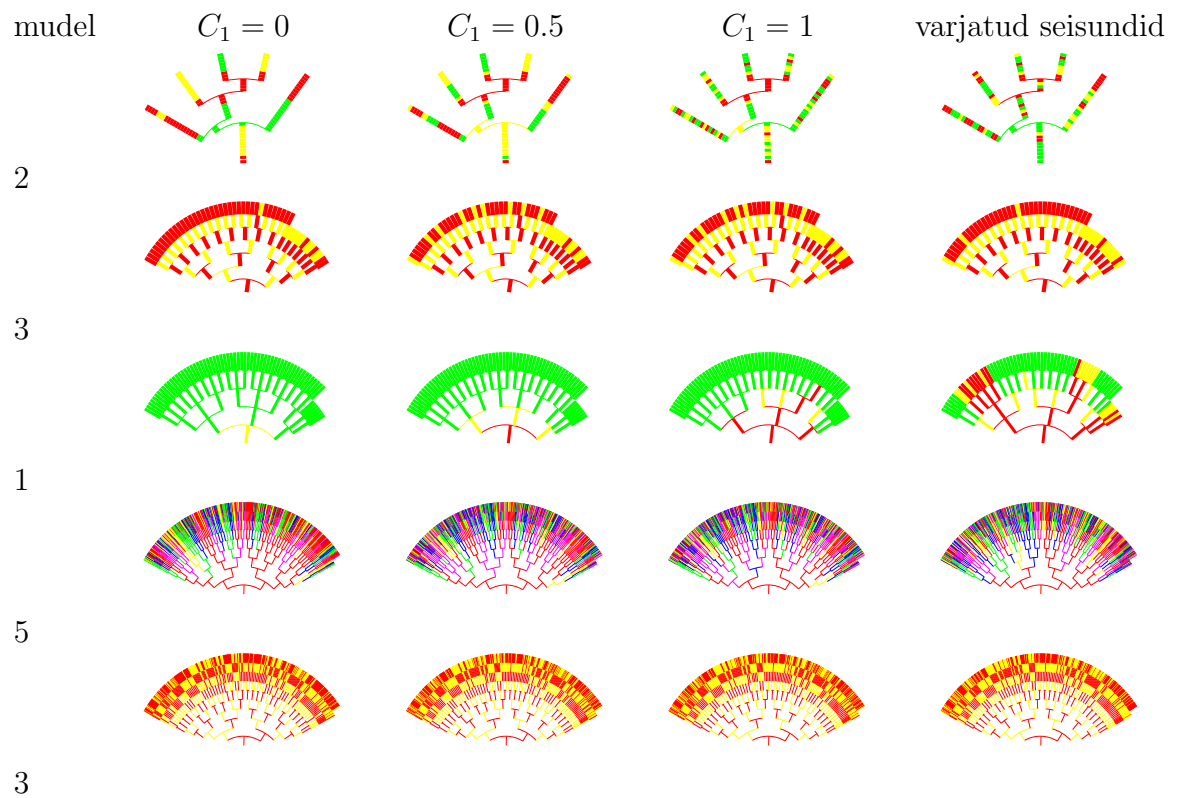
Joonise vasakpoolses veerus ($C_1 = 0$) on MAP joondused ja paremalt teises veerus ($C_1 = 1$) on PMAP joondused. Näeme, et MAP joondused kipuvad mudelite 1, 2 ja 5 puhul koosnema suurematest ühes seisundis olevatest tükkidest, samas kui PMAP joondused näevad välja kirjumad. See on oodatav, sest et nende mudelite korral on vanemaga samasse seisundisse jäämine kõige tõenäolisem valik, ja MAP joondus võtab seda arvesse. Seevastu näiteks mudeli 3 korral pole MAP ja PMAP joondustel suurt vahet, sest et emissioonijaotusi vaadates on seisund vaatluse põhjal niikuinii kergesti tuvastatav. Lisaks on selle mudeli puhul vanemaga samasse seisundisse jäämisest tõenäolisem seisundi vahetus, mis põhjustab vöödilise (või malelauamustrilise) seisundite konfiguratsiooni. Samuti on näha, et PMAP joonduses võib olla lubamatuid üleminekuid, nagu mudeli 1 puhul seisundist 1 (punane) seisundisse 3 (roheline).

Näide 10 (Hübriidklassifitseerijate empiiriline risk). Klassifitseerija g täpsuse kirjeldamiseks defineerime tegelike seisundite konfiguratsiooni $\bar{\mathbf{s}}$ ja vaatluste $\bar{\mathbf{x}}$ jaoks empiirilise riski

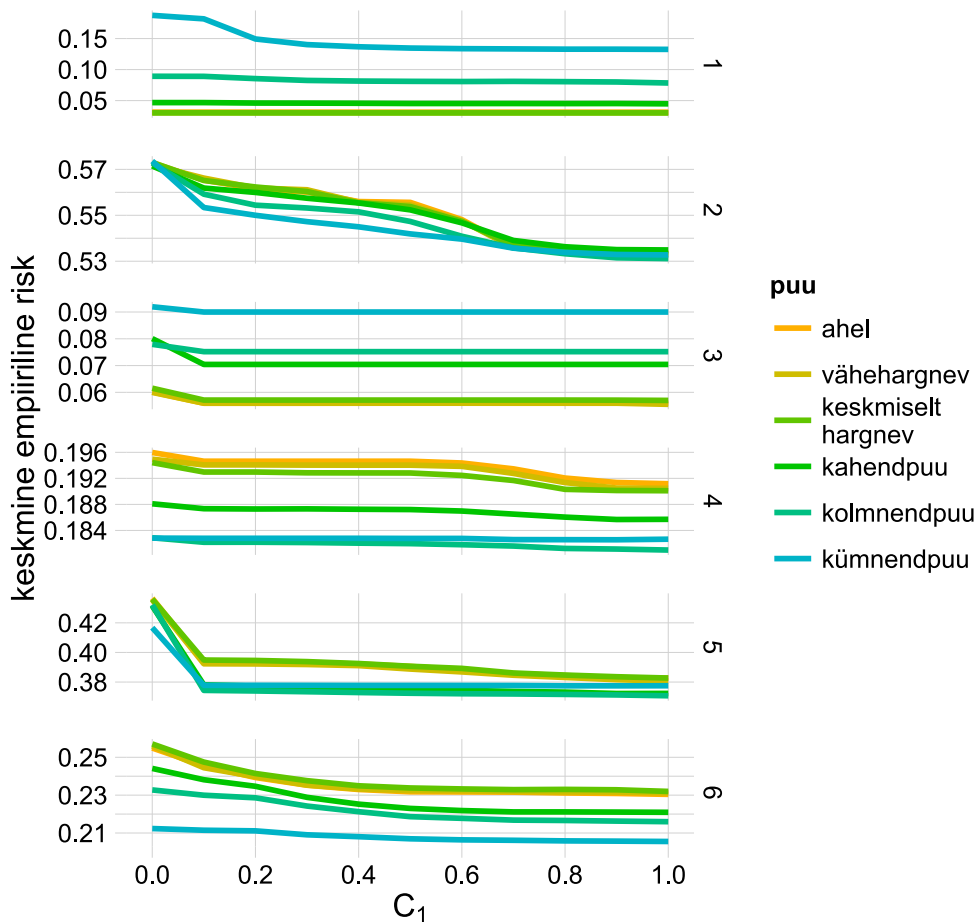
$$R_n(g) := \frac{1}{n} \sum_{u \in I} I_{s_u \neq g(\bar{\mathbf{x}})_u},$$

Tabel 6: Simulatsioonides kasutatud mudelid. Emissioonijaotus on kirjeldatud maatriksi $(f_j(x))_{j \in K, x \in \mathcal{X}}$ abil, kus reas j ja veerus x on tõenäosus, et fikseeritud seisundist j tuleb vaatlus x . Paneme tähele, et selle maatriksi veergude arv näitab erinevate vaatluste arvu. Esimene mudel on võetud Lember ja Koloydenko (2014) näitest 2.

number	algjaotus	üleminekumaatriks	emissioonijaotus
1	$\begin{pmatrix} 0.4 \\ 0.54 \\ 0.06 \end{pmatrix}$	$\begin{pmatrix} 0.6 & 0.4 & 0 \\ 0.1 & 0.1 & 0.8 \\ 0 & 0.02 & 0.98 \end{pmatrix}$	$\begin{pmatrix} 0.3 & 0.15 & 0.25 & 0.3 \\ 0.2 & 0.3 & 0.3 & 0.2 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0.5 \end{pmatrix}$
2	$\begin{pmatrix} 0.3 \\ 0.3 \\ 0.4 \end{pmatrix}$	$\begin{pmatrix} 0.5 & 0.25 & 0.25 \\ 0.25 & 0.5 & 0.25 \\ 0.25 & 0.25 & 0.5 \end{pmatrix}$	$\begin{pmatrix} 0.4 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.4 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.3 & 0.3 \end{pmatrix}$
3	$\begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$	$\begin{pmatrix} 0.1 & 0.9 \\ 0.9 & 0.1 \end{pmatrix}$	$\begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}$
4	$\begin{pmatrix} 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \end{pmatrix}$	$\begin{pmatrix} 0.6 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.6 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.6 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.6 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.6 \end{pmatrix}$	$\begin{pmatrix} 0.8 & 0.05 & 0.05 & 0.05 & 0.05 \\ 0.05 & 0.8 & 0.05 & 0.05 & 0.05 \\ 0.05 & 0.05 & 0.8 & 0.05 & 0.05 \\ 0.05 & 0.05 & 0.05 & 0.8 & 0.05 \\ 0.05 & 0.05 & 0.05 & 0.05 & 0.8 \end{pmatrix}$
5	$\begin{pmatrix} 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \end{pmatrix}$	$\begin{pmatrix} 0.6 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.6 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.6 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.6 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.6 \end{pmatrix}$	$\begin{pmatrix} 0.6 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.6 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.6 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.6 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.6 \end{pmatrix}$
6	$\begin{pmatrix} 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0.1 & 0.1 & 0.8 \\ 0 & 0 & 0.1 & 0.8 & 0.1 \\ 0 & 0.1 & 0.8 & 0.1 & 0 \\ 0.1 & 0.8 & 0.1 & 0 & 0 \\ 0.8 & 0.1 & 0.1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0.6 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.6 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.6 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.6 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.6 \end{pmatrix}$



Joonis 5: Erinevaid hübriidjoonduseid. Joonisel on nelja simulatsiooni tulemused: igal simulatsioonil kolm erinevat hübriidjoondust ja parempoolses veerus konkreetsel simulatsioonil realiseerunud tegelikud varjatud seisundid. Puu on kujutatud nii, et juur on all ja lehed üleval. Puu tipud asuvad radiaalselt paiknevatel lõikudel. Ühe tipu lapsi ühendab ringjoone kaar. Seisundile 1 vastab punane, seisundile 2 kollane, seisundile 3 roheline, seisundile 4 sinine ja seisundile 5 lilla värv.

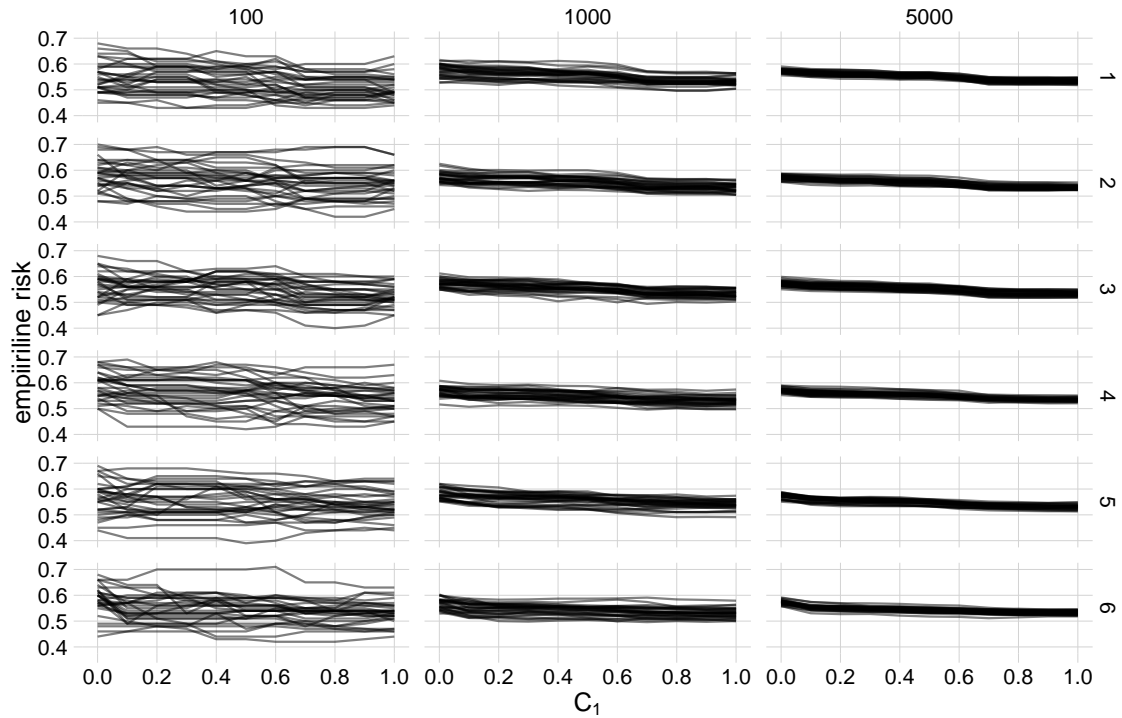


Joonis 6: Hübriidklassifitseerijate keskmised empiirilised riskid erinevate puude ja erinevate mudelite korral. Kasutatud on ainult puid tippude arvuga 5000. Igal mudeli (ridades: mudel 1 kuni mudel 6) ja puu tüübi (erinevad värvid) kombinatsioonil on teostatud 30 simulatsiooni, leitud hübriidklassifitseerijate ($C_1 = 0, 0.1, \dots, 1$) empiirilised riskid ja arvutatud aritmeetilised keskmised.

mis väljendab valesti klassifitseeritud punktide $u \in I$ osakaalu. Teame, et teoreetiliselt võiks seda empiirilist riski *keskmiselt* minimiseerida PMAP klassifitseerija. Kuidas käituvad teised hübriidklassifitseerijate pere liikmed? Uurime, kuidas sõltub hübriidklassifitseerijate empiiriline risk puu kujust ja tippude arvust.

Joonistelt 6 ja 7 näeme, et hübriidklassifitseerija keskmine empiiriline risk kaaneb monotoonselt koos konstandi C_1 kasvuga kõigi katsetatud puu tüüpide ja mudelite korral. See on kooskõlas teooriaga: teame, et PMAP klassifitseerija on see klassifitseerija, mis teeb keskmiselt kõige vähem vigu, ja empiiriline risk mõõdabki just tehtud vigade arvu. Seega on loogiline, et riski $\bar{R}_{C_1=1} := \bar{R}_{1,0}$ minimiseeriv klassifitseerija, mis ongi PMAP klassifitseerija, saavutab vähima empiirilise riski võrreldes mis iganes teiste klassifitseerijatega.

Jooniselt 6 ka on näha, et keskmine empiiriline risk sõltub puu hargnevusest. Mõnel juhul paistavad klassifitseerijad tegema rohkem vigu rohkelt hargneval puul (mudelite 1 ja 3 puhul), mõnel juhul aga hoopis vähem hargneval puul (mudelite 2, 4, 5 ja 6 puhul). Paneme tähele, et empiiriline risk (valesti klassifitseeritud kohtade



Joonis 7: Hübriidklassifitseerijate empiirilised riskid erinevate puude korral. Kasutatud on mudelit 2. Igal tippude arvu (veergudes: 100, 1000 ja 5000) ja puu tüübi (ridades: ahelast 1 kuni kümnenndpuuni 6) kombinatsioonil on teostatud 30 simulatsiooni ja leitud hübriidklassifitseerijate ($C_1 = 0, 0.1, \dots, 1$) empiirilised riskid. Üks must joon väljendab ühe konkreetse simulatsiooni (vektorite paari $(\bar{\mathbf{S}}, \bar{\mathbf{X}})$ ühe realisatsiooni) alusel tehtud erinevate joonduste empiirilisi riske.

osakaal) paistab *monotoonselt* sõltuvat puu hargnemiste arvust, ja seda iga klassifitseerija $\bar{R}_{C_1}, C_1 \in \{0, 0.1, \dots, 1\}$ korral. Näiteks mudelis 4 teevad klassifitseerijad kõige enam vigu ahelal, vähe hargneval puul on vigu veidi vähem, seejärel tuleb keskmiselt hargnev puu, siis kahendpuu ja kõige väiksema keskmise empiirilise riski saavutavad klassifitseerijad kolmendpuul (vaata joonist 6). Mudeli 3 puhul on olukord küll ümber pööratud, kuid järjekord siiski peaaegu sama.

Vaadates erinevate hübriidjoonduste käitumist konkreetsete mudeli realisatsioonide korral (joonis 7) näeme, et mõnikord võib parima tulemuse (kõige vähem vigu) anda hoopis mingi vahepealne klassifitseerija, mitte PMAP. Küll aga *keskmine* empiiriline risk ikkagi kahaneb sedamööda, kuidas konstant C_1 kasvab ja seega joendus (mingis mõttes) läheneb PMAP joondusele.

5 Parameetrite hindamine

5.1 Baum-Welchi algoritm

Eelnevas oleme kogu aeg eeldanud, et varjatud Markovi välja *mudel* on täielikult teada, küsimus on ainult varjatud kihi konfiguratsioonis. Ilmselt väga sageli see nii ei ole – vaatlustelt on tarvis hinnata ka mudeli parameetrid. Diskreetsete emissioonijaotustega ja puul defineeritud varjatud Markovi välja mudeli korral on vaja hinnata algjaotus, üleminekumaatriks ja emissioonijaotus kirjeldav stohhastiline maatriks. Järgnevalt tutvume ühe viisiga parameetrite hindamiseks.

Olgu eesmärgiks saada suurima tõepära hinnang huvipakkuva jaotuse (mitte tingimata varjatud Markovi välja mudeli) parameetervektorile käesoleva valimi alusel. Selleks on tarvis leida tõepärafunktsiooni maksimumkoht. Sõltuvalt konkreetsest jaotusmudelist võib see olla ülesanne olla väga erineva raskusega. Mõnikord aitab rasket ülesannet hoomatavamaks muuta ettekujutus, et lisaks käesolevale valimile on olemas ka mingi hulk mõõtmata jäänud juhuslikke suuruseid. Selline on olukord näiteks puuduvaid andmeid sisaldava valimi, segujaotusest pärineva valimi ja ka varjatud Markovi (välja) mudeli korral. Kui jaotusmudel on esitatav kätte saadud ja saamata jäänud vaatluste (ühis)jaotuse abil, siis EM algoritm on levinud võimaluseks iteratiivselt leida parameetervektorile suurima tõepära hinnang (või vähemasti jõuda tõepärafunktsiooni lokaalsesse maksimumi). EM algoritmi erijuhtu varjatud Markovi mudelite tarvis nimetatakse Baum-Welchi algoritmiks, kuigi need kaks algoritmi kujunesid ajalooliselt iseseisvalt (Koski, 2001). Kirjeldame Baum-Welchi algoritmi Bilmes jt. (1998) järgi ja seejärel esitame selle algoritmi vahetu üldistuse puudel antud varjatud Markovi välja mudelitele.

Olgu antud varjatud Markovi mudel vaatlustega $\bar{\mathbf{X}} = (X_t)_{t=1}^T$ ja varjatud kihiga $\bar{\mathbf{S}} = (S_t)_{t=1}^T$. Eeldame, et varjatud ahela seisundite ruumis K on k seisundit ja emissioonijaotusel on samuti lõplik arv l võimalikku väärtust hulgas \mathcal{X} . Samuti eeldame, et Markovi ahel $\bar{\mathbf{S}}$ on homogeenne – üleminekumaatriks on kõigi tippude jaoks üks ja seesama. Emissioonijaotuste “homogeensust” eeldatakse tihti vaikimisi. Nüüd on mudel määratud algjaotusega (tõenäosuste vektor $\pi := (\pi_j)_{j \in K} := (\mathbf{P}(S_1 = j))_{j \in K}$), üleminekumaatriksiga (stohhastiline maatriks $A := (a_{i,j})_{i \in K, j \in K} := (\mathbf{P}(S_t = j \mid S_{t-1} = i))_{i \in K, j \in K}$) ja stohhastilise maatriksiga $B := (b_j(x))_{j \in K, x \in \mathcal{X}} := \mathbf{P}(X_t = x \mid S_t = j) \forall t$, mis määrab emissioonijaotused.

Baum-Welchi algoritm töötab iteratiivselt. Kui on antud algühend (A, B, π) (ja otseloomulikult vaatlused $\bar{\mathbf{x}}$), siis kõigepealt arvutatakse selle algühendi põhjal silumistõenäosused $\mathbf{P}(S_t = j \mid \bar{\mathbf{X}} = \bar{\mathbf{x}})$, $j \in K$, ja suurused $\mathbf{P}(S_t = j, S_{t-1} = i \mid \bar{\mathbf{X}} = \bar{\mathbf{x}})$ iga $i, j \in K$ ja $u > 1$ jaoks. Neid mõlemaid on võimalik arvutada näiteks Durand jt. (2004) kirjeldatud algoritmiga.

Järgmisena leiatakse parandatud hinnangud. Uueks algjaotuse lähendiks saab

$$\tilde{\pi}_j = \mathbf{P}(S_1 = j \mid \bar{\mathbf{X}} = \bar{\mathbf{x}}), j \in K.$$

Tegu on sisuliselt ahela esimese lüli seisundis j viibimise marginaalse (järel)tõenäosusega.

Üleminekutõenäosuste uuteks lähenditeks saavad arvud

$$\tilde{a}_{i,j} = \frac{\sum_{t=1}^{T-1} \mathbf{P}(S_{t+1} = j, S_t = i \mid \bar{\mathbf{X}} = \bar{\mathbf{x}})}{\sum_{t=1}^{T-1} \mathbf{P}(S_t = i \mid \bar{\mathbf{X}} = \bar{\mathbf{x}})}, i, j \in K.$$

Paneme tähele, et iga $i, j \in K$ korral arv $a_{i,j}$ kujutab endast üleminekute $i \rightarrow j$ oodatava sageduse suhet selliste olukordade oodatavasse sagedusse, kus seisundist i üldse kuskile üle mindi.

Analoogiliselt uuendatakse ka lähendid emissioonijaotustele, valides

$$\tilde{b}_j(x) = \frac{\sum_{t=1}^T I_{x_t=x} \mathbf{P}(S_t = i \mid \bar{\mathbf{X}} = \bar{\mathbf{x}})}{\sum_{t=1}^T \mathbf{P}(S_t = i \mid \bar{\mathbf{X}} = \bar{\mathbf{x}})}, j \in K, x \in \mathcal{X}.$$

Olgu nüüd $(\bar{\mathbf{S}}, \bar{\mathbf{X}})$ varjatud Markovi välja mudel puul (I, N) , mis on samuti kirjeldatav algjaotuse π , üleminekumaatriksi A ja emissioonitõenäosuste maatriksi B abil. Lähtudes alglähendist (π, A, B) , arvutame jällegi $\mathbf{P}(S_u = j \mid \bar{\mathbf{X}} = \bar{\mathbf{x}})$, $j \in K, u \in I$ ja $\mathbf{P}(S_u = j, S_{\rho(u)} = i \mid \bar{\mathbf{X}} = \bar{\mathbf{x}})$, $i, j \in K, u > 1$ ning valime

$$\begin{aligned} \tilde{\pi}_j &= \mathbf{P}(S_1 = j \mid \bar{\mathbf{X}} = \bar{\mathbf{x}}), j \in K, \\ \tilde{a}_{i,j} &= \frac{\sum_{u>1} \mathbf{P}(S_u = j, S_{\rho(u)} = i \mid \bar{\mathbf{X}} = \bar{\mathbf{x}})}{\sum_{u>1} \mathbf{P}(S_{\rho(u)} = i \mid \bar{\mathbf{X}} = \bar{\mathbf{x}})}, i, j \in K, \\ \tilde{b}_j(x) &= \frac{\sum_{u \in I} I_{x_u=x} \mathbf{P}(S_u = i \mid \bar{\mathbf{X}} = \bar{\mathbf{x}})}{\sum_{u \in I} \mathbf{P}(S_u = i \mid \bar{\mathbf{X}} = \bar{\mathbf{x}})}, j \in K, x \in \mathcal{X}. \end{aligned}$$

5.2 Simulatsioonid

Üldjoontes on katse samasugune kui alapeatükis 4.3. Endiselt on simuleeritud diskreetse emissioonijaotusega varjatud Markovi välja mudeleid (mudelid kirjeldatud tabelis 6) kuue erineva hargnemisastmega puu peal: ahel, vähe hargnev puu, keskmiselt hargnev puu, kahendpuu, kolmendpuu ja kümnendpuu. Küll aga ei ole lastud hübriidklassifitseerijatel kasutada mudelite tegelikke parameetreid, vaid selle asemel parameetrite Baum-Welchi algoritmiga saadud hinnanguid.

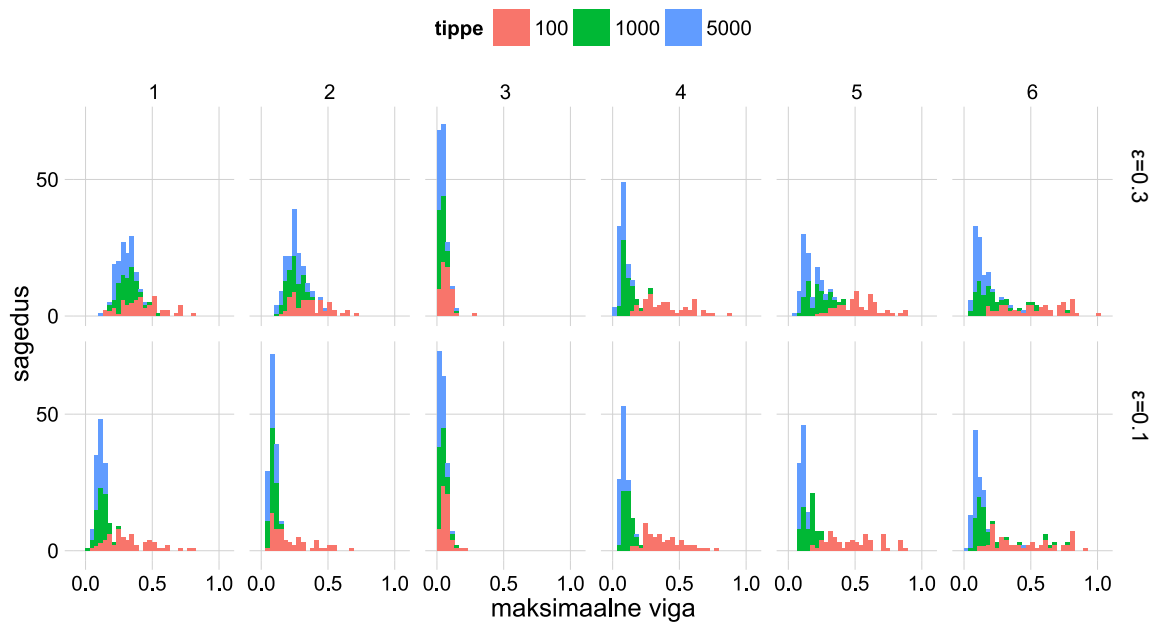
Baum-Welchi algoritm on iteratiivne algoritm, mis vajab algväärtuseid ja võib lõpetada tõepära lokaalses (mitte globaalses) maksimumis. Heade algväärtuste andmine Baum-Welchi algoritmile võib aga olla küllaltki keeruline. Käesolevas oleme läinud lihtsamat teed ja esitanud algväärtusteks mingil määral rikutud tegelikud väärtused. Täpsemalt, algjaotuse, üleminekutõenäosuste stohhastilise maatriksi ja emissioonijaotust kirjeldava stohhastilise maatriksi kõik elemendid p_i on asendatud sõltumatute juhuslike suuruste $P_i \sim U(\max(0, p - \varepsilon), \min(1, p + \varepsilon))$ realisatsioonidega ja seejärel on algjaotus ning kõik üleminekumaatriksi ja emissioonitõenäosuste maatriksi read normeeritud, et need summeeruksid üheks. Sisuliselt on igale elemendile p_i liidetud juhuslik viga jaotusega $U(-\varepsilon, \varepsilon)$, seega parameeter ε kontrollib rikutuse määra.

Antud olukorras on varjatud Markovi välja mudelil vabu parameetreid väga palju. Kui varjatud seisundeid on k ja emissioonijaotuse võimalikke väärtusi l , siis ilma täiendavaid kitsendusi lisamata oleks kokku vaja hinnata

$$(k - 1) + k(k - 1) + k(l - 1)$$

tõenäosust. Hinnangu täpsuse lühidaks kirjeldamiseks kasutame maksimaalset erinevust parameetrite hinnangute maatriksi ja tegelike väärtuste maatriksi elementide vahel:

$$\max_{i,j} |p_{i,j} - \hat{p}_{i,j}|,$$

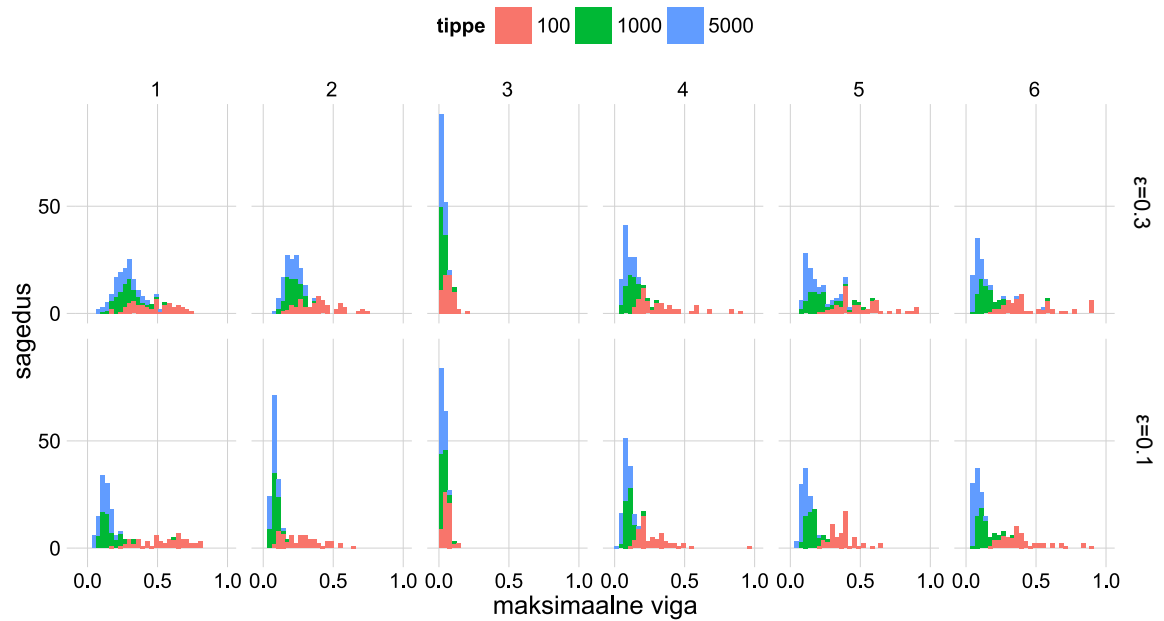


Joonis 8: Baum-Welchi algoritmi maksimaalne viga üleminekumaatriksi elementide hindamisel. Erinevad mudelid on veergudes (mudel 1 kuni mudel 6) ja puu suurused on eristatud värvidega. Alumises reas on Baum-Welchi algoritmile antud täpsem alg lähend. Erinevaid puu tüüpe ei ole eristatud. Igal puu tüübi, mudeli, puu suuruse ja alg lähendi täpsuse kombinatsioonil on tehtud 10 simulatsiooni.

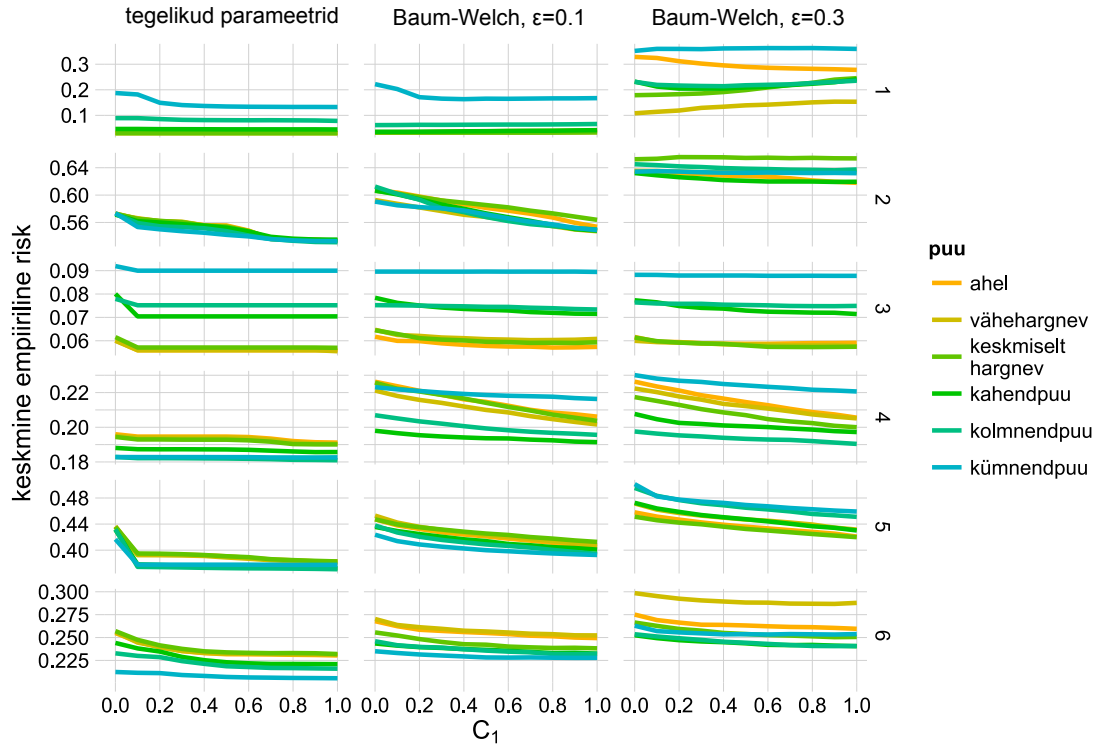
kus $(p_{i,j})_{i,j}$ on mingi parameetrite maatriks ja $(\hat{p}_{i,j})_{i,j}$ on vastavate hinnangute maatriks. Vaatame eraldi üleminekumaatriksi ja emissioonitõenäosuste maatriksi hinnangu täpsuseid. Algjaotuse hinnangud kippsid koonduma vektoriks, mille üks element on üks ja teised nullid – seetõttu nende hinnangute täpsust siin ei käsitle. Siiski on ka need algjaotuse hinnangud antud kasutada hübriidalgoritmile. Baum-Welchi algoritmi koondumiskriteeriumiks on lihtsuse mõttes valitud Raag (2009) eeskujul maksimaalne erinevus järjestikustel iteratsioonidel saadud parameetrite hinnangute vahel: kui see on väiksem kui mingi etteantud arv, siis otsingud lõpetatakse ja esitatakse viimati saadud lähend.

Näide 11 (Baum-Welchi algoritmiga saadud parameetrite hinnangute täpsus). Uurime üleminekumaatriksi ja emissioonijaotuse hinnangute täpsuseid, täpsuse all mõtleme maksimaalset erinevust hinnatud üleminekumaatriksi elementide ja vastavate tegelike elementide vahel. Joonistelt 9 ja 8 näeme, et hinnangute täpsus sõltub puu tippude arvust, mis ei ole üllatav – rohkem tippe, see tähendab rohkem vaatlusi, võimaldab täpsema hinnangu. Samuti paistab hinnangute täpsus mingil määral sõltuvat ka algväärtuste kvaliteedist, küll aga ei paista see sõltuvus olevat kuigi tugev. Vaid mudelite 1 ja 2 korral on erinevus silmnähtav. Paneme veel tähele, et mudel 3 võimaldab üsna täpselt hinnata oma parameetreid – arvatavasti seetõttu, et vaatlus reedab üsna üheselt varjatud seisundi (vaata mudeli 3 emissioonijaotust tabelis 6). Joonisel 8 on üleminekumaatriksi hinnangute maksimaalsete vigade

Näide 12 (Hübriidklassifitseerijate empiiriline risk mudeli parameetrite hinnangute



Joonis 9: Baum-Welchi algoritmi maksimaalne viga emissioonijaotust kirjeldava stohhastilise maatriksi elementide hindamisel. Erinevad mudelid on veergudes (mudel 1 kuni mudel 6) ja puu suurused on eristatud värvidega. Alumises reas on Baum-Welchi algoritmile antud täpsem alglähend. Erinevaid puu tüüpe ei ole eristatud. Igal puu tüübi, mudeli, puu suuruse ja alglähendi täpsuse kombinatsioonil on tehtud 10 simulatsiooni.



Joonis 10: Hübriidklassifitseerijate keskmised empiirilised riskid erinevate mudelite, puude ja erineva täpsusega teada olevate mudeli parameetrite korral. Kasutatud on 5000 tipuga puid. Iga puu tüübi (värvid), mudeli (ridades: mudel 1 kuni mudel 6) ja parameetrite teadaolemise taseme (veergudes) juures on tehtud 10 simulatsiooni, leitud iga simulatsiooni jaoks erinevad hübriidjoondused ($C_1 \in \{0, 0.1, \dots, 1\}$) ja arvutatud iga hübriidjoonduse keskmine empiiriline risk.

kasutamisel). Nüüd aga vaatame, kuidas mõjutab tegelike mudeli parameetrite asemel vaid hinnangute kasutamine hübriidklassifitseerijaid (nende empiirilisi riske). Joonisel 10 on võrreldud hübriidklassifitseerijate keskmisi empiirilisi riske kolmes erinevas olukorras:

1. hübriidklassifitseerija on saanud kasutada mudeli tegelikke parameetreid,
2. hübriidklassifitseerija on saanud kasutada Baum-Welchi algoritmi abil hinnatud parameetreid, kusjuures algväärtusteks on antud $\varepsilon = 0.1$ määral rikitud õiged parameetrid,
3. hübriidklassifitseerija on saanud kasutada Baum-Welchi algoritmi abil hinnatud parameetreid, kui algväärtusteks on antud $\varepsilon = 0.3$ määral rikitud õiged parameetrid.

Jooniselt 10 on näha, et hinnatud parameetrite kasutamine täpsuse asemel tõstab klassifitseerijate keskmist empiirilist riski (ehk vigade arvu), mis ei ole üllatav. Samuti on näha, et tulemust mõjutab ka Baum-Welchi algoritmile antavate algväärtuste kvaliteet, arvatavasti läbi hinnangute täpsuse. Muus osas paistab aga hinnangute kasutamine klassifitseerimiseks andma kvalitatiivselt sarnaseid tulemusi kui täpsuse

väärtuste kasutaminegi. Endiselt on PMAP klassifitseerija ($C_1 = 1$) kõige täpsem ja MAP klassifitseerija ($C_1 = 0$) antud hübriidklassifitseerijate seast kõige vähem täpne. Asjaolu, et paaril juhul ei kahane keskmine empiiriline risk monotoonselt, võib tuleneda väiksest valimimahust ja sellest, et esimese mudeli parameetrite hinnangud on küllaltki ebatäpsed (vaata jooniseid 9 ja 8).

Viited

- Petter Arnesen. Approximate recursive calculations of discrete Markov random fields. Master's thesis, Norwegian University of Science and Technology, 2010.
- Jeff A Bilmes jt. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. *International Computer Science Institute*, 4(510):126, 1998.
- Pierre Bremaud. *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*, kd. 31. Springer Science & Business Media, 1999.
- Nadezda A Bykova, Alexander V Favorov, ja Andrey A Mironov. Hidden Markov models for evolution and comparative genomics analysis. *PloS one*, 8(6):e65012, 2013.
- Ted Cox. An example of phase transition in countable one-dimensional Markov random fields. *Journal of Applied Probability*, lk 205–211, 1977.
- Jean-Baptiste Durand, Paulo Goncalves, ja Yann Guédon. Computational methods for hidden Markov tree models-an application to wavelet trees. *Signal Processing, IEEE Transactions on*, 52(9):2551–2560, 2004.
- Hans-Otto Georgii. *Gibbs measures and phase transitions*. Walter de Gruyter, 2. tr., 2011.
- Geoffrey Grimmett. *Probability on graphs: random processes on graphs and lattices*, kd. 1. Cambridge University Press, 2010.
- Jaime Huerta-Cepas, Joaquín Dopazo, ja Toni Gabaldón. ETE: a python environment for tree exploration. *BMC bioinformatics*, 11(1):24, 2010.
- Timo Koski. *Hidden Markov models for bioinformatics*. Springer Science & Business Media, 2001.
- Jüri Lember. Tehisõpe. loengukonspekt, 2013. internetiaadress: http://www.ms.ut.eesitesdefaultfiles/mstehisope_2013.pdf, vaadatud 07.05.2015.
- Jüri Lember ja Alexey A Koloydenko. Bridging Viterbi and posterior decoding: a generalized risk approach to hidden path inference based on hidden Markov models. *The Journal of Machine Learning Research*, 15(1):1–58, 2014.
- Jüri Lember, Alexey Koloydenko, ja Kristi Kuljus. Theory of segmentation. In Przemyslaw Dymarsky, toimetaja, *Hidden Markov Models, Theory and Applications*, ptk 3, lk 51–84. InTech, 2011.
- Stan Z Li. *Markov random field modeling in image analysis*. 3. tr., 2009.
- Leonid Peshkin ja Mikhail S Gelfand. Segmentation of yeast DNA using hidden Markov models. *Bioinformatics*, 15(12):980–986, 1999.

- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL <http://www.R-project.org/>.
- Mait Raag. Viterbi joonduse risk ja vigade arvu vähendamine. Master's thesis, Tartu Ülikool, 2009.
- Lawrence Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.
- Christian P. Robert ja George Casella. *Monte Carlo statistical methods*. Springer texts in statistics. Springer-Verlag New York, 1999.
- Sébastien Jocelyn Roch. *Markov models on trees: reconstruction and applications*. PhD thesis, University of California, Berkeley, 2007.
- Stefan Van Der Walt, S Chris Colbert, ja Gael Varoquaux. The NumPy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2):22–30, 2011.
- Hadley Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009. ISBN 978-0-387-98140-6. URL <http://had.co.nz/ggplot2/book>.
- Wang Xiang-yang, Sun Wei-wei, Wu Zhi fang, Yang Hong-ying, ja Wang Qin-yan. Color image segmentation using PDTDFB domain hidden Markov tree model. *Applied Soft Computing*, 2015.
- Stan Zachary. *Markov random fields and Markov chains on trees*. PhD thesis, Durham University, 1981.

Lisad

Lisa 1. Hübridaalgoritmi ja funktsioonid silumistõenäosuste arvutamiseks

```
import numpy as np
from math import log
from ete2 import *

def Marginal(puu, algjaotus, yleminekumaatriks):
    # Arvutab varjatud Markovi puu mudeli eeltõenäosused p(S).
    # Argumendid: ete2 Tree "puu",
    # juure marginaaljaotus NumPy maatriks "algjaotus" ja
    # üleminekutõenäosuste maatriks NumPy maatriks "yleminekumaatriks".

    puu.add_feature('marginal', algjaotus+0) # juure jaoks

    for tipp in puu.iter_descendants('preorder'): # teiste jaoks
        tipp.add_feature('marginal', np.dot(tipp.up.marginal, yleminekumaatriks))

def EmissionDist(vaatlus, maatriks=None):
    # Üldiselt, EmissionDist peaks tagastama antud vaatluse "vaatlus" jaoks tõepä
    # rade vektori,
    # üks arv iga varjatud seisundi kohta.
    # See konkreetne funktsioon siin on implementeeritud lõpliku arvu
    # võimalike väärtustega emissioonijaotuse jaoks, mis on
    # määratud stohhastilise maatriksiga "maatriks".

    # Varjatud seisundid peavad olema nummerdatud 0, 1, ..., k-1.

    return np.transpose(maatriks)[vaatlus]

def Upward(puu, yleminekumaatriks, emission_dist_maatriks=None, EmissionDistFun=
EmissionDist):
    # Durand et al. (2004) artiklist pärit upward recursion.
    # Kõik järgnevad "tähistused" on selle artikli põhjal.
    # Argumendid: ete2 Tree "puu", millele on rakendatud Marginal,
    # üleminekumaatriks NumPy maatriks "yleminekumaatriks",
    # emissioonijaotuse parameetrite maatriks NumPy maatriks "
    # emission_dist_maatriks",
    # funktsioon, mis "emission_dist_maatriks" alusel arvutab tõepärase vektori.

    # Kõrvalproduktina saame vaatluste log-tõepära. Initsialiseerime selle
    # ujukomaarvu.
    loglik = float(0)

    # Käime läbi kõik puu tipud, lehtedest juureni.
    for tipp in puu.traverse('postorder'):
        # Initsialiseerime vektori beta. Lehtede jaoks see saabki kohe beta-ks.
        beta = np.multiply(np.transpose(EmissionDistFun(tipp.vaatlus,
            emission_dist_maatriks)), tipp.marginal)
        if not tipp.is_leaf():
            # Kui käesolev tipp ei ole leht, siis on beta saamiseks vaja veel laste
            # beta_rho-sid.
            for laps in tipp.children:
                beta = np.multiply(beta, laps.beta_rho)
            N = np.sum(beta)
            loglik += np.log(N)
            tipp.add_feature('beta', np.divide(beta, N))
            # Arvutame beta_rho-d.
            tipp.add_feature('beta_rho', np.dot(np.dot(yleminekumaatriks, np.diag(1 /
                tipp.marginal)), tipp.beta))
    return loglik

def Downward(puu, yleminekumaatriks):
    # Durand et al. (2004) artiklist pärit downward recursion.
    # Argumendid: ete2 Tree "puu", millele on eelnevalt rakendatud Upward,
    # üleminekumaatriks NumPy maatriks "yleminekumaatriks".
```

```

# Mugavam on kasutada transponeeritud üleminekumaatriksit.
tyleminekumaatriks = np.transpose(yleminekumaatriks)

# Initsialiseerime juure jaoks silumistõenäosused.
puu.add_feature('smoothing', puu.beta+0)

# Käime läbi kõik ülejäänud tipud, lehed viimasena.
# Arvutame välja ka EM algoritmi jaoks vajalikud tõenäosused,
# millega üleminekutõenäosusi hinnata.
for tipp in puu.iter_descendants('preorder'):
    i_soltuv_osa = np.divide(tipp.up.smoothing, tipp.beta_rho)
    j_soltuv_osa = np.divide(tipp.beta, tipp.marginal)
    tipp.add_feature('ting_yl', np.multiply(np.transpose(np.multiply(
        i_soltuv_osa, tyleminekumaatriks)), j_soltuv_osa))
    tipp.add_feature('smoothing', np.sum(tipp.ting_yl, axis = 0))

def Hybrid(puu, C1, C2, algjaotus, yleminekumaatriks, emission_dist_maatriks=None,
joondusenimi='joondus', EmissionDistFun=EmissionDist):
    # Hübridalgoritm.
    # Argumendid: ete2 Tree "puu", millele on eelnevalt rakendatud Downward,
    # hübridriski konstandid (arvud) "C1" ja "C2",
    # juure marginaaljaotus NumPy maatriks "algjaotus",
    # üleminekumaatriks NumPy maatriks "yleminekumaatriks",
    # emissioonijaotuse parameetrite maatriks NumPy maatriks "
        emission_dist_maatriks",
    # joondusele antav nimi tekst "joondusenimi"
    # funktsioon, mis "emission_dist_maatriks" alusel arvutab tõepärase vektori.

    # Kontrollime, äkki on erijuht MAP või PMAP..
    # Kui üks parameeter on null, siis teise võib üheks võtta.
    if not C1: # if C1=0
        MAP(puu, algjaotus, yleminekumaatriks, emission_dist_maatriks, joondusenimi
            , EmissionDistFun)
    elif not C2: # if C2=0
        PMAP(puu, yleminekumaatriks, joondusenimi)
    else:
        # 1. arvutame delta-muutujad (analoogia Lember, Koloydenko (2014) artikliga
        ).
        for tipp in puu.traverse('postorder'):
            gamma = C1 * np.log(tipp.smoothing) + C2 * np.log(EmissionDistFun(tipp.
                vaatlus, emission_dist_maatriks))
            if not tipp.is_leaf():
                # Kui "tipp" ei ole leht, siis tuleb gammale midagi liita, et saada
                delta.
                for laps in tipp.children:
                    abimaatriks = C2 * np.log(yleminekumaatriks) + laps.delta
                    laps.add_feature('valik', np.argmax(abimaatriks, 1))
                    gamma = gamma + np.amax(abimaatriks, 1)
                # Vastasel juhul (kui tipp on leht), siis gamma ongi täpselt delta.
                tipp.add_feature('delta', gamma+0)

        # 2. optimaalse joonduse leidmine.
        puu.add_feature(joondusenimi, np.argmax(puu.delta + C2 * np.log(algjaotus))
            )

        for tipp in puu.iter_descendants('preorder'):
            tipp.add_feature(joondusenimi, tipp.valik[getattr(tipp.up, joondusenimi
                )]+0)

# Erijuhul kui C1=0, siis on tegu MAP joondusega ja kood lihtsustub veidi.
def MAP(puu, algjaotus, yleminekumaatriks, emission_dist_maatriks=None,
joondusenimi='joondus', EmissionDistFun=EmissionDist):

    for tipp in puu.traverse('postorder'):
        gamma = np.log(EmissionDistFun(tipp.vaatlus, emission_dist_maatriks))
        if not tipp.is_leaf():
            for laps in tipp.children:
                abimaatriks = np.log(yleminekumaatriks) + laps.delta
                laps.add_feature('valik', np.argmax(abimaatriks, 1))
                gamma = gamma + np.amax(abimaatriks, 1)
            tipp.add_feature('delta', gamma+0)

```

```

puu.add_feature(joondusenimi, np.argmax(puu.delta + np.log(algjaotus)))

for tipp in puu.iter_descendants('preorder'):
    tipp.add_feature(joondusenimi, tipp.valik[getattr(tipp.up, joondusenimi)
]+0)

# Erijuhul kui C2=0, siis on tegu PMAP joondusega ja kood lihtsustub veidi.
def PMAP(puu, yleminekumaatriks, joondusenimi='joondus'):
    # abimuutuja, et maatriksi kuju õige oleks.
    abi = np.multiply(0, yleminekumaatriks)
    for tipp in puu.traverse('postorder'):
        gamma = np.log(tipp.smoothing)
        if not tipp.is_leaf():
            for laps in tipp.children:
                abimaatriks = abi + laps.delta
                laps.add_feature('valik', np.argmax(abimaatriks, 1))
                gamma = gamma + np.amax(abimaatriks, 1)
            tipp.add_feature('delta', gamma+0)

    puu.add_feature(joondusenimi, np.argmax(puu.delta))

    for tipp in puu.iter_descendants('preorder'):
        tipp.add_feature(joondusenimi, tipp.valik[getattr(tipp.up, joondusenimi)
]+0)

```

Lisa 2. Funktsioonid simuleerimiseks

```

from ete2 import Tree
from random import random, randrange

def juhus(PropToenaosused):
    # Genereerib juhusliku täisarvu 0,1,2,...,n tõenäosusega, mis on
    # proportsionaalne mittenegatiivsete arvudega listis "PropToenaosused".
    i=0
    summa = PropToenaosused[0]
    juhuslikArv = random() * sum(PropToenaosused)
    while summa < juhuslikArv:
        i += 1
        summa += PropToenaosused[i]
    return i

def PuuGeneraator(lastearvujaotus, maxtippe=10000, fun=None):
    # Tekitab puu peaaegu nagu statsionaarse hargnemisprotsessi realisatsioonina.
    # Laste arvu jaotuse määrab list"lastearvujaotus", milles olevate
    # mittenegatiivsete
    # arvudega proportsionaalse tõenäosusega genereeritakse laste arv igale lehele.
    # Listi "lastearvujaotus" võib soovi korral eelnevalt modifitseerida
    # funktsiooniga "fun",
    # mis sõltub juba genereeritud tippude arvust ja maksimaalsest arvust.
    # Puu suurus on piiratud konstandiga "tippudearv" (kui protsess juba varem
    # välja ei sure).
    puu = Tree(name=1)
    tippe = 1 # tippude loendur
    tipud = [puu] # pikendatavate tippude list
    for tipp in tipud:
        if fun is None:
            lapsi = juhus(lastearvujaotus) # mitu last on tipul?
        else:
            lapsi = juhus(fun(lastearvujaotus, tippe, maxtippe))
        if lapsi: # Kui laste arv pole null, siis kirjutame tipule lapsi juurde
            for i in range(lapsi):
                tippe += 1
                if tippe > maxtippe:
                    break
                laps = Tree(name=tippe)
                tipud.append(laps) # käesolevat last peab ka for-tsükkel külastama
                tipp.add_child(laps)
        if tippe > maxtippe:
            break
    return puu

```

```

def EmissionDistGeneraator(hidden_state, emission):
    # Kui funktsioonile anda ette (üks) varjatud seisund (arvuna 0...k-1), siis
    # tagastab juhusliku arvu vastavast emissioonijaotusest.
    # Kui soovime m seisundiga diskreetset emissioonijaotust, siis
    # maatriks emission on stohhastiline k x m maatriks.
    # Kui soovime vaatlust mingist muust jaotusest, siis maatriksisse emission
    # tuleb mingil moel kirjutada emissioonijaotuste parameetrid (kuidas täpselt,
    # sõltub selle funktsiooni implementatsioonist.
    return juhus(emission[hidden_state])

def HMMGeneraator(puu, algjaotus, yleminekumaatriks, emission, Generaator=
EmissionDistGeneraator):
    # Vastavalt etteantud algjaotusele ja üleminekumaatriksile genereerib etteantud
    # (ete2 mooduli)
    # puu tippudele seisundi ja vaatluse maatriksiga emission määratud
    # emissioonijaotusest.
    puu.add_feature('seisund', juhus(algjaotus))
    puu.add_feature('vaatlus', Generaator(puu.seisund, emission))
    for tipp in puu.iter_descendants('preorder'):
        tipp.add_feature('seisund', juhus(yleminekumaatriks[tipp.up.seisund]))
        tipp.add_feature('vaatlus', Generaator(tipp.seisund, emission))
    return puu

# Potts'i mudeli genereerimine 1. peatüki näite jaoks
def LatticeNaabruskonnad(ridu, veerge):
    # Tekitab 1. järku naabruskonnad (listide listina) 2D võrel mõõtmega ridu x
    # veerge.
    tippe = ridu * veerge
    naabrid = []
    for tipp in range(tippe):
        naabrid.append([tipp-1, tipp+1, tipp-veerge, tipp+veerge])
    # Servade naabrite kärpimine
    # Ülemise korruse rahval pole ülemisi naabreid
    for i in range(veerge): naabrid[i][2] = None
    # Alumise korruse rahval pole alumisi naabreid.
    for i in range(tippe - veerge, tippe): naabrid[i][3] = None
    # Vasakpoolsel rahval pole vasakpoolseid naabeid.
    for i in range(0, tippe, veerge): naabrid[i][0] = None
    # Parempoolsel rahval pole parempoolseid naabeid.
    for i in range(veerge-1, tippe, veerge): naabrid[i][1] = None
    return naabrid

def alfa(minuSeisund, naabriSeisund, a):
    # Üldistatud Potts'i mudeli jaoks tarvilik funktsioon alfa.
    # Hetkel on implementeeritud siiski tavaline Potts'i mudel.
    return a * (minuSeisund == naabriSeisund)

def Potts'iUendaja(NaabriteKonfig, alfa, K, a):
    # Loosib uue seisundi seisundite [0,1,...,K-1] seast vastavalt (üldistatud)
    # Potts'i mudelile
    # parameetriga "a", arvestades naabrite seisundeid listis "NaabriteKonfig".
    LocalEnergy = []
    for seisund in range(K):
        LE = 0
        for naabriSeisund in NaabriteKonfig:
            LE += alfa(seisund, naabriSeisund, a)
        LocalEnergy.append(exp(LE))
    return juhus(LocalEnergy)

def GibbsSampler(Naabruskonnad, tsykleid, uendaja, a, K):
    # Genereerib Markovi välja realisatsiooni Gibbsi valikuga.
    # Graafi (I, N) naabruskonnad on määratud listide listiga "Naabruskonnad",
    # mudel on määratud funktsiooniga "uendaja",
    # (Potts'i mudelit silmas pidades) parameetriga "a" ja seisundite arvuga "K".
    # Gibbsi valiku toimimise illustreerimise huvides tagastab funktsioon
    # kõigi oma iteratsioonide tulemused kuni iteratsioonini "tsykleid".
    VanadSeisundid = []
    tippe = len(Naabruskonnad)
    for i in range(tippe):
        VanadSeisundid.append(randrange(K))
    KoikSeisundid = []
    for tsykkel in range(tsykleid):

```

```

Seisundid = VanadSeisundid[:]
for tipp in range(tippe):
    NaabriteSeisundid = [Seisundid[i] for i in Naabruskonnad[tipp] if i is
                        not None]
    Seisundid[tipp] = uuendaja(NaabriteSeisundid, alfa, K, a)
    KoikSeisundid.append(Seisundid)
    VanadSeisundid = Seisundid[:]
return KoikSeisundid

```

Lisa 3. Funktsioonid parameetrite hindamiseks

```

def ToenaosuseMuutja(toenaosus, maxmuutus):
    # Tagastab juhusliku arvu jaotusest
    # U(max(0, toenaosus-maxmuutus), min(1, toenaosus+maxmuutus))
    # Sisuliselt liidetakse arvule "toenaosus" juurde juhuslik arv jaotusest
    # U(-maxmuutus, maxmuutus) nii, et tulemus oleks ikka vahemikus (0,1).
    ylemine = min(1, toenaosus + maxmuutus)
    alumine = max(0, toenaosus - maxmuutus)
    return alumine + (ylemine-alumine)*random.random()

def AlgMudeliTekitaja(mudel, maxm=0.1):
    # Algväärtuste genereerimise funktsioon näiteks Baum-Welchi algoritmi jaoks.
    # Muudab etteantud mudeli (kolme elemendiga järjend:
    # algjaotus, üleminekumaatriks ja emissioonitõenäosuste maatriks)
    # parameetreid (tõenäosuseid) juhuslikult, maksimaalselt arvu "maxm" võrra.
    algjaotus0 = np.array([ToenaosuseMuutja(p, maxm) for p in mudel[0]])
    algjaotus0 = algjaotus0 / np.sum(algjaotus0)
    yleminek0 = np.array([[ToenaosuseMuutja(p, maxm) for p in rida] for rida in
                          mudel[1]])
    yleminek0 = np.transpose(np.transpose(yleminek0)/np.sum(yleminek0, 1))
    emission0 = np.array([[ToenaosuseMuutja(p, maxm) for p in rida] for rida in
                          mudel[2]])
    emission0 = np.transpose(np.transpose(emission0)/np.sum(emission0, 1))
    return [algjaotus0, yleminek0, emission0]

def HMMlogLik(puu, mudel, joondusenimi='joondus'):
    # Arvutab varjatud seisundite (mis võetakse joondusest) ja vaatluste
    # ühistõenäosuse.
    # Argumendid: ete2 Tree "puu", millele on rakendatud Downward ja mingi joondus,
    # kolme elemendiga järjend "mudel",
    # mille esimene element on algjaotust väljendav vektor np.array vormis,
    # teine element üleminekutõenäosusi väljendav stohhastiline maatriks
    # np.array vormis ja
    # kolmas element on emissioonijaotusi väljendav stohhastiline maatriks
    # np.array vormis.
    juureseisund = getattr(puu, joondusenimi)
    loglik = puu.Xloglik + np.log(puu.smoothing[juureseisund])
    for tipp in puu.iter_descendants():
        tipuseisund = getattr(tipp, joondusenimi)
        tipuvanemaseisund = getattr(tipp.up, joondusenimi)
        loglik += np.log(tipp.ting_yl[tipuvanemaseisund][tipuseisund]/tipp.
                        smoothing[tipuseisund])
    return loglik

def Vahed(mudel1, mudel2):
    # Leiab kahe mudeli parameetrite vahelise suurima erinevuse ja mediaanerinevuse
    # eraldi algjaotuse, üleminekumaatriksi ja emissioonijaotusi määrava maatriksi
    # elementide vahel.
    maxvahed = []
    medvahed = []
    for i in range(len(mudel1)):
        maxvahed.append(np.max(np.abs(mudel1[i] - mudel2[i])))
        medvahed.append(np.median(np.abs(mudel1[i] - mudel2[i])))
    return [maxvahed, medvahed]

def PeatumisIndikaator(uus_mudel, vana_mudel, epsilon):
    # Baum-Welchi algoritmi peatamise määramine (Mait Raagi magistritöö järgi)
    # kui suurim muutus uue ja vana mudeli parameetrite vahel on väiksem kui arv "
    # epsilon",
    # siis tagastab True.
    return max(Vahed(uus_mudel, vana_mudel)[0]) < epsilon

```

```

def BaumWelchDiskreetneSamm(puu, vana_mudel):
    # Uuendab vanad parameetrite hinnangud Baum-Welchi algoritmiga erijuhul, kui
    # emissioonijaotus on diskreetne
    # ja vaatlused on 0, 1, 2, ..., k-1.
    Marginal(puu, vana_mudel[0], vana_mudel[1])
    loglik = Upward(puu, vana_mudel[1], vana_mudel[2])
    Downward(puu, vana_mudel[1])
    # Algjaotuse hinnang
    uus_algjaotus = puu.smoothing + 0 # Et Python teeks koopia, mitte uue viite...
    # Üleminekumaatriksi ja emissioonijaotuse hinnang
    # Initsialiseerime üleminekumaatriksi ja emissioonijaotuse hindamiseks
    # vajalikud vektorid/maatriksid
    smoothing_summa = 0*puu.smoothing
    smoothing_summa2 = puu.smoothing + 0
    ting_yl_summa = 0*puu.children[0].ting_yl
    vaatluste_arv = np.shape(vana_mudel[2])[1]
    emission_summa = np.array([(puu.vaatlus == k)*puu.smoothing[:] for k in range(
        vaatluste_arv)])
    for tipp in puu.iter_descendants():
        smoothing_summa += tipp.up.smoothing
        ting_yl_summa += tipp.ting_yl
        smoothing_summa2 += tipp.smoothing
        emission_summa[tipp.vaatlus] += tipp.smoothing
    # Hinnangute rehkendamine ja väljastamine
    uus_yleminek = np.transpose(np.transpose(ting_yl_summa) / smoothing_summa)
    uus_emission = emission_summa / smoothing_summa2
    return [uus_algjaotus, uus_yleminek, np.transpose(uus_emission)]

def BaumWelchDiskreetne(puu, mudel, epsilon=0.01, maxiter=1000):
    # Baum-Welchi algoritm HMM parameetrite hindamiseks erijuhul, kui
    # emissioonijaotus on diskreetne
    # ja vaatlused on 0, 1, 2, ..., k-1.
    i = 0
    vana_mudel = mudel[:]
    while i < maxiter:
        uus_mudel = BaumWelchDiskreetneSamm(puu, vana_mudel)
        if PeatumisIndikaator(uus_mudel, vana_mudel, epsilon):
            break
        vana_mudel = uus_mudel[:]
        i += 1
    return uus_mudel

```

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Mark Gimbutas,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose **“Varjatud Markovi välja mudelid puudel”**, mille juhendaja on Jüri Lember,
 - 1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
 - 1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartu 13.05.2015