

Simuloidun jäädytyksen suppenemislause

Antti Luoto

Matematiikan pro gradu

Jyväskylään yliopisto
Matematiikan ja tilastotieteen laitos
Kevät 2013
2013-06-19

Tiivistelmä: Antti Luoto, *Simuloidun jäähtymisen suppenemislause* (engl. *Convergence theorem for simulated annealing*), matematiikan pro gradu -tutkielma, 82 s., Jyväskylän yliopisto, Matematiikan ja tilastotieteen laitos, kevät 2013.

Tämä pro gradu -tutkielma käsittelee simuloitu jäähtymys -nimisen kombinatorisen optimointimenetelmän teoriaa ja käytäntöä. Esimerkiksi kuvankäsittelyssä sovelletun algoritmin ideana on löytää annetulla joukolla määritellyn reaaliarvoisen energiefunktion globaali minimikohta sallimalla – ei pelkästään energiaa vähentäviä – vaan myös energiaa kasvattavia siirtymiä lähtöjoukon alkioiden välillä. Tilastolliseen fysiikkaan analogian omaavan, Gibbsin jakauman ominaisuuksiin pohjautuvan menetelmän matemaattisena perustana toimivat epähomogeeniset Markovin ketjut, joiden suppenemista tarkastellaan Dobrushinin kontraktiokerroinmenetelmän avulla.

Simuloidun jäähtymisen suppenemislause, joka takaa epähomogeenisen Markov Chain Monte Carlo -ketjun suppenemisen minimienergiatilojen joukkoon, todistetaan aluksi Gibbs-otannan tapauksessa sekä deterministisille että satunnaisille päivitysjonoille. Tätä varten tehdään katsaus satunnaiskenttien, naapurustojärjestelmien, klikkien ja potentiaalien teoriaan.

Suppenemislause todistetaan myös yleisempiin tilanteisiin soveltuvan Metropolis-otannan tapauksessa. Metropolis-algoritmilla ajettavaa simuloitua jäähtymistä sovelletaan lopuksi konkreettiseen ongelmaan, jossa pyritään muodostamaan suorakulmion muotoiselle tenttisalille vilpin estämiseksi optimaalinen istumajärjestys. Simulointikokeiden yhteydessä pohditaan menetelmän käytännön soveltamiseen liittyvää problematiikkaa ja pohditaan lopuksi sitä, kuinka hyvin jäähtymys soveltuu tenttisaliongelman ratkaisemiseen.

Avainsanat: Gibbs-otanta, epähomogeeninen Markovin ketju, Metropolis-algoritmi, satunnaiskenttä, simuloitu jäähtymys

Sisältö

Johdanto	1
Luku 1. Esitietoja ja merkintöjä	3
1.1. Todennäköisyysavaruudet	3
1.2. Äärellinen todennäköisyysavaruus	5
1.3. Äärellisten todennäköisyysjakaumien avaruus	5
1.4. Stokastisten matriisien avaruus	9
Luku 2. Satunnaiskenttien alkeita	13
2.1. Satunnaiskentän jakaumat, Markovin satunnaiskenttä	13
2.2. Satunnaiskentän marginaalijakaumat ja ehdolliset jakaumat	14
2.3. Gibbsin kentät	16
Luku 3. Epähomogeenisen Markovin ketjun heikko ja vahva ergodisuus	21
3.1. Epähomogeeninen Markovin ketju	21
3.2. Ääretön tulo	26
3.3. Epähomogeenisen Markovin ketjun suppeneminen	28
3.4. Heikko ergodisuus	29
3.5. Vahva ergodisuus	34
Luku 4. Gibbs-otanta	39
4.1. Gibbs-otannan päivitysjonoista ja suppenemisesta	40
4.2. Menetelmän nimestä ja alkuperästä	41
4.3. Suppenemistulokset, johdanto	41
4.4. Deterministinen päivitysjono	42
4.5. Satunnainen päivitysjono	46
Luku 5. Simuloidun jäähtytyksen teoria ja Gibbs-jäähtytyksen suppeneminen	49
5.1. Taustoja ja algoritmin esittelyä	49
5.2. Gibbs-jäähtytyksen algoritmi	53
5.3. Deterministinen päivitysjono	53
5.4. Satunnainen päivitysjono	56
Luku 6. Metropolis-dynamiikalla varustettu simuloitu jäähtytys	59
6.1. Metropolis-algoritmi	59
6.2. Metropolis-algoritmin ja Metropolis-jäähtytyksen suppeneminen	61
6.3. Optimaalinen vakio jäähtytysaikataululle	66
Luku 7. Tenttialiongelma ja simuloitu jäähtytys käytännössä	69
7.1. Tenttialiongelman sanallinen kuvailu	69
7.2. Tenttialiongelmaan liittyvän matemaattisen mallin muodostaminen	69

7.3. Metropolis-jäähdytysketjun matemaattinen kuvailu	71
7.4. Empiiriset tarkastelut	73
7.4.1. Energiafunktion skaalaaminen ja jäähdytysvakion valinta	74
7.4.2. Simulointikokeita eri jäähdytysvakioiden arvoilla	75
7.4.3. Ehdotusmatriisin ja alkukonfiguraation vaikutus	76
7.5. Johtopäätökset	79
Kirjallisuutta	81

Johdanto

Tämä tutkielman tarkoituksena on johdattaa lukija simuloitu jäähdytys -nimisen kombinatorisen optimointimenetelmän teoriaan ja käytäntöön. Markov Chain Monte Carlo -menetelmiin pohjautuvan algoritmin avulla voidaan (ainakin teoriassa) ratkaista ongelmia, joissa reaaliarvoinen funktio $U : E \rightarrow \mathbb{R}$ pyritään minimoimaan annetun äärellisen joukon E yli. Ensimmäisen matemaattisen todistuksen simuloitun jäähdytyksen suppenemiselle antoivat veljekset S. ja D. Geman vuonna 1984 julkaisussa kuvankäsittelyä, Gibbs-otantaa ja Bayes-tilastotiedettä koskevassa artikkelissaan [8]. Siinä simuloitun jäähdytyksen algoritmin pohjana käytettiin Gibbs-otannan varhaista versiota, jossa siirtymät alkioiden välillä suoritettiin poimimalla alkioita äärellisellä karteesisella tulojoukolla määritellyn Gibbsin jakauman täysistä ehdollisista jakaumista.

Edellä mainitun tuloksen todistamista ja täsmällistä muotoilua varten tarvitaan jonkin verran stokastiikan ja tilastotieteen eri osa-alueisiin liittyviä käsitteitä ja tuloksia. Luvussa 1 käydään aluksi läpi todennäköisyysteorian perusasioita ja merkintöjä, jonka jälkeen siirrytään tarkastelemaan äärellisiin todennäköisyysavaruuksiin liittyvien jakaumien ja stokastisten matriisien ominaisuuksia. Lisäksi käydään läpi jakaumien välisten etäisyyksien mittaamiseen käytettävän kokonaisvaihteluetaisyyden keskeisiä ominaisuuksia.

Satunnaiskenttiä käsittelevä Luku 2 toimii matemaattisena alustana Gibbs-otannan ja siihen perustuvan simuloitun jäähdytyksen algoritmeille. Luvussa esiteltävää liittyvää käsitteistöä sovelletaan myös Luvussa 7 tarkasteltavan käytännön ongelman matemaattisen mallin muotoilussa.

Simuloitun jäähdytyksen matemaattisena perustana toimivat epähomogeeniset Markovin ketjut, joiden teoriaa ja suppenemisominaisuuksia käsitellään Luvussa 3. Jos epähomogeenisen Markovin ketjun $(X_n)_{n \geq 0}$ riippuvuus alkujakaumasta häviää asympotoottisesti (heikko ergodisuus), ja mikäli sen siirtymämatriisien $(P_n)_{n \geq 1}$ invarianttien jakaumien jono $(\mu_n)_{n \geq 1}$ suppenee riittävän nopeasti johonkin rajajakaumaan μ_∞ , Lauseen 3.5.2 nojalla $\mathbb{P}_\nu(X_n = x) \xrightarrow{n \rightarrow \infty} \mu_\infty(x)$ kaikilla alkujakaumilla ν ja tilaavaruuden E alkioilla x . Käsiteltävien simuloitun jäähdytyksen suppenemistulosten todistukset perustuvat juuri tähän, R. L. Dobrushinin mukaan nimettyyn tulokseen.

Gibbs-otantaa käsittelevässä Luvussa 4 perehdytään kyseisen menetelmän teoriaan ja todistetaan suppenemislauseet sopivasti valituille deterministisille ja satunnaisille päivitysjonuille. Myös Gibbs-jäähdytystä koskeva suppenemislause todistetaan sekä determinisen että satunnaisen päivitysjonon tapauksessa Luvussa 5. Artikkelin [8] varsin tiiviiseen formaattiin puetut, deterministisiä jonoja koskevat todistukset on pyritty päivittämään lukijaystävällisempään muotoon mukaillen kirjan [30] kontraktiokertoimen käyttöön perustuvaa lähestymistapaa. Satunnaista päivitystä koskevat,

huomattavasti suoraviivaisemmat tulokset todistusideoineen ovat peräisin samasta kirjasta [30].

Kaksi viimeistä lukua käsittelevät tutkielman toista pääteemaa eli konkreettista ongelmaa, jossa pyritään optimoimaan tenttialin istumajärjestys vilpin estämiseksi. Eri istumajärjestyksiä vastaavien alkioiden arvottamista varten muodostetaan tenttialin lokaaleihin ominaisuuksiin perustuva energiafunktio. Koska istumajärjestyksistä koostuvaa konfiguraatioavaruutta ei voida esittää karteesisessa tulomuodossa, päädytään soveltamaan varhaisimmista simuloitua jäähdytystä käsittelevistä artikkeleista [18] ja [4] peräisin olevaa, menetelmän Metropolis-algoritmiin perustuvaa versiota. Luku 6 perustuukin kyseisen jäähdytysalgoritmin Gibbs-jäähdytykselle analogisen suppenemislauseen todistamiselle. Luvussa 7 rakennetaan aluksi matemaattinen malli tenttialiongelmalle, jolle edellä mainitun Lauseen 6.2.5 oletukset toteutuvat. Tämän jälkeen suoritetaan joukko logaritmiseen jäähdytysaikatauluun perustuvan Metropolis-jäähdytyksen ominaisuuksia havainnollistavia simulointikokeita. Saatujen tulosten tulkintojen ja johtopäätösten teon ohessa pohditaan simuloidun jäähdytyksen käytännön soveltamiseen liittyviä etuja ja ongelmia.

Tutkielman ensisijaisena lähdemateriaalina on toiminut G. Winklerin Gibbsin kenttiä, kuva-analyysiä ja MCMC-menetelmiä käsittelevä kirja [30]. Nyrkkisääntönä voidaan pitää, että jollei erikseen ole mainittu, likipitäen kaikki tutkielmassa esiintyvät epätriviaalit tulokset ovat joko kokonaan, osittain tai ainakin ideatasolla peräisin tästä kirjasta. Luvuissa 4 ja 5 esiintyvät todistukset perustuvat artikkelin [8] ohella käytännössä yksinomaan kirjan [30, Chapter 5] esitykseen. Lisäksi kokonaisvaihteluetaisyyteen, satunnaiskenttiin, kontraktiokertoimeen, heikkoon ja vahvaan ergodisuuteen ja Metropolis-jäähdytykseen liittyvät tulokset ovat enemmän tai vähemmän peräisin tästä kirjasta.

Muista keskeisistä lähdekirjoista mainittakoot P. van Laarhovenin ja E. Aartsin [19] (erityisesti simuloidun jäähdytyksen teoriaa koskeva osuus), D. Isaacsonin ja R. Madsenin [14] (äärettömät tulot, heikon ja vahvan ergodisuuden teoria), P. Brémaudin [3] (satunnaiskentät) ja lisäksi D. Levinin, Y. Peresin ja E. Wilmerin [21] sekä O. Häggströmin [15] (Markovin ketjujen teoria). Luvun 1 sisältö on koostettu suurimmaksi osaksi luentomonisteissa [17], [9], [29], [25] sekä [16] käsiteltävistä asioista.

Kirjoittajan itse todistamista tuloksista merkittävimpinä voitaneen pitää artikkeleista [8] pois jätettyjä tuloksia vastaavia Lemmoja 4.4.4 ja 5.1.2. Tämän lisäksi muutamien esimerkkien, huomautusten, täydennysten ja liki triviaalien pikkutulosten todistusten laatimisen ohella tutkielman tekijän kädenjälki näkyy lähinnä käytännön ongelmaa käsittelevässä Luvussa 7. Erityismaininta Lemman 7.3.2 todistuksen idean antamisesta kuuluu H. Hartikaiselle.

Kaikki simulaatiot suoritettiin käyttämällä R-ohjelmaa (versio 2.15.1) [26], jota sovellettiin Paint -ohjelman (Microsoft Windows 7, versio 6.1) ohella myös kuvien piirtämiseen. Kaikki istumajärjestyksiä ilmentävät ruudukkokuvaajat on tehty käyttämällä R:n Plotrix-pakettia [20]. Esimerkin 3.1.4 simuloinneissa hyödynnettiin L. Leskelän kirjoittamia R-funktioita. Muiden simulaatioiden koodia laadittaessa suurena apuna toimi A. Penttisen luentomoniste [24].

Esitietoja ja merkintöjä

1.1. Todennäköisyysavaruudet

Olkoon Ω mielivaltainen joukko ja olkoon $\mathcal{P}(\Omega) = \{A : A \subset \Omega\}$ kaikkien joukon Ω osajoukkojen kokoelma; otosavaruuden *potenssijoukko*.

MÄÄRITELMÄ 1.1.1 (Sigma-algebra). Kokoelma $\mathcal{F} \subset \mathcal{P}(\Omega)$ on *sigma-algebra* joukolla Ω , mikäli seuraavat ehdot ovat voimassa:

- (i) $\emptyset, \Omega \in \mathcal{F}$
- (ii) $A^C := \Omega \setminus A \in \mathcal{F}$ kaikilla $A \in \mathcal{F}$
- (iii) Jos $A_1, A_2, \dots \in \mathcal{F}$, niin $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

MÄÄRITELMÄ 1.1.2 (Todennäköisyys). Olkoon \mathcal{F} sigma-algebra joukolla Ω . Tällöin joukkofunktio $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ on *todennäköisyyssmitta* tai *todennäköisyys*, mikäli

- (i) $\mathbb{P}(\Omega) = 1$
- (ii) $\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$, kun $A_1, A_2, \dots \in \mathcal{F}$ ja $A_i \cap A_j = \emptyset$ kaikilla $i \neq j$.

HUOMAUTUS 1.1.3. Tapahtumien $A, B \in \mathcal{F}$ leikkauksen todennäköisyydelle käytetään usein myös merkintää $\mathbb{P}(A, B)$.

Kolmikkoa $(\Omega, \mathcal{F}, \mathbb{P})$, missä \mathcal{F} on otosavaruuden Ω sigma-algebra ja \mathbb{P} todennäköisyys sigma-algebralla \mathcal{F} , kutsutaan *todennäköisyysavaruudeksi*. Joukosta Ω käytetään nimitystä *otosavaruus* ja sigma-algebran \mathcal{F} osajoukkoja $A \subset \mathcal{F}$ kutsutaan *tapahtumiksi*. Paria (Ω, \mathcal{F}) kutsutaan *mitalliseksi avaruudeksi*.

MÄÄRITELMÄ 1.1.4 (Riippumattomuus). Olkoon $(\Omega, \mathcal{F}, \mathbb{P})$ todennäköisyysavaruus ja $\emptyset \neq I$ jokin mielivaltainen indeksijoukko. Tällöin tapahtumat $(A_i)_{i \in I} \subset \mathcal{F}$ ovat riippumattomia, mikäli kaikille äärellisille, eri indekseistä koostuville jonoille $i_1, \dots, i_n \in I$ on voimassa ehto

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_n}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \dots \mathbb{P}(A_{i_n}).$$

MÄÄRITELMÄ 1.1.5 (Ehdollinen todennäköisyys). Olkoon $A, B \in \mathcal{F}$ tapahtumia, ja olkoon $\mathbb{P}(B) > 0$. Tapahtuman A *ehdollinen todennäköisyys* ehdolla tapahtuma B on

$$\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

HUOMAUTUS 1.1.6. Olkoot $A, B, C \in \mathcal{F}$ ja $\mathbb{P}(B) > 0$. Tällöin

- (i) Kuvaus $\mathbb{Q} : \mathcal{F} \rightarrow \mathbb{R}$; $\mathbb{Q}(A) = \mathbb{P}(A | B)$ on todennäköisyysmitta avaruudella (Ω, \mathcal{F}) .
- (ii) Jos tapahtumat A ja B ovat riippumattomia, niin $\mathbb{P}(A | B) = \mathbb{P}(A)$.
- (iii) Sanotaan, että tapahtumat A ja C ovat *ehdollisesti riippumattomia ehdolla tapahtuma B* , mikäli $\mathbb{P}(A, C | B) = \mathbb{P}(A | B)\mathbb{P}(C | B)$.

MÄÄRITELMÄ 1.1.7 (Satunnaismuuttuja ja -alkio). Olkoot (Ω, \mathcal{F}) ja (E, \mathcal{E}) mittallisia avaruuksia. Funktio $X : \Omega \rightarrow \mathbb{R}$ on *satunnaismuuttuja*, jos $X^{-1}(B) \in \mathcal{F}$ kaikilla $B \in \mathcal{B}(\mathbb{R})$, missä $\mathcal{B}(\mathbb{R})$ on reaaliakselin avoimien joukkojen virittämä sigma-algebra, niin sanottu *Borelin sigma-algebra*. Kuvaus $Y : \Omega \rightarrow E$ on *satunnaisalkio*, jos puolestaan $Y^{-1}(B) \in \mathcal{F}$ kaikilla $B \in \mathcal{E}$.

HUOMAUTUS 1.1.8. Yksinkertaistetaan kielenkäyttöä kutsumalla jatkossa E -arvoisia satunnaisalkioita E -arvoisiksi satunnaismuuttujiksi.

MÄÄRITELMÄ 1.1.9 (Indikaattorifunktio). Olkoon $(\Omega, \mathcal{F}, \mathbb{P})$ todennäköisyysavaruus ja olkoon $A \in \mathcal{F}$. Määritellään satunnainen *indikaattorifunktio* $\mathbb{1}_A : \Omega \rightarrow \mathbb{R}$ asettamalla

$$\mathbb{1}_A(\omega) = \begin{cases} 1, & \text{jos } \omega \in A \\ 0, & \text{jos } \omega \notin A \end{cases}.$$

Määritellään edelleen mielivaltaisen joukon E osajoukkoa $A \subset E$ vastaava deterministinen indikaattorifunktio $\mathbb{1}_A : E \rightarrow \mathbb{R}$ asettamalla

$$\mathbb{1}_A(x) = \begin{cases} 1, & \text{jos } x \in A \\ 0, & \text{jos } x \notin A \end{cases}.$$

HUOMAUTUS 1.1.10.

- (i) Indikaattorifunktion argumentti jätetään toisinaan kirjoittamatta mikäli se on helposti pääteltävissä asiayhteydestä.
- (ii) Indikaattorifunktiosta $\mathbb{1}\{\tilde{\omega} \in \Omega : X(\tilde{\omega}) \in B\}(\omega)$ käytetään lyhennysmerkintää $\mathbb{1}_{\{X \in B\}}$, missä X on satunnaismuuttuja avaruudella $(\Omega, \mathcal{F}, \mathbb{P})$ ja $B \in \mathcal{B}(\mathbb{R})$.
- (iii) Kirjoitetaan jatkossa hankalat, muotoa $\mathbb{1}\{\tilde{x} \in E : \tilde{x} \text{ toteuttaa ehdon } \mathcal{P}\}(x)$ olevat funktiot lyhyemmin muodossa $\mathbb{1}\{x \text{ toteuttaa ehdon } \mathcal{P}\}$.

MÄÄRITELMÄ 1.1.11 (Odotusarvo). Määritellään todennäköisyysavaruuden $(\Omega, \mathcal{F}, \mathbb{P})$ satunnaismuuttujan X *odotusarvo* asettamalla

$$\mathbb{E}X := \int_{\Omega} X(\omega) d\mathbb{P}(\omega), \text{ mikäli } \mathbb{E}|X| < \infty.$$

Jos $\mathbb{E}|X| = \infty$, odotusarvoa ei määritellä.

HUOMAUTUS 1.1.12. Jos X on jatkuva satunnaismuuttuja *tiheysfunktioaan* f , muuttujanvaihto (ks. esim. [27, s. 196-197]) antaa

$$\mathbb{E}X = \int_{\mathbb{R}} xf(x) dx.$$

Diskreetille satunnaismuuttujalle X , jonka arvojoukko on $\{x_1, x_2, \dots, x_n\}$, ja jolla on *pistetodennäköisyysfunktio* p , saadaan puolestaan

$$\mathbb{E}X = \sum_{i=1}^n x_i p(x_i).$$

Siirrytään seuraavaksi tarkastelemaan äärellisiä todennäköisyysavaruuksia.

1.2. Äärellinen todennäköisyysavaruus

Olkoon Ω otosavaruus, jolle $|\Omega| < \infty$. Todetaan, että tällöin $\mathcal{P}(\Omega)$ täyttää sigma-algebran ehdot. Näin on, sillä se on kaikki yksiöt $\{\omega\}_{\omega \in \Omega}$ sisältävä äärellinen $2^{|\Omega|}$ joukon kokoelma. Tästä syystä äärelliselle otosavaruudelle voidaan aina valita (ja valitaan) tapahtumien joukoksi eli sigma-algebraksi potenssijoukko $\mathcal{P}(\Omega)$. Olkoon sitten funktio $p : \Omega \rightarrow [0, 1]$, jolla on ominaisuudet

$$(i) \quad p(\omega) \geq 0 \text{ kaikilla } \omega \in \Omega$$

$$(ii) \quad \sum_{\omega \in \Omega} p(\omega) = 1.$$

Tällaista funktiota kutsutaan *pistetodennäköisyysfunktioiksi*. Funktio p määrää yksikäsitteisen todennäköisyysmitan \mathbb{P} tapahtumien joukolle $\mathcal{P}(\Omega)$, kun asetetaan

$$\mathbb{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]; \quad \mathbb{P}(A) := \sum_{\omega \in A} p(\omega) \text{ kaikilla } A \subset \Omega.$$

Voidaan nyt koota *äärellinen todennäköisyysavaruus* $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. Jatkossa siis todennäköisyysavaruus oletetaan aina äärelliseksi, mikäli siihen liittyvä sigma-algebra on otosavaruuden potenssijoukko. Todennäköisyysavaruutta $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ voidaan tällöin merkitä myös lyhyemmin parina (Ω, \mathbb{P}) .

MÄÄRITELMÄ 1.2.1 (Satunnaismuuttuja ja -alkio). Olkoon $E \neq \emptyset$ mielivaltainen joukko. Kuvausta $X : \Omega \rightarrow E$ äärelliseltä otosavaruudelta Ω kutsutaan E -arvoiseksi satunnaismuuttujaksi (tai satunnaisalkioksi).

HUOMAUTUS 1.2.2. Tästä lähtien puhuttaessa äärellisen otosavaruuden E -arvoisesta satunnaismuuttujasta oletetaan implisiittisesti, että $X(\Omega) = E$, ts. maaliavaruus ei sisällä sellaisia pisteitä, joihin mikään $\omega \in \Omega$ ei kuvaudu. Tämä oletus takaa, että otosavaruuden Ω äärellisyyden nojalla myös E on äärellinen, joten voidaan aina varustaa E sigma-algebralla $\mathcal{P}(E)$, jolloin edelleen X on $(\mathcal{P}(\Omega), \mathcal{P}(E))$ -mitallinen. Siten Määritelmä 1.2.1 on sopusoinnussa Määritelmän 1.1.7 kanssa.

Oletetaan nyt, että X on E -arvoinen satunnaismuuttuja äärellisellä todennäköisyysavaruudella (Ω, \mathbb{P}) . Määritellään joukkofunktio $\mathbb{P}_X : \mathcal{P}(E) \rightarrow \mathbb{R}$ asettamalla

$$\mathbb{P}_X(B) := \mathbb{P}(X \in B) = \mathbb{P}(\{\omega \in \Omega : X \in B\}) = \sum_{\omega \in \Omega} p_X(\omega) \mathbf{1}_{\{X \in B\}}(\omega),$$

kaikilla $B \subset E$, missä p_X on *satunnaismuuttujan X pistetodennäköisyysfunktio*. Sen indusoimaa todennäköisyyttä \mathbb{P}_X kutsutaan satunnaismuuttujan X *jakaumaksi*. Ei ole vaikeaa osoittaa, että kolmikko $(E, \mathcal{P}(E), \mathbb{P}_X)$ muodostaa äärellisen todennäköisyysavaruuden.

1.3. Äärellisten todennäköisyysjakaumien avaruus

Edustakoon $E \neq \emptyset$ tämän luvun ajan kiinnitettyä mielivaltaista äärellistä joukkoa. Tarkastellaan funktioiden avaruutta $\mathbb{R}^E = \{f \mid f : E \rightarrow \mathbb{R}\}$. Voidaan osoittaa, että \mathbb{R}^E on tällöin äärellisulotteinen vektoriavaruus. Sen alkiot eli vektorit ovat

siis kuvauksia joukolta E reaalitylukujen joukolle, joita voidaan merkitä myös lyhyesti muodossa $f = (f(x))_{x \in E}$. Varustamalla joukko \mathbb{R}^E L^1 -normilla

$$\|f\| := \|f\|_{L^1} = \sum_{x \in E} |f(x)|$$

saadaan lopputuloksena täydellinen normiavaruus eli *Banachin avaruus* $(\mathbb{R}^E, \|\cdot\|)$.¹ L^1 -normi indusoi joukkoon \mathbb{R}^E metriikan, niin sanotun *kokonaisvaihteluetaisyden*:

$$d_{TV}(f, g) := \|f - g\| = \sum_{x \in E} |f(x) - g(x)|.$$

Määritellään sitten joukko

$$\mathcal{D}_E := \left\{ \mu : E \rightarrow \mathbb{R} \mid \sum_{x \in E} \mu(x) = 1 \text{ ja } \mu(x) \geq 0 \text{ kaikilla } x \in E \right\}$$

Banach-avaruuden $(\mathbb{R}^E, \|\cdot\|)$ yksikköpallon reunan $\partial B(0, 1) = \{f \in \mathbb{R}^E \mid \|f\| = 1\}$ osajoukko \mathcal{D}_E koostuu siis kaikista pistetodennäköisyysfunktioista joukolla E . Vektoriiksi $(\mu(x))_{x \in E}$ tulkituista funktioista $\mu \in \mathcal{D}_E$ käytetään jatkossa nimityksiä *jakauma* tai *todennäköisyysvektori*.²

Tämän tutkielman keskiössä ovat joukon E todennäköisyysjakaumien muodostamat jonot $(\mu_n)_{n \geq 1} \subset \mathcal{D}_E$. Niiden suppenemista tarkastellaan d_{TV} -metriikan avulla. Funktiojonon $(f_n) \subset \mathbb{R}^E$ suppeneminen kokonaisvaihteluetaisyden mielessä on vahva suppenemisen muoto. Esimerkiksi numeroituvan tila-avaruuden tapauksessa pisteittäinen suppeneminen ei vielä takaa suppenemista metriikan d_{TV} mielessä. Äärellisten tila-avaruuksien tilanteessa nämä suppenemisen muodot ovat kuitenkin ekvivalentteja.

LEMMA 1.3.1. *Olkoon $(f_n)_{n \geq 1}$ joukon \mathbb{R}^E jono, ja olkoon $f \in \mathbb{R}^E$. Tällöin seuraavat kaksi ehtoa ovat yhtäpitäviä:*

- (i) $d_{TV}(f_n, f) \xrightarrow{n \rightarrow \infty} 0$.
- (ii) $|f_n(x) - f(x)| \xrightarrow{n \rightarrow \infty} 0$ kaikilla $x \in E$.

TODISTUS. Todistus on triviaali joten se sivuutetaan. □

HUOMAUTUS 1.3.2. Lemman 1.3.1 ehdolle (i) käytetään jatkossa myös lyhennysmerkintää $f_n \xrightarrow{TV} f$.

Metrisen avaruuden (\mathcal{D}_E, d_{TV}) tärkeä ominaisuus on, että kaikki sen Cauchy-jonot suppenevat, kuten seuraava lemma osoittaa.

¹Avaruuden $(\mathbb{R}^E, \|\cdot\|)$ täydellisyys on suora seuraus normiavaruuden $(\mathbb{R}, \|\cdot\|)$ täydellisyydestä. Voidaan nimittäin helposti osoittaa, että äärellisen joukon E tapauksessa jono $(f_n)_n = ((f_n(x))_{x \in E})_n \subset \mathbb{R}^E$ on Cauchy-jono L^1 -normin mielessä jos ja vain jos marginaalijonot $(f_n(x))_n \subset \mathbb{R}$ ovat Cauchy-jonoja L^1 -mielessä kaikilla $x \in E$.

²Erityisesti Markovin ketjuja käsittelevässä kirjallisuudessa paljon käytetty nimitys *jakauma* on kieltämättä hieman hankala, sillä kuten todettua, samaa nimitystä käytetään satunnaismuuttujan X todennäköisyysmitasta \mathbb{P}_X , joka on joukkofunktio. Tutkielman kirjoittaja pyrkii pitämään tämän asian mielessä ja välttämään mahdollisia sekaannuksia.

LEMMA 1.3.3. \mathcal{D}_E on Banach-avaruuden $(\mathbb{R}^E, \|\cdot\|)$ suljettu osajoukko. Erityisesti siten (\mathcal{D}_E, d_{TV}) on täydellinen metrinen avaruus.

TODISTUS. Osoitetaan, että $\overline{\mathcal{D}_E} \subset \mathcal{D}_E$: olkoon $\nu \in \overline{\mathcal{D}_E}$. Tällöin on olemassa suppeneva jono $(\nu_n)_{n=1}^\infty \subset \mathcal{D}_E$ siten, että $\nu_n \xrightarrow{TV} \nu$, josta Lemman 1.3.1 nojalla seuraa, että $\nu(x) = \lim_{n \rightarrow \infty} \nu_n(x)$ kaikilla $x \in E$.

Koska $\nu_n(x) \in [0, 1]$ kaikilla $n \geq 1$ ja $x \in E$, ja koska $[0, 1]$ on suljettu reaaliakselin osajoukko, niin myös $\nu(x) \in [0, 1]$ eli $\nu(x) \geq 0$ kaikilla $x \in E$. Koska lisäksi

$$\sum_{x \in E} \nu(x) = \sum_{x \in E} \lim_{n \rightarrow \infty} \nu_n(x) = \lim_{n \rightarrow \infty} \sum_{x \in E} \nu_n(x) = \lim_{n \rightarrow \infty} 1 = 1,$$

niin $\nu \in \mathcal{D}_E$. On siten osoitettu, että $\overline{\mathcal{D}_E} \subset \mathcal{D}_E$, joten \mathcal{D}_E on suljettu.

Olkoon sitten $(\mu_n)_{n \geq 1}$ Cauchy-jono joukossa \mathcal{D}_E . Koska $(\mathbb{R}^E, \|\cdot\|)$ on täydellinen ja koska $\mathcal{D}_E \subset \mathbb{R}^E$, on olemassa raja-arvo $f \in \mathbb{R}^E$. Koska \mathcal{D}_E on suljettu, sisältää se suppenevien jonojensa raja-arvot [29, s. 79], joten $f \in \mathcal{D}_E$. Siten myös avaruus (\mathcal{D}_E, d_{TV}) on täydellinen. \square

HUOMAUTUS 1.3.4. On hyvä huomata, että koska \mathcal{D}_E ei ole vektoriavaruus, joten sen yhteydessä ei voida puhua normiavaruudesta.

Kootaan seuraavaksi jatkoon kannalta tarpeelliset kokonaisvaihteluetaisyyden ominaisuudet seuraavaan lauseeseen.³

LAUSE 1.3.5. Olkoon μ ja $\nu \in \mathcal{D}_E$. Tällöin

- (i) $\|\mu - \nu\| = 2 \sum_{x \in E} (\mu(x) - \nu(x))^+ = 2 \sum_{x \in E} (\mu(x) - \nu(x))^-$.
- (ii) $\|\mu - \nu\| \leq 2$, missä yhtäsuuruus on voimassa jos ja vain jos jakaumilla μ ja ν on erilliset kantajat, ts.
 $\{x \in E : \mu(x) > 0\} \cap \{x \in E : \nu(x) > 0\} = \emptyset$.
- (iii) $\|\mu - \nu\| = 2 \left(1 - \sum_x \mu(x) \wedge \nu(x)\right)$.
- (iv) $\|\mu - \nu\| = \max_{|h| \leq 1} \left| \sum_x h(x) (\mu(x) - \nu(x)) \right|$.

TODISTUS. Kohtien (i), (iii) ja (iv) todistukset löytyvät kirjasta [30, s. 81–82]. Todistetaan kohta (ii).

Todetaan, että kohdan (ii) epäyhtälö on aina voimassa, sillä itseisarvon kolmioepäyhtälön nojalla voidaan kirjoittaa kaikille $\mu, \nu \in \mathcal{D}_E$

$$\|\mu - \nu\| = \sum_{x \in E} |\mu(x) - \nu(x)| \leq \sum_{x \in E} |\mu(x)| + \sum_{x \in E} |\nu(x)| = 2.$$

Kiinnitetään sitten μ ja $\nu \in \mathcal{D}_E$. Määritellään jakaumien μ ja ν kantajat asettamalla $A := \{x \in E : \mu(x) > 0\}$ ja $B := \{x \in E : \nu(x) > 0\}$. Koska E voidaan esittää

³Määritellään reaalfunktion f positiivi- ja negatiiviosat asettamalla $f^+(x) := \max\{f(x), 0\}$ ja $f^-(x) := \max\{-f(x), 0\}$.

erillisten joukkojen $A \setminus B$, $B \setminus A$ ja $A \cap B$ yhdisteenä, soveltamalla kohtaa (i) voidaan kirjoittaa

$$\begin{aligned} \|\mu - \nu\| &= 2 \left(\sum_{x \in A \setminus B} (\mu(x) - \nu(x))^+ + \sum_{x \in B \setminus A} (\mu(x) - \nu(x))^+ + \sum_{x \in A \cap B} (\mu(x) - \nu(x))^+ \right) \\ (1.1) \quad &= 2 \left(\sum_{x \in A \setminus B} \mu(x) + \sum_{x \in A \cap B} (\mu(x) - \nu(x))^+ \right). \end{aligned}$$

Jos $A \cap B = \emptyset$, niin $A \setminus B = A$, ja lauseke (1.1) yksinkertaistuu muotoon

$$2 \left(\sum_{x \in A} \mu(x) + 0 \right) = 2 \sum_{x \in E} \mu(x) = 2.$$

Siten $\|\mu - \nu\| = 2$ jos $A \cap B = \emptyset$. Jos puolestaan $A \cap B \neq \emptyset$, niin $(\mu(x) - \nu(x))^+ < \mu(x)$ kaikilla $x \in A \cap B$, joten lauseketta 1.1 arvioimalla saadaan aito epäyhtälö;

$$2 \left(\sum_{x \in A \setminus B} \mu(x) + \sum_{x \in A \cap B} (\mu(x) - \nu(x))^+ \right) < 2 \left(\sum_{x \in A \setminus B} \mu(x) + \sum_{x \in A \cap B} \mu(x) \right) = 2.$$

On siten voimassa myös implikaatio: $A \cap B = \emptyset$, jos $\|\mu - \nu\| = 2$. \square

HUOMAUTUS 1.3.6. Toisinaan (esimerkiksi kirjoissa [15] ja [21]) kokonaisvaihteluetäisyys määritellään lisäämällä eteen kerroin $\frac{1}{2}$, ts. $d_{TV}(\mu, \nu) = \frac{1}{2} \|\mu - \nu\|$, mikä on ymmärrettävää edellisen lauseen kohdan (ii) valossa.

Myöhemmin todistettavan epähomogeenisten Markovin ketjujen suppenemislauseen (Lause 3.5.2) kannalta seuraavat kaksi tulosta ovat tärkeitä.

LAUSE 1.3.7. *Olkoon $(\mu_n)_{n \geq 1}$ jono avaruuden (\mathcal{D}_E, d_{TV}) jakaumia siten, että*

$$(1.2) \quad \sum_{n=1}^{\infty} \|\mu_n - \mu_{n+1}\| < \infty.$$

Tällöin $(\mu_n)_{n \geq 1}$ on Cauchy-jono, jolloin erityisesti on olemassa raja-arvo $\mu_\infty \in \mathcal{D}_E$ siten, että $\mu_n \xrightarrow{TV} \mu_\infty$.

TODISTUS. Olkoon $\varepsilon > 0$. Oletuksen (1.2) nojalla osasummien jono $(S_k)_{k \geq 1}$, missä $S_k = \sum_{n=1}^k \|\mu_{n+1} - \mu_n\|$, $k \geq 1$, suppenee johonkin lukuun $S \in \mathbb{R}$, joten on olemassa indeksi $k_0 = k_0(\varepsilon) \geq 1$ siten, että

$$S - S_{k-1} = \sum_{n=k}^{\infty} \|\mu_{n+1} - \mu_n\| < \varepsilon \text{ kaikilla } k > k_0,$$

jolloin metriikan d_{TV} kolmioepäyhtälöä soveltamalla saadaan

$$\|\mu_k - \mu_l\| \leq \sum_{n=l}^{k-1} \|\mu_{n+1} - \mu_n\| \leq \sum_{n=l}^{\infty} \|\mu_{n+1} - \mu_n\| < \varepsilon,$$

kun $k > l > k_0$. Tämä tarkoittaa, että $(\mu_n)_{n \geq 1}$ on Cauchy-jono, ja koska Lemman 1.3.3 nojalla \mathcal{D}_E on täydellinen, on olemassa raja-arvo $\mu_n \xrightarrow{TV} \mu_\infty \in \mathcal{D}_E$. \square

HUOMAUTUS 1.3.8. Kaikki Cauchy-jonot eivät toteuta summautumisehtoa (1.2). Vastaesimerkiksi kelpaa mm. Cauchy-jono $(\nu_n)_{n \geq 1} \subset \mathcal{D}_E$, jolle $\|\nu_n - \nu_{n+1}\| = n^{-1}$ kaikilla $n \geq 1$. Tullaan huomaamaan, että Lauseen 3.5.2 todistuksessa käsiteltävälle rajajakaumien jonolle pelkkä Cauchyn ehdon toteutuminen ei riitä, vaan suppenemisnopeuden on oltava vähintään ehdon (1.2) mukaista. Seuraava Lemma 1.3.9 antaa hyödyllisen tavan näyttää ehto (1.2) toteen annetulle jakaumajonolle $(\mu_n) \subset \mathcal{D}_E$.

LEMMA 1.3.9. *Olkoon $(\mu_n)_{n \geq 1} \subset \mathcal{D}_E$. Jos kaikilla $x \in E$ kuvaus $n \mapsto \mu_n(x)$ on monotoninen jostain lähtien, niin jono $(\mu_n)_{n \geq 1}$ toteuttaa ehdon (1.2).*

TODISTUS. Oletuksen nojalla jokaiselle $x \in E$ on olemassa $n_x \geq 1$ siten, että joko $\mu_n(x) \leq \mu_{n+1}(x)$ tai $\mu_n(x) \geq \mu_{n+1}(x)$ kaikilla $n \geq n_x$. Koska

$$(\mu_n(x) - \mu_{n+1}(x))^+ = \begin{cases} 0, & \text{jos } \mu_n(x) \leq \mu_{n+1}(x) \\ \mu_n(x) - \mu_{n+1}(x), & \text{jos } \mu_n(x) > \mu_{n+1}(x) \end{cases},$$

niin jokaiselle $x \in E$

$$\sum_{n=n_x}^m (\mu_n(x) - \mu_{n+1}(x))^+ \leq 1, \quad \text{kun } m > n_x.$$

Erityisesti siten myös

$$\max_{x \in E} \sum_{n=N}^m (\mu_n(x) - \mu_{n+1}(x))^+ \leq 1, \quad \text{kun } m > N := \max_{x \in E} n_x.$$

Siispä kaikilla $m > N$ on voimassa

$$\begin{aligned} S_m &:= \sum_{n=1}^m \|\mu_n - \mu_{n+1}\| = 2 \sum_{x \in E} \sum_{n=1}^m (\mu_n(x) - \mu_{n+1}(x))^+ \\ &\leq 2 \sum_{x \in E} \left(\sum_{n=1}^{N-1} (\mu_n(x) - \mu_{n+1}(x))^+ + 1 \right) \\ &\leq 2|E|(N-1+1) \leq 2|E|N, \end{aligned}$$

joten osasummien jono $(S_n)_{n \geq 1}$ on rajoitettu. Koska se on myös kasvava, on olemassa raja-arvo $S := \lim_{n \rightarrow \infty} S_n < \infty$. Tämä todistaa väitteen. \square

1.4. Stokastisten matriisien avaruus

MÄÄRITELMÄ 1.4.1 (Stokastinen matriisi). Reaaliarvoinen neliömatriisi $P = [a_{ij}]_{n \times n}$ on *stokastinen matriisi*, mikäli

- (i) $a_{ij} \geq 0$ kaikilla $i, j = 1, 2, \dots, n$
- (ii) $\sum_{j=1}^n a_{ij} = 1$ kaikilla $i = 1, 2, \dots, n$.

Olkoon $E \neq \emptyset$ jälleen äärellinen tila-avaruus. Koska joukon E alkioita voivat olla lukuja, vektoreita, matriiseja, funktioita ja niin edelleen, ei tiloja $x \in E$ ole useinkaan tarpeellista numeroida. Markovin ketjujen teoriassa $|E| \times |E|$ -matriisin P alkioon

viitattaessa onkin tavanomaista kirjoittaa yksinkertaisesti $P(x, y)$, missä tila $x \in E$ vastaa riviä ja tila $y \in E$ saraketta. Olkoon siis

$$\mathcal{M}_{E \times E} := \{P \mid P \text{ on stokastinen } |E| \times |E| \text{-matriisi}\}.$$

Olkoot $\mu \in \mathcal{D}_E$, $P, Q \in \mathcal{M}_{E \times E}$. Käytetään vastedes lyhennysmerkintöjä

$$(i) \quad \mu P(y) := (\mu P)(y) = \sum_{x \in E} \mu(x) P(x, y)$$

$$(ii) \quad PQ(x, y) := (PQ)(x, y) = \sum_{z \in E} P(x, z) Q(z, y).$$

Kohta (i) vastaa tilannetta, jossa jakauma μ $1 \times |E|$ -matriisiksi eli $|E|$ -ulotteiseksi rivivektoriksi tulkittuna kerrotaan $|E| \times |E|$ -matriisilla, jolloin tulona saadaan $1 \times |E|$ -matriisi eli $|E|$ -ulotteinen rivivektori μP . Siisteyden vuoksi pidättäydytään transpoo-simerkitöjen käytöstä, sillä asiayhteydestä on aina mahdollista päätellä, onko kyseessä rivi- vai sarakevektori.

Kohta (ii) puolestaan vastaa tavanomaista kahden $|E| \times |E|$ -matriisin matriisitulua. Ei ole vaikeaa todeta, että $\mu P \in \mathcal{D}_E$ ja $PQ \in \mathcal{M}_{E \times E}$. Siten myös yleisemmin n matriisille $P_1, P_2, \dots, P_n \in \mathcal{M}_{E \times E}$ tulo $\mu P_1 \cdots P_n \in \mathcal{D}_E$ on joukon \mathcal{D}_E jakauma ja $P_1 \cdots P_n$ joukon $\mathcal{M}_{E \times E}$ stokastinen matriisi.

LAUSE 1.4.2. *Olkoon $\mu \in \mathcal{D}_E$ ja $P_1, P_2, \dots, P_n \in \mathcal{M}_{E \times E}$. Tällöin*

$$(a) \quad \mu P_1 \cdots P_n(y) = \sum_{z_0, \dots, z_{n-1} \in E} \mu(z_0) P_1(z_0, z_1) \cdots P_n(z_{n-1}, y)$$

$$(b) \quad P_1 \cdots P_n(x, y) = \sum_{z_0, \dots, z_{n-1} \in E} P_1(x, z_1) P_2(z_1, z_2) \cdots P_n(z_{n-1}, y).$$

$$(c) \quad \mu P_1 \cdots P_n \in \mathcal{D}_E \quad \text{ja} \quad P_1 \cdots P_n \in \mathcal{M}_{E \times E}.$$

TODISTUS. Todistus on yksinkertainen induktiotodistus, joten se sivuutetaan. \square

Tehdään vielä tämän luvun päätteeksi muutama stokastisiin matriiseihin liittyvä huomio Luvussa 3 tarkasteltavan heikon ergodisuuden myöhempää havainnollistamista silmällä pitäen.

MÄÄRITELMÄ 1.4.3 (Vakiorivimatriisi). Sanotaan, että $P \in \mathcal{M}_{E \times E}$ on *vakiorivimatriisi*, mikäli matriisin P rivit ovat identtiset, toisin sanoen $P(x, \cdot) = p$ kaikilla $x \in E$ jollekin jakaumalle $p \in \mathcal{D}_E$.

LAUSE 1.4.4. *Olkoon $P \in \mathcal{M}_{E \times E}$ vakiorivimatriisi, jolle $P(x, \cdot) = p \in \mathcal{D}_E$ kaikilla $x \in E$. Tällöin kaikille jakaumille $\mu \in \mathcal{D}_E$ ja stokastisille matriiseille $Q \in \mathcal{M}_{E \times E}$*

(i) $\mu P = p$, (ii) $QP = P$ ja (iii) PQ on vakiorivimatriisi.

TODISTUS. Kohta (i):

$$\mu P(y) = \sum_{z \in E} \mu(z) P(z, y) = \sum_{z \in E} \mu(z) p(y) = p(y) \quad \text{kaikilla } y \in E.$$

Kohta (ii):

$$QP(x, y) = \sum_{z \in E} Q(x, z)P(z, y) = \sum_{z \in E} Q(x, z)p(y) = p(y) = P(x, y) \text{ kaikilla } x, y \in E.$$

Kohta (iii): olkoon $q : E \rightarrow \mathbb{R}$; $q(y) := \sum_{z \in E} p(z)Q(z, y) = pQ(y)$, jolloin Lauseen 1.4.2 kohdan (c) nojalla pätee $q \in \mathcal{D}_E$. Koska lisäksi

$$PQ(x, y) = \sum_{z \in E} P(x, z)Q(z, y) = \sum_{z \in E} p(z)Q(z, y) = q(y) \text{ kaikilla } x, y \in E,$$

väite on todistettu. □

LUKU 2

Satunnaiskenttien alkeita

Tässä luvussa muotoillaan satunnaiskentän käsite sekä sen tärkeinä erikoistapauksina Markovin satunnaiskenttä ja Gibbsin kenttä. Todistetaan lisäksi muutamia hyödyllisiä tuloksia myöhempiä lukuja – erityisesti Gibbs-otantaa ja Gibbs-otantaan perustuvaa simuloitua jäähtytystä koskevia Lukuja 4 ja 5 – varten.

Olkoon S niin sanottu *solmujen joukko*, jolle $|S| < \infty$. Lähtökohtaisesti joukolta S ei vaadita minkäänlaista erityisrakennetta, vaan se toimii ainoastaan eräänlaisena järjestämättömänä indeksijoukkona. Satunnaiskenttien sovelluksissa S on usein jokin avaruuden \mathbb{R}^2 osajoukko, esimerkiksi äärellinen *neliöhila*

$$\{(i, j) \in \mathbb{Z}^2 : i, j = 1, \dots, m\} \text{ jollekin } m \in \mathbb{N}.$$

Olkoon sitten jokaista solmua $s \in S$ kohden annettu äärellinen *faasiavaruus* $\Lambda_s \subset \mathbb{R}$. Tuloavaruutta

$$\prod_{s \in S} \Lambda_s = \left\{ x : S \rightarrow \bigcup_{s \in S} \Lambda_s \mid x(s) \in \Lambda_s \text{ kaikilla } s \in S \right\}$$

kutstutaan tällöin *konfiguraatioavaruudeksi*, jonka alkiot, *konfiguraatiot*, ovat siis kuvauksia, joissa jokaiseen solmuun $s \in S$ liittyy jokin *faasi* $\lambda \in \Lambda_s$.

Otetaan seuraavaksi käyttöön merkinnät konfiguraatioiden rajoittumille. Määritellään konfiguraation $x \in \prod_{s \in S} \Lambda_s$ *rajoittuma* osajoukkoon $\emptyset \neq A \subset S$ yhdistettynä kuvauksena

$$x_A := p_A \circ x : S \rightarrow \prod_{s \in A} \Lambda_s,$$

missä p_A on *projektiokuvaus*, joka kuvaa tuloavaruuden $\prod_{s \in S} \Lambda_s$ alkion $x = (x(s))_{s \in S}$ tuloavaruudelle $\prod_{s \in A} \Lambda_s$ siten, että $p_A(x) = (x(s))_{s \in A}$. Merkitään jatkossa kaikille yksiöille $\{s\} \subset S$ $x_s := x_{\{s\}} = x(s)$, jolloin voidaan myös merkitä konfiguraatiota hieman helpommassa muodossa $x = (x_s)_{s \in S}$. Merkitään edelleen $x = (x_A, x_{S \setminus A})$ epätyhjille osajoukoille $A \subset S$. Konfiguraatioavaruutta merkitään tästä lähtien (poislukien Luvut 6 ja 7) aina kirjaimella E , jolloin on perusteltua käyttää myös lyhennysmerkintää $E_A = \prod_{s \in A} \Lambda_s$, kun $A \subset S$.

2.1. Satunnaiskentän jakaumat, Markovin satunnaiskenttä

MÄÄRITELMÄ 2.1.1 (Satunnaiskenttä). Todennäköisyysavaruuden (Ω, \mathbb{P}) satunnaisuuttujista $X_s : \Omega \rightarrow \Lambda_s, s \in S$, koostuva perhe $X = (X_s)_{s \in S}$ on *satunnaiskenttä*, mikäli

$$\mathbb{P}(X = x) = \mathbb{P}(X_s = x_s \forall s \in S) > 0 \quad \text{kaikilla } x = (x_s)_{s \in S} \in E = \prod_{s \in S} \Lambda_s.$$

ESIMERKKI 2.1.2. Olkoon Määritelmän 2.1.1 tilanteessa solmujen joukkona $S = \{e_i : i = 1, \dots, n\}$, missä e_i on euklidisen avaruuden \mathbb{R}^n i :s kantavektori. Olkoon lisäksi $\Lambda = \{\lambda_1, \dots, \lambda_L\} \subset \mathbb{R}$ kaikille solmuille $e_i \in S$ yhteinen faasiavaruus. Tällöin satunnaiskenttä $(X_s)_{s \in S}$ voidaan tulkita todennäköisyysavaruuden (Ω, \mathbb{P}) *satunnaisvektorina* $X : \Omega \rightarrow \mathbb{R}^n$; $X(\omega) = (X_1(\omega), \dots, X_n(\omega))$ tekemällä samaistukset

$$(X_s)_{s \in S} = (X_{e_i})_{i=1}^n \approx \sum_{i=1}^n e_i \cdot X_i = (X_1, \dots, X_n) \quad \text{ja} \quad \Lambda^S \approx \Lambda^n \subset \mathbb{R}^n.$$

2.2. Satunnaiskentän marginaalijakaumat ja ehdolliset jakaumat

Tutkitaan seuraavaksi satunnaiskentän jakaumaa ja otetaan käyttöön jatkon kannalta tarvittavia lyhennysmerkintöjä. Olkoon satunnaiskenttä X kuten Määritelmässä 2.1.1 joillekin äärellisille S ja E . Satunnaiskentän X jakauma \mathbb{P}_X on satunnaisuuttujakokoelman $(X_s)_{s \in S}$ *yhteisjakauma*:

$$\mathbb{P}_X : \mathcal{P}(E) \rightarrow \mathbb{R}; \quad \mathbb{P}_X(A) := \mathbb{P}(X \in A) = \sum_{x \in A} \mathbb{P}(X = x),$$

missä kullekin konfiguraatiolle $x = (x_s)_{s \in S} \in E$

$$\mathbb{P}(X = x) = \mathbb{P}(X_s = x_s, s \in S) = \mathbb{P}\left(\bigcap_{s \in S} \{\omega \in \Omega : X_s(\omega) = x_s\}\right).$$

Tavallisesti satunnaiskentän X jakauma \mathbb{P}_X määräytyy vasta jälkikäteen jonkin pistetodennäköisyysfunktion $\pi \in \mathcal{D}_E$ indusoimana:

$$\mathbb{P}_X(A) := \sum_{x \in A} \pi(x) \quad \text{kaikilla } A \in \mathcal{P}(E).$$

Myös tätä funktiota π kutsutaan jatkossa satunnaiskentän X *jakaumaksi*. Määritellään sitten osakokoelmaa $(X_s)_{s \in A}$ vastaava satunnaiskentän X *marginaalijakauma* $\pi_A : E_A \rightarrow \mathbb{R}$ tavalliseen tapaan asettamalla

$$\pi_A(x_A) := \mathbb{P}(X_A = x_A) = \sum_{x_{S \setminus A} \in E_{S \setminus A}} \mathbb{P}(X = (x_A, x_{S \setminus A})) = \sum_{x_{S \setminus A} \in E_{S \setminus A}} \pi(x_A, x_{S \setminus A})$$

kaikilla $x_A \in E_A$. Erityisesti siten yhden satunnaisuuttujan X_s , $s \in S$, marginaalijakauman π_s todennäköisyydet voidaan esittää muodossa

$$\pi_s(x_s) = \sum_{x_{S \setminus \{s\}}} \pi(x_s, x_{S \setminus \{s\}}), \quad \text{missä summa käy yli joukon } E_{S \setminus \{s\}}.$$

Olkoot sitten $A, B \subset S$ erillisiä joukon S epätyhjiä osajoukkoja. Kullekin kiinnitetylle konfiguraatorajoittumalle $x_B \in E_B$ voidaan tällöin määritellä *ehdollinen jakauma* $\pi(\cdot | x_B) : E_A \rightarrow \mathbb{R}$ asettamalla

$$\begin{aligned} \pi(x_A | x_B) &:= \mathbb{P}(X_A = x_A | X_B = x_B) = \frac{\mathbb{P}(X_A = x_A, X_B = x_B)}{\mathbb{P}(X_B = x_B)} \\ (2.1) \quad &= \frac{\sum_{x_{S \setminus (A \cup B)}} \mathbb{P}(X = (x_{A \cup B}, x_{S \setminus (A \cup B)}))}{\sum_{x_{S \setminus B}} \mathbb{P}(X = (x_B, x_{S \setminus B}))} \\ &= \frac{\sum_{x_{S \setminus (A \cup B)}} \pi(x_{A \cup B}, x_{S \setminus (A \cup B)})}{\sum_{x_{S \setminus B}} \pi(x_B, x_{S \setminus B})}, \end{aligned}$$

missä yhtälöiden (2.1) summat otetaan yli rajoittumajoukkojen $E_{S \setminus (A \cup B)}$ ja $E_{S \setminus B}$. Erityisen tarkastelun kohteeksi nousevat myöhemmin niin kutstut *täydet ehdolliset jakaumat* $\pi(\cdot \mid x_{S \setminus \{s\}})$, joiden todennäköisyyksille saadaan yhtälöitä (2.1) käyttämällä saadaan siisti esitys

$$(2.2) \quad \pi(x_s \mid x_{S \setminus \{s\}}) = \frac{\pi(x)}{\sum_{\lambda \in \Lambda_s} \pi(\lambda, x_{S \setminus \{s\}})}.$$

HUOMAUTUS 2.2.1. Määritelmään 2.1.1 sisällytetty *positiivisuusehto*

$$(2.3) \quad \mathbb{P}(X = x) > 0 \text{ kaikilla } x \in E$$

takaa satunnaiskentälle hyviä ominaisuuksia, minkä vuoksi se toisinaan otetaan satunnaiskentän määritelmään mukaan. Erityisesti se takaa, että kaikki satunnaiskentän X marginaalijakaumat ovat aidosti positiivisia, sillä kaikille $\emptyset \neq A \subset S$ ja $x_A \in E_A$ on tällöin voimassa

$$\mathbb{P}(X_A = x_A) \geq \max_{\substack{x \in E \\ x_A = x_A}} \mathbb{P}(X = x) > 0.$$

Tämä puolestaan implikoi, että kaikki satunnaiskentän X ehdolliset jakaumat voidaan määritellä ja lisäksi myös ne ovat aidosti positiivisia.

Tarkastellaan seuraavaksi suuntaamatonta verkkoa $\mathcal{G} = (S, \mathcal{E})$, jossa joukko \mathcal{E} sisältää informaation verkon linkeistä. Toisin sanoen $e = \{s, t\} \in \mathcal{E}$, jos solmujen s ja t välillä on linkki. Toisinaan on helpompi tarkastella suuntaamattomia verkkoja (S, \mathcal{N}) , joissa linkkien joukko on korvattu *naapurustojärjestelmällä* \mathcal{N} .

MÄÄRITELMÄ 2.2.2 (Naapurustojärjestelmä). Kokoelma $\mathcal{N} = (\mathcal{N}_s)_{s \in S}$ solmujen joukon S osajoukkoja \mathcal{N}_s on *naapurustojärjestelmä*, mikäli

- (i) $s \notin \mathcal{N}_s$ kaikilla $s \in S$
- (ii) jos $s \in \mathcal{N}_t$, niin $t \in \mathcal{N}_s$ kaikilla $s, t \in S$.

Sanotaan, että solmut s ja t ovat naapureita, ja merkitään $s \sim t$, jos $s \in \mathcal{N}_t$. Joukkoa \mathcal{N}_s kutsutaan solmun s *naapurustoksi*.

Voidaan helposti osoittaa, että linkkien joukko indusoi verkkoon naapurustojärjestelmän ja päinvastoin. Jatkossa puhutaankin suuntaamattomasta verkosta aina parina (S, \mathcal{N}) .

MÄÄRITELMÄ 2.2.3 (Markovin satunnaiskenttä). E -arvoinen satunnaiskenttä on *Markovin satunnaiskenttä* (tai *Markovin kenttä*) *naapurustojärjestelmän* $\mathcal{N} = (\mathcal{N}_s)_{s \in S}$ *suhteen*, mikäli ehto

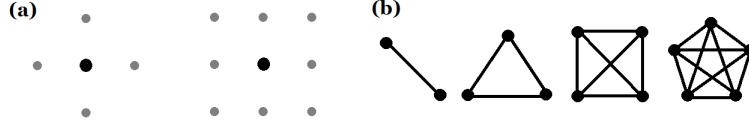
$$(2.4) \quad \mathbb{P}(X_s = x_s \mid X_{S \setminus \{s\}} = x_{S \setminus \{s\}}) = \mathbb{P}(X_s = x_s \mid X_{\mathcal{N}_s} = x_{\mathcal{N}_s}),$$

on voimassa kaikilla $x \in E$ ja $s \in S$.

HUOMAUTUS 2.2.4. Ehtoa 2.4 kutsutaan *lokaaliksi Markov-ominaisuudeksi*. Se tarkoittaa, että kaikille $s \in S$ satunnaiskentän komponentti X_s on *ehdollisesti riippumaton sulkeuman* $\mathcal{N}_s \cup \{s\}$ ulkopuolisista komponenteista $\{X_t : t \notin \mathcal{N}_s \cup \{s\}\}$, kun muuttujien $\{X_t : t \in \mathcal{N}_s\}$ arvot tunnetaan. Satunnaiskentille voidaan määritellä myös useita muita Markov-ominaisuuksia. Niiden välisistä riippuvuussuhteista löytyy lisää informaatiota mm. kirjasta [30, s. 71].

HUOMAUTUS 2.2.5. Voidaan osoittaa, että Markovin satunnaiskentän jakauma π voidaan määrätä yksikäsitteisesti mikäli täydet ehdolliset jakaumat $\pi(\cdot \mid x_{S \setminus \{s\}})$, jotka yksinkertaistuvat niin kutsutuiksi *lokaaleiksi tunnuksiksi* $\pi(\cdot \mid x_{\mathcal{N}_s})$, tunnetaan (ks. [3, s. 255–256]). Lokaalien tunnusten kokoelmaa kutsutaankin satunnaiskentän *lokaaliksi spesifikaatioksi*.

2.3. Gibbsin kentät



KUVA 2.1. Kuva (a): neljän ja kahdeksan alkion naapurustot kaksiulotteisella hilalla. Harmaat pallot edustavat mustaa palloa vastaavaa solmun naapureita. Kuva (b): esimerkkikuvat 2-5 alkion klikeistä.

MÄÄRITELMÄ 2.3.1 (Klikki). Olkoon (S, \mathcal{N}) suuntaamaton verkko. Osajoukko $C \subset S$ on *klikki* jos kaikki joukon C eri alkiot ovat keskenään naapureita. Erityisesti siten tyhjä joukko \emptyset ja yksiöt ovat klikkejä. Naapurustojärjestelmän \mathcal{N} indusoimalle klikkikokoelmalle käytetään jatkossa merkintää $\mathcal{C}_{\mathcal{N}}$.

Kuvassa 2.1 on esitelty yksinkertaisimpia kaksiulotteisia naapurustoja ja klikkejä.

MÄÄRITELMÄ 2.3.2 (Potentiaali, naapuripotentiaali, paripotentiaali). Kokoelma $\mathcal{V} = (V_A)_{A \subset S} = \{V_A : E \rightarrow \mathbb{R} \mid A \subset S\}$ on *potentiaali*, mikäli

- (i) $V_{\emptyset} = 0$ ja
- (ii) jos $x_A = y_A$, niin $V_A(x) = V_A(y)$.

Potentiaali $\mathcal{V} = \mathcal{V}_{\mathcal{N}}$ on *naapuripotentiaali* naapurustojärjestelmän \mathcal{N} suhteen, mikäli $V_A = 0$ aina kun $A \notin \mathcal{C}_{\mathcal{N}}$. Lisäksi sanotaan, että potentiaali \mathcal{V} on *paripotentiaali*, mikäli $V_A = 0$ kaikille $A \subset S, |A| > 2$.

Kuvausta $U : E \rightarrow \mathbb{R}, U(x) = \sum_{A \subset S} V_A(x)$ kutsutaan potentiaalin $\mathcal{V} = (V_A)_{A \subset S}$ määrittämiseksi *energiafunktioksi* tai *energiaksi*.

HUOMAUTUS 2.3.3. Määritelmän 2.3.2 ehdosta (ii) seuraa, että funktiot V_A eivät saa lisäinformaatiota konfiguraatioiden $x \in E$ osajoukon $A \subset S$ ulkopuolisista arvoista $(x_s)_{s \in S \setminus A}$.

HUOMAUTUS 2.3.4. Naapuripotentiaali $\mathcal{V}_{\mathcal{N}}$ voidaan kirjoittaa muodossa $(V_C)_{C \in \mathcal{C}_{\mathcal{N}}}$, kun jätetään pois turhat nollakuvaukset $V_A, A \notin \mathcal{C}_{\mathcal{N}}$. Lisäksi naapuripotentiaalin $\mathcal{V}_{\mathcal{N}}$ energiefunktio saadaan, kun summataan ainoastaan naapurustojärjestelmän \mathcal{N} määrittämien klikkien $C \in \mathcal{C}_{\mathcal{N}}$ yli; $U(x) = \sum_{C \in \mathcal{C}_{\mathcal{N}}} V_C(x)$ kaikille $x \in E$.

ESIMERKKI 2.3.5 (Ising-malli). Olkoon $S = \{(i, j) \in \mathbb{Z}^2 : 1 \leq i, j \leq m\}$ jollakin $m \geq 2$, ja olkoon $\Lambda = \{-1, 1\}$ solmuille yhteinen faasiavaruus. Määritellään kunkin solmun $s = (s_1, s_2) \in S$ naapurusto asettamalla

$$\mathcal{N}_s := \{t = (t_1, t_2) \in S : 0 < \|(s_1, s_2) - (t_1, t_2)\|_{L^2} \leq 1\}, \text{ missä}$$

$$\|(s_1, s_2)\|_{L^2} = \left(\sum_{i=1}^2 s_i^2 \right)^{1/2} \text{ on } L^2\text{-normi joukolla } S.$$

Naapurustojärjestelmän $\mathcal{N} = (\mathcal{N}_s)_{s \in S}$ määräämä kokoelma $\mathcal{C}_{\mathcal{N}}$ sisältää siten (tyhjän joukon lisäksi) kolmen tyyppisiä klikkejä: yksiöitä sekä vierekkäisistä tai päällekkäisistä solmuista koostuvia kaksioita. Tarkastellaan naapuripotentialiaa $\mathcal{V} = (V_C)_{C \in \mathcal{C}_{\mathcal{N}}}$, jonka potentiaalifunktiot ovat muotoa

$$V_{\{s,t\}} = -\frac{J}{k_B T} \sum_{\{s,t\} \in \mathcal{C}} x_s x_t \quad \text{ja} \quad V_{\{s\}} = -\frac{mB}{k_B T} \sum_{\{s\} \in \mathcal{C}} x_s.$$

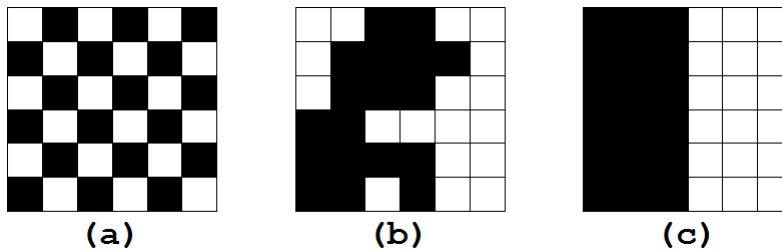
Niin sanotun 2-ulotteisen *Ising-mallin* (E. Ising, 1925) energiafunktio voidaan esittää tällöin muodossa

$$(2.5) \quad U(x) = -\frac{J}{k_B T} \sum_{\{s,t\} \in \mathcal{C}} x_s x_t - \frac{mB}{k_B T} \sum_{\{s\} \in \mathcal{C}} x_s,$$

missä J ja m ovat materiaaliin liittyviä vakioita, k_B on Boltzmannin vakio, T on absoluuttinen lämpötila ja B on ulkoisen magneettikentän intensiteetti. Jos kyseessä on ferromagneetti, $J > 0$ (ks. Kuva 2.2), kun taas antiferromagneetille $J < 0$.

Mallissa arvot $x_s \in \{-1, 1\}$ vastaavat hilapisteeseen s liittyvää fysikaalista suuretta *spiniä*. Yhtälön (2.5) oikean puolen ensimmäinen termi vastaa spin-parien interaktioenergiaa ja vasen puoli edustaa ulkoisen kentän magneettikentän energiakontribuutiota. Jos T on pieni (tai J suuri), spinit pyrkivät järjestäytymään samansuuntaisesti. Suuri T (tai pieni J) puolestaan vastaa heikkoa interaktiota spinien välillä. G. Winkler kirjoittaa Ising-mallista seuraavasti:

“The *Ising model* seems to be simple at a first glance. But it exhibits a variety of fundamental and typical phenomena shared by many large complex systems. Hence it is *the* test model for substantial questions about Markov fields.” [30, s. 62]



KUVA 2.2. Ising-malli 6×6 -hilalla. Parittaisia vaihtoja tekevällä Metropolis-jäähdytyksellä realisoituja konfiguraatioita mallille, jossa $\frac{J}{k_B T} = 1$ ja $\frac{mB}{k_B T} = 0$. Kuva (a) on alkukonfiguraatio, $U = 10$; (b) $n = 50$, $U = -28$; (c) $n = 650$, $U = -78$. Valkoinen väri edustaa negatiivista ja musta positiivista spiniä.

MÄÄRITELMÄ 2.3.6 (Gibbsin kenttä). Olkoon (S, \mathcal{N}) suuntaamaton verkko. E -arvoinen satunnaiskenttä X verkolla (S, \mathcal{N}) on *Gibbsin kenttä potentiaalın \mathcal{V} suhteen*, jos

$$\mathbb{P}(X = x) = \pi(x) = \frac{1}{Z} \exp(-U(x)) = \frac{\exp(-U(x))}{\sum_{y \in E} \exp(-U(y))}$$

kaikilla $x \in E$, missä U on potentiaalista \mathcal{V} johdettu energiafunktio.

HUOMAUTUS 2.3.7. Gibbsin kenttään liittyvä potentiaali \mathcal{V} ei ole yksikäsitteinen. Jokaiselle Gibbsin kentälle on kuitenkin olemassa yksikäsitteinen *normalisoitu* potentiaali, niin sanottu *tyhjiöpotentiaali*. Asiaa käsitellään tarkemmin muun muassa kirjassa [30, s. 66–72].

Määritelmässä 2.3.6 esiintyvää jakaumaa π kutstutaan *Gibbsin* tai *Boltzmannin jakaumaksi*, jonka normalisoivaa vakiota $Z = \sum_{y \in E} \exp(-U(y))$ kutsutaan *partitiiofunktiksi*. Konfiguraatiota $x \in E$ vastaavaa painokerrointa $\exp(-U(x))$ kutsutaan puolestaan nimellä *Boltzmannin tekijä*. Gibbsin jakauman ominaisuuksia tutkitaan tarkemmin Luvuissa 4 ja 5.

HUOMAUTUS 2.3.8. Naapuripotentiaaliin $\mathcal{V}_{\mathcal{N}}$ liittyvän Gibbsin kentän jakauman π tärkeä ominaisuus on, että se *faktorisoituu* yli klikkien joukon $\mathcal{C}_{\mathcal{N}}$. Toisin sanoen se voidaan esittää muodossa

$$\pi(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}_{\mathcal{N}}} \phi_C(x),$$

missä kukin epänegatiivinen funktio $\phi_C : E \rightarrow \mathbb{R}$ riippuu konfiguraatioiden arvoista ainoastaan joukolla C . Voidaan nimittäin asettaa $\phi_C(x) := \exp(-V_C(x))$ kaikilla $C \in \mathcal{C}_{\mathcal{N}}$.

Todistetaan seuraavaksi tärkeä tulos koskien Gibbsin jakauman täysien ehdollisten jakaumien laskemista.

LAUSE 2.3.9. *Olkoon (S, \mathcal{N}) suuntaamaton verkko ja olkoon X Gibbsin kenttä naapuripotentiaalın $(V_C)_{C \in \mathcal{C}_{\mathcal{N}}}$ suhteen arvojoukkonaan $E = \prod_{s \in S} \Lambda_s$. Tällöin kentän X jakauman π lokaaleille tunnuksille on voimassa esitys*

$$(2.6) \quad \pi(x_s \mid x_{S \setminus \{s\}}) = \frac{\exp(-\sum_{C \in \mathcal{A}_s} V_C(x))}{\sum_{\lambda_s \in \Lambda_s} \exp(-\sum_{C \in \mathcal{A}_s} V_C(\lambda_s, x_{S \setminus \{s\}}))},$$

missä $\mathcal{A}_s = \{C \in \mathcal{C}_{\mathcal{N}} : s \in C\}$ on solmun $s \in S$ sisältävien klikkien kokoelma.

TODISTUS. Olkoot $s \in S$, $\mathcal{A}_s = \{C \in \mathcal{C}_{\mathcal{N}} : s \in C\}$ ja $\mathcal{B}_s = \{C \in \mathcal{C}_{\mathcal{N}} : s \notin C\}$, jolloin $\mathcal{A}_s \cap \mathcal{B}_s = \emptyset$ ja $\mathcal{C}_{\mathcal{N}} = \mathcal{A}_s \cup \mathcal{B}_s$. Ositetaan siis klikkien kokoelma erillisiin osakokoelmiin, joista vain toisen osakokoelman klikit sisältävät solmun $s \in S$. Tällöin $U(x) = \sum_{C \in \mathcal{A}_s} V_C(x) + \sum_{C \in \mathcal{B}_s} V_C(x)$ kaikilla $x \in E$. Lisäksi $V_C(x) = V_C(\lambda_s, x_{S \setminus \{s\}})$ kaikilla $\lambda_s \in \Lambda_s$, kun $C \in \mathcal{B}_s$, sillä konfiguraatiot x ja $(\lambda_s, x_{S \setminus \{s\}})$ yhtenevät solmun $s \notin C$ ulkopuolella.

Huomautuksen 2.2 nojalla kaikille $x \in E$ voidaan nyt kirjoittaa

$$\begin{aligned} \pi(x_s \mid x_{S \setminus \{s\}}) &= \frac{\pi(x)}{\sum_{\lambda_s \in \Lambda_s} \pi(\lambda_s, x_{S \setminus \{s\}})} = \frac{Z}{Z} \cdot \frac{\exp(-U(x))}{\sum_{\lambda_s \in \Lambda_s} \exp(-U(\lambda_s, x_{S \setminus \{s\}}))} \\ &= \frac{\exp(-\sum_{C \in \mathcal{A}_s} V_C(x) - \sum_{C \in \mathcal{B}_s} V_C(x))}{\sum_{\lambda_s \in \Lambda_s} \exp(-\sum_{C \in \mathcal{A}_s} V_C(\lambda_s, x_{S \setminus \{s\}}) - \sum_{C \in \mathcal{B}_s} V_C(\lambda_s, x_{S \setminus \{s\}}))} \\ &= \frac{\exp(-\sum_{C \in \mathcal{B}_s} V_C(x)) \exp(-\sum_{C \in \mathcal{A}_s} V_C(x))}{\exp(-\sum_{C \in \mathcal{B}_s} V_C(x)) \sum_{\lambda_s \in \Lambda_s} \exp(\sum_{C \in \mathcal{A}_s} V_C(\lambda_s, x_{S \setminus \{s\}}))} \\ &= \frac{\exp(-\sum_{C \in \mathcal{A}_s} V_C(x))}{\sum_{\lambda_s \in \Lambda_s} \exp(-\sum_{C \in \mathcal{A}_s} V_C(\lambda_s, x_{S \setminus \{s\}}))}. \end{aligned}$$

□

SEURAUUS 2.3.10. *Gibbsin kenttä naapuripotentialin \mathcal{V}_N suhteen on Markovin satunnaiskenttä naapurustojärjestelmän \mathcal{N} suhteen.*

TODISTUS. On näytettävä, että naapuripotentialiin \mathcal{V}_N liittyvän Gibbsin kentän X jakauman π täydet ehdolliset jakaumat yhtenevät lokaalien tunnuksien kanssa, toisin sanoen lokaali Markovin ehto (2.4) on voimassa.

Yhtälön (2.6) oikean puolen osamäärälausekkeessa esiintyy ainoastaan klikkejä, jotka kuuluvat kokoelmaan $\mathcal{A}_s = \{C \in \mathcal{C}_N : s \in C\}$. Koska vastaavat potentiaalifunktiot $(V_C)_{C \in \mathcal{A}_s}$ puolestaan riippuvat konfiguraatioiden arvoista ainoastaan kokoelman \mathcal{A}_s klikkeihin kuuluvissa solmuissa, riittää osoittaa, että

$$\bigcup_{C \in \mathcal{A}_s} C = \mathcal{N}_s \cup \{s\}.$$

Inklusio $\mathcal{N}_s \cup \{s\} \subset \bigcup_{C \in \mathcal{A}_s} C$ on selvä, sillä triviaalisti $s \in \bigcup_{C \in \mathcal{A}_s} C$, ja jos $t \in \mathcal{N}_s$, niin $\{s, t\} \in \mathcal{A}_s$.

Olkoon sitten $t \in \bigcup_{C \in \mathcal{A}_s} C$. Jos $t = s$, niin triviaalisti $t \in \mathcal{N}_s \cup \{s\}$. Voidaan siten olettaa, että $t \neq s$. Tällöin olemassa klikki $\tilde{C} \in \mathcal{A}_s$ siten, että $t \in \tilde{C}$. Koska lisäksi $s \in \tilde{C}$, niin $t \in \mathcal{N}_s$. Väite on siten todistettu. □

HUOMAUTUS 2.3.11. Ollaan täten saatu verrattain konkreettinen tapa muodostaa Markovin satunnaiskenttiä. Riittää (tavalla tai toisella) muodostaa naapuripotentiali \mathcal{V}_N , josta johdettu Gibbsin jakauma π toteuttaa Seurauksen 2.3.10 nojalla lokaalin (naapurustosysteemiin \mathcal{N} liittyvän) Markov-ominaisuuden. Satunnaiskenttä X , jonka jakauma on edellä saatu π , on siten myös Markovin satunnaiskenttä.

Tilastotieteilijäpiireissä paljon huomiota saaneessa, aikanaan vaille julkaisua jääneessä artikkelissaan vuodelta 1971 J. Hammersley ja P. Clifford todistivat ensimmäisinä toisen teoreettisesti varsin mielenkiintoisen tuloksen. Sopivasti muotoiltuna se ilmaisee, että jokainen Markovin satunnaiskenttä on Gibbsin kenttä.

LAUSE 2.3.12 (Hammersleyn–Cliffordin lause). *Satunnaiskenttä X on Markovin satunnaiskenttä naapurustojärjestelmän \mathcal{N} suhteen jos ja vain jos se on Gibbsin kenttä jonkin naapuripotentialin $\mathcal{V} = (V_C)_{C \in \mathcal{C}_N}$ suhteen.*

TODISTUS. “ \Rightarrow ”: Seuraus 2.3.10. “ \Leftarrow ”: Möbiuksen kääntökaavaan perustuva G. R. Grimmetin todistus [10] löytyy myös mm. kirjoista [3] ja [30]. □

Epähomogeenisen Markovin ketjun heikko ja vahva ergodisuus

3.1. Epähomogeeninen Markovin ketju

Tässä luvussa käsitellään simuloidun jäähtymisen ja tiettyjen Gibbsin algoritmien matemaattisena perustana toimivia *epähomogeenisia Markovin ketjuja* ja niiden perusominaisuuksia. Käydään lyhyesti läpi myös homogeenisiin Markovin ketjuihin ja siirtymämatriiseihin liittyviä käsitteitä ja tuloksia myöhempiä lukuja varten. Luvun jälkimmäisen puolen aikana perehdytään melko syvällisesti epähomogeenisen ketjun suppenemisen muotoihin; heikkoon ja vahvaan ergodisuuteen. Tavoitteena on löytää simuloidun jäähtymisen suppenemislauseita varten riittävät ehdot epähomogeenisten Markovin ketjujen tasapainojakaumaan suppenemiselle.

Koko tämän luvun ajan $E \neq \emptyset$ tarkoittaa mielivaltaista äärellistä joukkoa.

MÄÄRITELMÄ 3.1.1 (Stokastinen prosessi). Todennäköisyysavaruuden $(\Omega, \mathcal{F}, \mathbb{P})$ E -arvoisista satunnaismuuttujista koostuvaa indeksoitua perhettä $(X_t)_{t \in T}$, $T \subset \mathbb{R}$, kutsutaan *stokastiseksi prosessiksi* tai *satunnaisprosessiksi tila-avaruudella E* .

Mikäli $T = \{0, 1, 2, \dots\}$, kyseessä on *diskreettiaikainen prosessi*. Jos puolestaan $T = [0, \infty)$, puhutaan *jatkuva-aikaisesta prosessista*. Kiinnitetylle $\omega \in \Omega$ kuvausta $t \mapsto X_t(\omega)$ kutsutaan prosessin $(X_t)_{t \in T}$ *poluksi*.

ESIMERKKI 3.1.2. Ehkä tyypillisin esimerkki stokastisesta prosessista (X_t) on tilanne, jossa satunnaismuuttujat X_t ovat keskenään riippumattomia. Standardiesimerkki riippumattomasta diskreettiaikaisesta stokastisesta prosessista on jono $(X_n)_{n \geq 0}$, jossa kukin $X_n \in \{\text{KR}, \text{KL}\}$ edustaa riippumattoman kolikonheiton tulosta hetkellä $n \geq 1$. Ideaaliselle kolikolle $\mathbb{P}(X_n = \text{KL}) = \mathbb{P}(X_n = \text{KR}) = 0.5$ kaikilla $n \geq 1$. Esimerkki prosessin (X_n) *polusta* on ääretön jono $(\text{KR}, \text{KR}, \text{KL}, \text{KR}, \text{KL}, \text{KL}, \dots)$.

MÄÄRITELMÄ 3.1.3 (Markovin ketju). Olkoon $(X_n)_{n \geq 0}$ todennäköisyysavaruuden $(\Omega, \mathcal{F}, \mathbb{P})$ diskreettiaikainen stokastinen prosessi äärellisellä tila-avaruudella E . Prosessia kutsutaan tällöin *Markovin ketjuksi*, mikäli se toteuttaa *Markovin ehdon*:

$$\mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1} \cdots X_0 = x_0) = \mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1})$$

kaikilla $n \geq 1$ ja kaikilla $x_i \in E$, $n \geq i \geq 0$.

Jakaumaa $\mu = (\mathbb{P}(X_0 = x))_{x \in E}$ kutsutaan Markovin ketjun *alkujakaumaksi*. Lisäksi siirtymätodennäköisyyksistä koottua matriisia $P_n := [\mathbb{P}(X_n = y \mid X_{n-1} = x)]_{x, y \in E}$ kutsutaan ketjun *siirtymämatriisiksi* (hetkellä n). Todetaan oitis, että näin määriteltynä $\mu \in \mathcal{D}_E$ ja $P_n \in \mathcal{M}_{E \times E}$ kaikilla $n \geq 1$.

Jos $P_n = P$ kaikilla $n \geq 1$ jollakin $P \in \mathcal{M}_{E \times E}$, $(X_n)_{n \geq 0}$ on *homogeeninen* Markovin ketju. Muussa tapauksessa kyseessä on *epähomogeeninen* Markovin ketju.

ketju prosessin (Y_n, Y_{n+1}, \dots) siirtymätodennäköisyyksien tulee olla riippumattomia prosessin aiemmista arvoista $(Y_0, Y_1, \dots, Y_{n-1})$.

LAUSE 3.1.5 (Marginaalijakaumat ja Chapmanin–Kolmogorovin yhtälö). *Olkoon $(X_n)_{n \geq 0}$ epähomogeeninen Markovin ketju siirtymämatriiseina $(P_n)_{n \geq 1} \subset \mathcal{M}_{E \times E}$ ja alkujakaumanaan $\mu \in \mathcal{D}_E$. Tällöin*

$$(i) \quad \mathbb{P}(X_n = x) = \mu P_1 \cdots P_n(x) \text{ kaikilla } x \in E, n \geq 1.$$

$$(ii) \quad \mathbb{P}(X_n = y \mid X_m = x) = P_{m+1} \cdots P_n(x, y) \text{ kaikilla } x, y \in E, n > m \geq 0.$$

TODISTUS. Todistetaan kohta (i) induktiolla. Todistuksen idea on peräisin kirjasta [15, s. 11–12], jossa tulos (i) todistetaan homogeeniselle Markovin ketjulle. Kohta (ii) menee oleellisesti samaan tapaan, joten sen todistus sivuutetaan.

Tapaus $n = 1$: olkoon $x \in E$ ja $A := \{x_0 \in E : \mathbb{P}(X_0 = x_0) > 0\}$. Tällöin

$$\mathbb{P}(X_1 = x) = \sum_{x_0 \in E} \mathbb{P}(X_1 = x, X_0 = x_0) = \sum_{x_0 \in A} \mathbb{P}(X_1 = x, X_0 = x_0), \text{ sillä}$$

$0 = \mathbb{P}(X_0 = x_0) \geq \mathbb{P}(X_1 = x, X_0 = x_0) \geq 0$ kaikilla $x_0 \in E \setminus A$ ja $x \in E$. Koska lisäksi $\mathbb{P}(X_0 = x_0) > 0$, kun $x_0 \in A$, Määritelmä 1.1.5 implikoi yhtälön

$$\mathbb{P}(X_1 = x, X_0 = x_0) = \mathbb{P}(X_1 = x \mid X_0 = x_0) \mathbb{P}(X_0 = x_0)$$

kaikille $x_0 \in A$ ja $x \in E$. Siten edelleen

$$\begin{aligned} \sum_{x_0 \in A} \mathbb{P}(X_1 = x, X_0 = x_0) &= \sum_{x_0 \in A} \mathbb{P}(X_1 = x \mid X_0 = x_0) \mathbb{P}(X_0 = x_0) \\ &= \sum_{x_0 \in E} \mathbb{P}(X_1 = x \mid X_0 = x_0) \mathbb{P}(X_0 = x_0) \\ &= \sum_{x_0 \in E} \mu(x_0) P_1(x_0, x) = \mu P_1(x). \end{aligned}$$

Tehdään sitten induktio-oletus ja oletetaan, että väite on tosi, kun $n = m$. Tapaus $n = m + 1$: olkoon $x \in E$ ja $B := \{x_m \in E : \mathbb{P}(X_m = x_m) > 0\} \subset E$, jolloin puolestaan $\mathbb{P}(X_{m+1} = x, X_m = x_m) = 0$, kun $x_m \in E \setminus B$. Etenemällä kuten todistuksen alkuosassa ja käyttämällä induktio-oletusta saadaan

$$\begin{aligned} \mathbb{P}(X_{m+1} = x) &= \sum_{x_m \in B} \mathbb{P}(X_{m+1} = x, X_m = x_m) \\ &= \sum_{x_m \in B} \mathbb{P}(X_{m+1} = x \mid X_m = x_m) \mathbb{P}(X_m = x_m) \\ &= \sum_{x_m \in E} \mathbb{P}(X_{m+1} = x \mid X_m = x_m) \mathbb{P}(X_m = x_m) \\ &= \sum_{x_m \in E} \left(\mu P_1 \cdots P_m(x_m) \right) P_{m+1}(x_m, x) \\ &= \mu P_1 \cdots P_{m+1}(x). \end{aligned}$$

□

HUOMAUTUS 3.1.6. Kaikki (epä)homogeeniseen Markovin ketjuun liittyvät äärellisulotteiset jakaumat saadaan laskettua alkujakauman ja siirtymämatriisien avulla.

HUOMAUTUS 3.1.7. Myöhemmin tullaan törmäämään useasti jakaumiin muotoa

$$(3.1) \quad \mu P_i \cdots P_n = \sum_{z \in E} \mathbb{P}(X_n = \cdot \mid X_{i-1} = z) \mu(z), \quad \text{missä } \mu \in \mathcal{D}_E, n > i \geq 1.$$

Tulkinta: jakauma (3.1) kertoo hetken n todennäköisyydet sellaiselle epähomogeeniselle ketjulle $(X_k)_{k \geq 0}$, joka on käynnistetty hetkellä $i - 1$ jakauman μ mukaisesti. Esimerkiksi Diracin pistemassajakaumalle

$$\delta_x : E \rightarrow \mathbb{R}; \quad \delta_x(y) = \mathbf{1}_{\{x\}}(y)$$

jakauma $\delta_x P_2 \cdots P_7 = \mathbb{P}(X_7 = \cdot \mid X_1 = x)$ kertoo todennäköisyydet, kun lähdetään liikkeelle tilasta $x \in E$ hetkellä 1 ja tehdään 6 askelta siirtymämatriisien P_2, \dots, P_7 mukaisesti.

LAUSE 3.1.8 (Markovin ketjun olemassaolo). *Olkoon E äärellinen joukko, $\mu \in \mathcal{D}_E$ ja $(P_n)_{n \geq 1} \subset \mathcal{M}_{E \times E}$. Tällöin on olemassa todennäköisyysavaruus $(\Omega, \mathcal{F}, \mathbb{P})$ ja epähomogeeninen Markovin ketju $(X_n)_{n \geq 0}$ tila-avaruudella E siten, että*

- (i) $\mathbb{P}(X_0 = x) = \mu(x)$,
- (ii) $\mathbb{P}(X_n = y \mid X_{n-1} = x) = P_n(x, y)$ kaikilla $x, y \in E$ ja $n \geq 1$.

TODISTUS. Sivuuetaan. Todistus löytyy muun muassa kirjasta [30, s.77-78]. \square

HUOMAUTUS 3.1.9. Lauseen 3.1.8 ketju $(X_n)_{n \geq 0}$ ei suinkaan ole yksikäsitteinen.

Käydään seuraavaksi lyhyesti läpi tärkeimpiä siirtymämatriiseihin liittyviä käsitteitä.

MÄÄRITELMÄ 3.1.10. Olkoon $P \in \mathcal{M}_{E \times E}$. Sanotaan, että P on

- *aidosti positiivinen* ja merkitään $P > 0$, jos $P(x, y) > 0$ kaikilla $x, y \in E$
- *pelkistymätön*, jos kaikilla $x, y \in E$ on olemassa luku n siten, että $P^n(x, y) > 0$
- *jaksoton*, jos $\text{syt}\{n \geq 1 : P^n(x, x) > 0\} = 1$ kaikilla $x, y \in E$ ¹
- *symmetrinen*, jos $P(x, y) = P(y, x)$ kaikilla $x \in E$
- *primitiivinen*, jos on olemassa luku τ siten, että $P^\tau(x, y) > 0$ kaikilla $x, y \in E$
- *kääntyvä* jakauman $\nu \in \mathcal{D}_E$ suhteen, jos

$$(3.2) \quad \nu(x)P(x, y) = \nu(y)P(y, x) \quad \text{kaikilla } x, y \in E.$$

HUOMAUTUS 3.1.11. Yhtälöitä (3.2) kutsutaan *yksityiskohtaisiksi tasapainoyhtälöiksi*. Lisäksi sanotaan, että $\pi \in \mathcal{D}_E$ on siirtymämatriisin P *invariantti* tai *stationaarinen jakauma*, jos

$$(3.3) \quad \pi P(y) = \sum_{x \in E} \pi(x)P(x, y) = \pi(y) \quad \text{kaikilla } y \in E, \text{ ts. } \pi P = \pi.$$

Yhtälöä (3.3) kutsutaan puolestaan *globaaliksi tasapainoyhtälöksi*.

¹syt = suurin yhteinen tekijä

MÄÄRITELMÄ 3.1.12 (Tilojen kommunikointi). Olkoon $x, y \in E$ ja olkoon $P \in \mathcal{M}_{E \times E}$. Sanotaan, että tilat x ja y *kommunikoiivat*, mikäli on olemassa *polku* eli äärellinen jono z_0, z_1, \dots, z_m joukon E tiloja siten, että $x = z_0, y = z_m$ ja $P(z_{j-1}, z_j) > 0$ kaikilla $j = 1, 2, \dots, m$.

HUOMAUTUS 3.1.13. On varsin suoraviivaista osoittaa, että $P \in \mathcal{M}_{E \times E}$ on pelkistymätön jos ja vain jos jokainen $x \in E$ kommunikoi jokaisen $y \in E$ kanssa.

Kootaan lopuksi yhteen tärkeitä siirtymämatriisien ominaisuuksia.

LAUSE 3.1.14. *Olkoon $P \in \mathcal{M}_{E \times E}$.*

- (i) *P on primitiivinen jos ja vain jos P on pelkistymätön ja jaksoton.*
- (ii) *Jos P on aidosti positiivinen, niin P on primitiivinen.*
- (iii) *Jos P on kääntyvä jakauman $\pi \in \mathcal{D}_E$ suhteen, niin tällöin π on matriisin P invariantti jakauma.*
- (iv) *Primitiiviselle P on olemassa yksikäsitteinen ja aidosti positiivinen invariantti jakauma.*
- (v) *Jos P on primitiivinen invarianttina jakaumanaan $\pi \in \mathcal{D}_E$, niin $\nu P^n \xrightarrow{TV} \pi$ kaikilla alkujakaumilla $\nu \in \mathcal{D}_E$.*

LEMMA 3.1.15. *Olkoon $P \in \mathcal{M}_{E \times E}$, jolle $P^k > 0$ jollakin $k \geq 1$. Tällöin $P^{k+n} > 0$ kaikilla $n \geq 1$.*

TODISTUS. Kiinnitetään $n \geq 1$ ja tilat $x, y \in E$. Olkoon lisäksi $z \in E$, jolle $P^n(x, z) > 0$ (tällainen tila löytyy aina; voi olla $z = x$ tai $z = y$). Oletuksesta $P^k > 0$ seuraa tällöin, että $P^{k+n}(x, y) \geq P^n(x, z)P^k(z, y) > 0$. Nyt koska x, y ja n valittiin mielivaltaisesti, $P^{k+n} > 0$ kaikilla $n \geq 1$. \square

LAUSEEN 3.1.14 TODISTUS: Kohta (i) : “ \Leftarrow ”: ks. esim. [15, s. 26]. “ \Rightarrow ”: matriisin P primitiivisyyden nojalla on olemassa luku $\tau \geq 1$ siten, että $P^\tau > 0$. Pelkistymättömyys seuraa siten välittömästi.

Jaksottomuuden todistus on peräisin kirjasta [13, s. 517]. Kiinnitetään $x \in E$ ja merkitään $L_x = \{n \geq 1 : P^n(x, x) > 0\}$. Primitiivisyyden ja Lemman 3.1.15 nojalla $P^{\tau+n}(x, x) > 0$ kaikilla $n \geq 1$, jolloin $\{\tau, \tau + 1, \tau + 2, \dots\} \subset L_x$. On siis $\text{sytl}_x = 1$, joten P on myös jaksoton.

Kohta (ii) : väite seuraa, kun asetetaan primitiivisyyden määritelmässä $\tau = 1$.

Kohta (iii) : olkoon $y \in E$. Väite seuraa alla olevasta päättelyketjusta:

$$\pi P(y) = \sum_{x \in E} \pi(x) P(x, y) = \sum_{x \in E} \pi(y) P(y, x) = \pi(y) \sum_{x \in E} P(y, x) = \pi(y).$$

Kohta (iv) : katso [30, s. 86].

Kohta (v) : todistetaan myöhemmin (Seuraus 3.4.8). \square

Ennen kuin siirrytään tarkastelemaan Markovin ketjujen suppenemista, tehdään pienimuotoinen ekskursio äärettömien tulojen maailmaan.

3.2. Ääretön tulo

Tarkastellaan hetki *ääretöntä tuloa*

$$(3.4) \quad \prod_{k=1}^{\infty} (1 - a_k) = (1 - a_1)(1 - a_2) \cdots, \text{ missä } a_k \in [0, 1] \text{ kaikilla } k \geq 1.$$

Kysymys kuuluu: mitä tarkoittaa, että ääretön tulo (3.4) suppenee? Tavallisten sarjojen eli äärettömien summien teoriasta tiedetään, että suppeneminen tai hajaantuminen riippuu ainoastaan sarjan *hännästä*: äärellisen monen termin poistaminen ei muuta hajaantuvaa sarjaa suppenevaksi tai päinvastoin. Tavallisesti ei siten haluta antaa myöskään tulosarjojen tapauksessa äärellisen monen termin vaikuttaa sen rajakäyttäytymiseen. Voidaan todeta, että kyseinen häntäominaisuus sisältyy Määritelmään 3.2.1.

Otetaan osatuloille $(1 - a_i) \cdots (1 - a_n)$ käyttöön merkintä

$$A_{i,n} := \prod_{k=i}^n (1 - a_k).$$

Tällöin kaikilla $i \geq 1$ $(A_{i,n})_{n \geq i} = (A_{i,i}, A_{i,i+1}, \dots)$ on rajoitettu ja vähenevä jono reaalityyppisiä lukuja välillä $[0, 1]$, joten on olemassa raja-arvo $\lim_{n \rightarrow \infty} A_{i,n} \in [0, 1]$ riippumatta siitä, mitä arvoja luvut a_k saavat. Äärettömän tulon suppeneminen ja hajaantuminen perustuu näihin raja-arvoihin, kuten seuraavasta määritelmästä nähdään.

MÄÄRITELMÄ 3.2.1. Olkoon $(a_k)_{k \geq 1} \subset [0, 1]$ lukujono. Sanotaan, että *ääretön tulo* $\prod_{k=1}^{\infty} (1 - a_k)$ *suppenee*, jos on olemassa luku $\alpha \in [0, 1]$ siten, että

$$(3.5) \quad \lim_{n \rightarrow \infty} A_{i,n} = \alpha \quad \text{jollekin } \alpha \in [0, 1],$$

jolloin sen *arvo* määritellään asettamalla

$$\prod_{k=1}^{\infty} (1 - a_k) := \left(\prod_{k=1}^{i-1} (1 - a_k) \right) \lim_{n \rightarrow \infty} \left(\prod_{k=1}^n (1 - a_k) \right) = A_{1,i-1} \cdot \lim_{n \rightarrow \infty} A_{i,n} \geq 0.$$

Jos sen sijaan

$$\lim_{n \rightarrow \infty} A_{i,n} = 0 \quad \text{kaikilla } i \geq 1,$$

sanotaan, että sarja *hajaantuu* nolliin.

HUOMAUTUS 3.2.2. Ilmaisua “hajaantuu nolliin” saattaa särähtää lukijan korvaan. Tätä ilmaisua voidaan kuitenkin perustella tarkastelemalla luonnollisia logaritmeja, sillä (olettaen, että $a_k \in [0, 1]$ kaikilla $k \geq 1$) voidaan kirjoittaa

$$\prod_{k=i}^n (1 - a_k) = \exp \left(\log \left(\prod_{k=i}^n (1 - a_k) \right) \right) = \exp \left(\sum_{k=i}^n \log(1 - a_k) \right).$$

Ääretön tulo $\prod_{k=1}^{\infty} (1 - a_k)$ hajaantuu nolliin täsmälleen silloin, kun logaritmien summa $\sum_{k=i}^{\infty} \log(1 - a_k) = -\infty$ kaikilla $i \geq 1$, ts. silloin, kun summa $\sum_{k=1}^{\infty} \log(1 - a_k)$ hajaantuu kohti miinus ääretöntä.

SEURAUUS 3.2.3. Olkoon $(a_k)_{k \geq 1} \subset [0, 1]$. Tällöin ääretön tulo $\prod_{k=1}^{\infty} (1 - a_k)$

(a) suppenee lukuun $L \in (0, 1]$ tai hajaantuu nolnaan, jos $a_k < 1$ kaikilla $k \geq 1$.

(b) suppenee tai hajaantuu nolnaan, jos $\#\{k \geq 1 : a_k = 1\} < \infty$.

(c) hajaantuu nolnaan, jos $a_k = 1$ äärettömän monella $k \geq 1$.

TODISTUS. Kohdat (a) – (c) seuraavat suoraan Määritelmästä 3.2.1. \square

Todistetaan seuraavaksi tärkeää Lausetta 3.2.5 varten pieni aputullos.

LEMMA 3.2.4.

$$1 - x \leq e^{-x} \quad \text{kun } x \in (0, 1].$$

TODISTUS. Olkoon $x \in (0, 1]$. Käyttämällä eksponenttifunktion sarjakehitelmää voidaan arvioida

$$\begin{aligned} -1 + x + e^{-x} &= -1 + x + \sum_{k=0}^{\infty} \frac{(-x)^k}{k!} = \frac{x^2}{2!} - \frac{x^3}{3!} + \frac{x^4}{4!} - \frac{x^5}{5!} + \dots \\ &\geq x^3 \left(\frac{1}{2!} - \frac{1}{3!} \right) + x^5 \left(\frac{1}{4!} - \frac{1}{5!} \right) + \dots = \sum_{k=1}^{\infty} x^{2k+1} \left(\frac{1}{(2k)!} - \frac{1}{(2k+1)!} \right) \geq 0, \end{aligned}$$

joten $e^{-x} \geq 1 - x$, ja väite on siten todistettu. \square

LAUSE 3.2.5. Olkoon $(a_k)_{k \geq 1} \subset [0, 1]$ lukujono. Tällöin ääretön tulo $\prod_{k=1}^{\infty} (1 - a_k)$ hajaantuu nolnaan jos ja vain jos $\sum_{k=1}^{\infty} a_k = \infty$.

TODISTUS. Lauseen todistus on peräisin kirjasta [14, s. 37–38] Oletetaan, että $\sum_{k=1}^{\infty} a_k = \infty$. Jos $a_k = 1$ äärettömän monella $k \geq 1$, ääretön tulo hajaantuu kohti nolaa, jolloin väite on tosi. Oletetaan siis, että on olemassa luku m siten, että $a_k < 1$ kaikilla $k \geq m$. Lemmaa 3.2.4 soveltamalla voidaan tällöin kirjoittaa

$$0 \leq \prod_{k=m}^n (1 - a_k) \leq \prod_{k=m}^n \exp(-a_k) = \exp\left(-\sum_{k=m}^n a_k\right) \quad \text{kaikilla } n \geq m, \text{ joten}$$

$$0 \leq \lim_{n \rightarrow \infty} \prod_{k=m}^n (1 - a_k) \leq \lim_{n \rightarrow \infty} \exp\left(-\sum_{k=m}^n a_k\right) = 0,$$

sillä $\sum_{k=m}^n a_k$ hajaantuu kohti äärettömyyttä. Siten myös tässä tapauksessa ääretön tulo hajaantuu kohti nolaa.

Oletetaan sitten, että ääretön tulo hajaantuu nolnaan. Tehdään antiteesi ja oletetaan, että $\sum_{k=1}^{\infty} a_k < \infty$ (muistetaan, että $a_k \in [0, 1]$). Tämä tarkoittaa, että kaikilla $\varepsilon > 0$ on olemassa luku M siten, että $\sum_{k=M}^{\infty} a_k < \varepsilon$. Helppo induktio termien lukumäärän suhteen osoittaa, että

$$\prod_{k=M}^n (1 - a_k) \geq 1 - \sum_{k=M}^n a_k \quad \text{kaikilla } n \geq M \geq 1.$$

Tätä tietoa hyödyntämällä nähdään, että

$$\prod_{k=M}^n (1 - a_k) \geq 1 - \sum_{k=M}^n a_k \geq 1 - \sum_{k=M}^{\infty} a_k > 1 - \varepsilon \quad \text{kun } n \geq M.$$

Siten

$$\lim_{n \rightarrow \infty} \prod_{k=M}^n (1 - a_k) \geq 1 - \varepsilon > 0.$$

Tämä tarkoittaa, että ääretön tulo $\prod_{k=1}^{\infty} (1 - a_k)$ suppenee, mikä on ristiriita. \square

3.3. Epähomogeenisen Markovin ketjun suppeneminen

Siirrytään seuraavaksi tarkastelemaan epähomogeenisten Markovin ketjujen suppenemisen kahta muotoa: *heikkoa* ja *vahvaa ergodisuutta*. Esitetään aluksi näiden käsitteiden määritelmät ja muutama niihin liittyvä perustulos, jonka jälkeen ryhdytään tarkastelemaan niitä erikseen yksityiskohtaisemmin Kohdissa 3.4 ja 3.5.

MÄÄRITELMÄ 3.3.1 (Heikko ergodisuus). Olkoon $(X_n)_{n \geq 0}$ epähomogeeninen Markovin ketju tila-avaruudella E siirtymämatriiseina $(P_n)_{n \geq 1} \subset \mathcal{M}_{E \times E}$. Sanotaan, että ketju $(X_n)_{n \geq 0}$ on tällöin *heikosti ergodinen*, mikäli

$$\sup_{\mu, \nu \in \mathcal{D}_E} \left| \left| \mu P_i \cdots P_n - \nu P_i \cdots P_n \right| \right| \xrightarrow{n \rightarrow \infty} 0 \quad \text{kaikilla } i \geq 1.$$

MÄÄRITELMÄ 3.3.2 (Vahva ergodisuus). Olkoon $(X_n)_{n \geq 0}$ epähomogeeninen Markovin ketju tila-avaruudella E siirtymämatriiseina $(P_n)_{n \geq 1} \in \mathcal{M}_{E \times E}$. Jos on olemassa jakauma $\pi \in \mathcal{D}_E$ siten, että

$$\sup_{\mu \in \mathcal{D}_E} \left| \left| \mu P_i \cdots P_n - \pi \right| \right| \xrightarrow{n \rightarrow \infty} 0 \quad \text{kaikilla } i \geq 1,$$

sanotaan, että Markovin ketju on *vahvasti ergodinen*.

HUOMAUTUS 3.3.3. Kolmioepäyhtälön avulla helposti nähdään, että jos $(X_n)_{n \geq 0}$ on vahvasti ergodinen, niin tasapainojakauma π on yksikäsitteinen.

LAUSE 3.3.4. *Olkoon $(P_n)_{n \geq 1} \subset \mathcal{M}_{E \times E}$ ja $\pi \in \mathcal{D}_E$. Tällöin kaikilla $n > i \geq 1$*

$$\begin{aligned} \sup_{\mu, \nu \in \mathcal{D}_E} \left| \left| \mu P_i \cdots P_n - \nu P_i \cdots P_n \right| \right| &= \max_{x, y \in E} \left| \left| P_i \cdots P_n(x, \cdot) - P_i \cdots P_n(y, \cdot) \right| \right| \quad \text{ja} \\ \sup_{\mu \in \mathcal{D}_E} \left| \left| \mu P_i \cdots P_n - \pi \right| \right| &= \max_{x \in E} \left| \left| P_i \cdots P_n(x, \cdot) - \pi \right| \right|. \end{aligned}$$

TODISTUS. Sivuutetaan. Homogeenisen ketjun tapaus (joka on helposti yleistettävissä epähomogeenisille ketjuille) käsitellään kirjassa [21, s. 329–330]. \square

SEURAUUS 3.3.5. *Olkoon $(P_n)_{n \geq 1} \subset \mathcal{M}_{E \times E}$ epähomogeenisen Markovin ketjun siirtymämatriisit.*

(i) *Jos $\left| \left| \mu P_i \cdots P_n - \nu P_i \cdots P_n \right| \right| \xrightarrow{n \rightarrow \infty} 0$ kaikilla $i \geq 1$ ja $\mu, \nu \in \mathcal{D}_E$, on ketju heikosti ergodinen.*

(ii) *Jos on olemassa $\pi \in \mathcal{D}_E$ siten, että $\mu P_i \cdots P_n \xrightarrow{TV} \pi$ kaikilla $i \geq 1$ ja $\mu \in \mathcal{D}_E$, on ketju vahvasti ergodinen.*

TODISTUS. Todistetaan ensimmäinen väite. Toinen väite todistetaan käytännössä samalla tavalla. Olkoon $\varepsilon > 0$. Oletuksen nojalla tällöin on olemassa jokaista paria $x, y \in E$ kohden luku $N_{x,y} \geq 1$ siten, että

$$\left| \left| P_i \cdots P_n(x, \cdot) - P_i \cdots P_n(y, \cdot) \right| \right| = \left| \left| \delta_x P_i \cdots P_n - \delta_y P_i \cdots P_n \right| \right| < \varepsilon,$$

kun $n \geq N_{x,y}$, sillä $\{\delta_x : E \rightarrow \mathbb{R} \mid x \in E\} \subset \mathcal{D}_E$. Siten Lauseen 3.3.4 nojalla

$$\sup_{\mu, \nu \in \mathcal{D}_E} \left| \mu P_i \cdots P_n - \nu P_i \cdots P_n \right| = \max_{x, y \in E} \left| P_i \cdots P_n(x, \cdot) - P_i \cdots P_n(y, \cdot) \right| < \varepsilon,$$

kun $n \geq \max\{N_{x,y} \mid x, y \in E\}$. Väite seuraa tästä. \square

3.4. Heikko ergodisuus

Epähomogeenisten Markovin ketjujen suppenemistarkasteluiden ja erityisesti heikon ergodisuuden osoittamisen kannalta varsin käteväksi osoittautuu seuraavaksi määriteltävä apuväline.

MÄÄRITELMÄ 3.4.1. Matriisin $P \in \mathcal{M}_{E \times E}$ *kontraktiokerroin* on

$$c(P) := \frac{1}{2} \max_{x, y} \left| P(x, \cdot) - P(y, \cdot) \right| = \frac{1}{2} \max_{x, y} \sum_{z \in E} |P(x, z) - P(y, z)|.$$

HUOMAUTUS 3.4.2. Kontraktiokerroin on peräisin R.L. Dobrushinin artikkelista [5] vuodelta 1956, jossa määritellään *ergodinen kerroin* (ergodic coefficient) $\alpha(P)$ jatkuvan tila-avaruuden tapauksessa. Äärellisten tila-avaruuksien vastine tälle luvulle on

$$\alpha(P) := 1 - \frac{1}{2} \max_{x, y \in E} \left| P(x, \cdot) - P(y, \cdot) \right| =: 1 - c(P).$$

Varmasti osaksi tästä syystä kontraktiokertoimen $c(P)$ "oikeasta" nimestä ei vaikuta vallitsevan kovin suurta yksimielisyyttä: G. Winkler kutsuu sitä nimellä *contraction coefficient* [30, s. 82], P. Brémaud puolestaan nimellä *Dobrushin's ergodic coefficient* [3, s. 235], kun taas D. Isaacsonin ja R. Madsenin kirjassa [14, s. 144] käytetään nimitystä *delta coefficient*.

Siirrytään nyt tutkimaan kontraktiokertoimen ominaisuuksia. Kaikille $P \in \mathcal{M}_{E \times E}$ pätee $c(P) \in [0, 1]$, sillä $\|\mu - \nu\| \in [0, 2]$ kaikilla $\mu, \nu \in \mathcal{D}_E$. Erityisesti $c(P) = 1$ täsmälleen silloin, kun vähintään kahdella matriisin P rivillä $P(x, \cdot)$ ja $P(y, \cdot)$ jakaumiksi tulkittuna on erillisten kantajat (Lause 1.3.5). Epäyhtälö $c(P) < 1$ pätee kaikille aidosti positiivisille $P \in \mathcal{M}_{E \times E}$ myöhemmin todistettavan Lemman 3.4.6 nojalla.

Yhtälö $c(P) = 0$ on puolestaan voimassa jos ja vain jos P on vakiorivimatriisi. Tähän asiaan palataan vielä myöhemmin. Sitä ennen lasketaan yksi esimerkki ja todistetaan muutamia hyödyllisiä kontraktiokertoimen ominaisuuksia.

ESIMERKKI 3.4.3. Olkoon $E = \{x, y, z\}$ ja

$$P = \begin{bmatrix} P(x, x) & P(x, y) & P(x, z) \\ P(y, x) & P(y, y) & P(y, z) \\ P(z, x) & P(z, y) & P(z, z) \end{bmatrix} = \begin{bmatrix} 0.4 & 0.5 & 0.1 \\ 0.2 & 0.4 & 0.4 \\ 0.7 & 0.2 & 0.1 \end{bmatrix}.$$

Lasketaan rivien väliset kokonaisvaihteluetaisyydet:

$$\begin{aligned} \left| P(x, \cdot) - P(y, \cdot) \right| &= |0.4 - 0.2| + |0.5 - 0.4| + |0.1 - 0.4| = 0.6 \\ \left| P(x, \cdot) - P(z, \cdot) \right| &= |0.4 - 0.7| + |0.5 - 0.2| + |0.1 - 0.1| = 0.6 \\ \left| P(y, \cdot) - P(z, \cdot) \right| &= |0.2 - 0.7| + |0.4 - 0.2| + |0.4 - 0.1| = 1.0. \end{aligned}$$

Siten

$$c(P) = \frac{1}{2} \max_{e, \tilde{e} \in E} \|P(e, \cdot) - P(\tilde{e}, \cdot)\| = \frac{1}{2} \|P(y, \cdot) - P(z, \cdot)\| = \frac{1}{2}.$$

Kontraktiokertoimen tehokkuus perustuu ensisijaisesti seuraavaksi todistettavaan Lauseeseen 3.4.4.

LAUSE 3.4.4. *Olkoon $\mu, \nu \in \mathcal{D}_E$ ja $P, Q \in \mathcal{M}_{E \times E}$. Tällöin*

$$(i) \quad \|\mu P - \nu P\| \leq c(P) \|\mu - \nu\| \quad \text{ja}$$

$$(ii) \quad c(PQ) \leq c(P)c(Q).$$

TODISTUS. Todistus on päällisin puolin esitetty kirjassa [30, s.82-83]. Aloitetaan ottamalla käyttöön seuraavat lyhennysmerkinnät:

$$\mu(f) := \sum_x f(x)\mu(x) \quad \text{ja} \quad Pf(x) := \sum_y f(y)P(x, y), \quad \text{missä } f : E \rightarrow \mathbb{R}.$$

Edellä $\mu(f)$ voidaan tulkita funktion f odotusarvoksi jakauman μ suhteen. Merkintä $Pf(x)$ puolestaan vastaa tilannetta, jossa funktion f odotusarvo otetaan rivivektoria vastaavan jakauman $P(x, \cdot)$ suhteen.

Osoitetaan sitten, että kaikilla $\mu, \nu \in \mathcal{D}_E$ ja $f : E \rightarrow \mathbb{R}$ on voimassa epäyhtälö

$$(3.6) \quad |\mu(f) - \nu(f)| \leq \frac{1}{2} \max_{x, y} |f(x) - f(y)| \|\mu - \nu\|.$$

Kiinnitetään nyt $\mu, \nu \in \mathcal{D}_E$ ja funktio f . Olkoon lisäksi $l = \frac{1}{2}(f_{\max} - f_{\min})$, jolloin suoraviivaisten laskujen avulla voidaan näyttää, että

$$(3.7) \quad \max_x |f(x) - l| = \frac{1}{2} \max_{x, y} |f(x) - f(y)|.$$

Koska

$$\sum_x l\mu(x) - \sum_x l\nu(x) = l\left(\sum_x \mu(x) - \sum_x \nu(x)\right) = 0, \quad \text{niin}$$

$$\begin{aligned} |\mu(f) - \nu(f)| &= \left| \sum_x f(x)\mu(x) - \sum_x f(x)\nu(x) - \sum_x l\mu(x) + \sum_x l\nu(x) \right| \\ &= \left| \sum_x (f(x) - l)\mu(x) - \sum_x (f(x) - l)\nu(x) \right| \\ &= \left| \sum_x (f(x) - l)(\mu(x) - \nu(x)) \right| \leq \left| \max_x (f(x) - l) \sum_x (\mu(x) - \nu(x)) \right| \\ &\leq \max_x |f(x) - l| \|\mu - \nu\| = \frac{1}{2} \max_{x, y} |f(x) - f(y)| \|\mu - \nu\|, \end{aligned}$$

joten epäyhtälö (3.6) on tosi. Lauseen 1.3.5 kohdan (iv) nojalla voidaan kirjoittaa

$$(3.8) \quad \begin{aligned} \|\mu P - \nu P\| &= \max_{|h| \leq 1} \left| \sum_x h(x)(\mu P(x) - \nu P(x)) \right| \\ &= \max_{|h| \leq 1} \left| \sum_x h(x)\mu P(x) - \sum_x h(x)\nu P(x) \right| \end{aligned}$$

$$(3.9) \quad = \max_{|h| \leq 1} \left| \mu P(h) - \nu P(h) \right|.$$

Koska edelleen

$$\begin{aligned}
 \mu P(h) &= \sum_x h(x) \mu P(x) = \sum_x h(x) \left(\sum_y \mu(y) P(y, x) \right) = \sum_y \mu(y) \left(\sum_x h(x) P(y, x) \right) \\
 (3.10) \quad &= \sum_y P h(y) \mu(y) = \mu(P h),
 \end{aligned}$$

ja siten tietysti myös $\nu P(h) = \nu(P h)$, saadaan epäyhtälön (3.6), yhtälöiden (3.7), (3.8) ja (3.10) nojalla

$$\begin{aligned}
 \max_{|h| \leq 1} |\mu P(h) - \nu P(h)| &= \max_{|h| \leq 1} |\mu(P h) - \nu(P h)| \leq \frac{1}{2} \max_{|h| \leq 1} \max_{x, y} |P h(x) - P h(y)| \|\mu - \nu\| \\
 &= \frac{1}{2} \max_{x, y} \max_{|h| \leq 1} \left| \sum_z h(z) (P(x, z) - P(y, z)) \right| \|\mu - \nu\| \\
 &= \frac{1}{2} \max_{x, y} \|P(x, \cdot) - P(y, \cdot)\| \|\mu - \nu\| = c(P) \|\mu - \nu\|.
 \end{aligned}$$

Kohdan (i) väite $\|\mu P - \nu P\| \leq c(P) \|\nu - \mu\|$ on siten yhtälön (3.8) nojalla todistettu.

Olkoon sitten P ja $Q \in \mathcal{M}_{E \times E}$. Todetaan aluksi, että $PQ(z, \cdot) = P(z, \cdot)Q$. Näin on, sillä kaikilla $x \in E$ pätee

$$PQ(z, x) = \sum_y P(z, y)Q(y, x) = (P(z, \cdot)Q)(x),$$

missä $P(z, \cdot)$ tulkitaan alkujakaumana, kun siirrytään yhden askeleen verran siirtymämatriisiin Q mukaan. Täten

$$\begin{aligned}
 c(PQ) &= \frac{1}{2} \max_{x, y} \|PQ(x, \cdot) - PQ(y, \cdot)\| = \frac{1}{2} \max_{x, y} \|P(x, \cdot)Q - P(y, \cdot)Q\| \\
 &\leq \frac{1}{2} \max_{x, y} \|P(x, \cdot) - P(y, \cdot)\| c(Q) = c(P)c(Q).
 \end{aligned}$$

Tämä todistaa kohdan (ii). □

SEURAUS 3.4.5. *Jakaumille $\mu, \nu \in \mathcal{D}_E$ ja matriiseille $P, Q \in \mathcal{M}_{E \times E}$ on voimassa*

$$\|\mu P - \nu P\| \leq \|\mu - \nu\| \quad \text{ja} \quad \|\mu P - \nu P\| \leq 2c(P).$$

TODISTUS. Koska $c(P) \leq 1$ ja $\|\mu - \nu\| \leq 2$, väite seuraa Lauseesta 3.4.4. □

LEMMA 3.4.6. *Jokaiselle siirtymämatriisille $P \in \mathcal{M}_{E \times E}$ on voimassa epäyhtälöt*

$$c(P) \leq 1 - |E| \min_{x, y \in E} P(x, y) \leq 1 - \min_{x, y \in E} P(x, y).$$

Erityisesti, jos P on aidosti positiivinen, niin $c(P) < 1$.

TODISTUS. Lauseen 1.3.5 kohdan (iii) nojalla voidaan arvioida

$$\begin{aligned} c(P) &= \frac{1}{2} \max_{x,y} \|P(x, \cdot) - P(y, \cdot)\| = \frac{1}{2} \max_{x,y} 2 \left(1 - \sum_z P(x, z) \wedge P(y, z) \right) \\ &= 1 - \min_{x,y} \sum_z P(x, z) \wedge P(y, z) \leq 1 - |E| \min_{x,y,z} \{P(x, z), P(y, z)\} \\ &= 1 - |E| \min_{x,y} P(x, y) \leq 1 - \min_{x,y} P(x, y). \end{aligned}$$

Lisäksi, jos P on aidosti positiivinen, niin $\min_{x,y} P(x, y) > 0$, joten edellä todistetun tuloksen nojalla $c(P) < 1$. \square

LEMMA 3.4.7. *Jokaiselle siirtymämatriisille $P \in \mathcal{M}_{E \times E}$ jono $(c(P^n))_{n \geq 1}$ on vähenevä. Jos P on lisäksi primitiivinen, niin $c(P^n) \xrightarrow{n \rightarrow \infty} 0$.*

TODISTUS. Todistuksien ideat löytyvät kirjasta [30, s. 84]. Vähenevyys seuraa suoraan käyttämällä Lausetta 3.4.4:

$$c(P^{n+1}) \leq c(P^n)c(P) \leq c(P^n) \text{ kaikilla } n \geq 1.$$

Tarkastellaan sitten primitiivistä matriisia $P \in \mathcal{M}_{E \times E}$, jolle $Q = P^\tau$ on aidosti positiivinen jollain $\tau \geq 1$. Määritellään kuvaus $K : \mathbb{N} \rightarrow \mathbb{N} \cup \{0\}$ asettamalla

$$K(n) = \max\{k \geq 0 : k\tau \leq n\} = \left\lfloor \frac{n}{\tau} \right\rfloor.$$

Kaikille $n \geq 1$ saadaan siten Lausetta 3.4.4 soveltamalla yläraja²

$$c(P^n) = c(P^{K(n)\tau} P^{n-K(n)\tau}) = c(Q^{K(n)} P^{n-K(n)\tau}) \leq c(Q^{K(n)})c(P^{n-K(n)\tau}) \leq c(Q)^{K(n)}.$$

Lemman 3.4.6 nojalla $c(Q) < 1$. Koska lisäksi $K(n) \xrightarrow{n \rightarrow \infty} \infty$, joten $\lim_{n \rightarrow \infty} c(Q)^{K(n)} = 0$. Väite seuraa päättelyketjusta

$$0 \leq \lim_{n \rightarrow \infty} c(P^n) \leq \limsup_{n \rightarrow \infty} c(P^n) \leq \limsup_{n \rightarrow \infty} c(Q)^{K(n)} = \lim_{n \rightarrow \infty} c(Q)^{K(n)} = 0.$$

\square

SEURAUUS 3.4.8. *Olkoon $P \in \mathcal{M}_{E \times E}$ primitiivinen homogeenisen Markovin ketjun siirtymämatriisi, jolla on invariantti jakauma $\pi \in \mathcal{D}_E$. Tällöin kaikilla alkujakaumilla $\nu \in \mathcal{D}_E$*

$$\nu P^n \xrightarrow{TV} \pi.$$

TODISTUS. Käyttämällä Lausetta 3.4.4 ja Seurausta 3.4.5 saadaan jokaiselle alkujakaumalle $\nu \in \mathcal{D}_E$ johdettua yläraja

$$\|\nu P^n - \pi\| = \|\nu P^n - \pi P^n\| \leq \|\nu - \pi\| c(P^n) \leq 2c(P^n) \text{ kaikilla } n \geq 1.$$

Väite seuraa, sillä siirtymämatriisin P primitiivisyyden ja Lemman 3.4.7 nojalla $c(P^n) \xrightarrow{n \rightarrow \infty} 0$. \square

²Sovitaan, että $P^0 := I \in \mathcal{M}_{E \times E}$ missä I on identtinen matriisi. Koska $c(I) = 1$, epäyhtälö on siten voimassa myös silloin, kun $n < \tau$.

HUOMAUTUS 3.4.9. Seurauksen 3.4.8 todistusta mukailemalla saadaan ylärajan primitiivistä matriisia P vastaavan homogeenisen ketjun sekoittumisajalle. Määritellään ketjun *sekoittumisaika* kuten kirjassa [21]:

$$n_{\text{mix}} := \min \left\{ n \geq 1 : \sup_{\nu \in \mathcal{D}_E} \left| \nu P^n - \pi \right| \leq \frac{1}{4} \right\}.$$

Olettaen, että $c(P) > 0$, on nimittäin voimassa (karkea) yläraja

$$(3.11) \quad n_{\text{mix}} \leq \frac{\log(\frac{1}{8})}{\log(c(P))}.$$

Näin on, sillä jos kiinnitetyle $n \geq 1$ pätee

$$n \geq \frac{\log 8}{\log(1/c(P))} = \frac{\log(\frac{1}{8})}{\log(c(P))},$$

niin yhtäpitävästi $c(P)^n \leq \frac{1}{8}$, ja edelleen

$$\sup_{\nu \in \mathcal{D}_E} \left| \nu P^n - \pi \right| \leq 2c(P^n) \leq 2c(P)^n \leq \frac{1}{4},$$

joten $n \geq n_{\text{mix}}$. Tämä implikoi epäyhtälön (3.11).

Saatua ylärajaa hyödyntämällä saadaan esimerkiksi kontraktiokertoimien $c(P)$ arvoja 0.5, 0.8, 0.9 ja 0.99 vastaavien sekoittumisaikojen ylärajoiksi luvut 3, 10, 20 ja 207 (tässä järjestyksessä).

Tietyissä tilanteissa (esimerkiksi jos P on kääntyvä jakauman π suhteen) sekoittumisajalle voidaan saada parempia arvioita tarkastelemalla matriisin P ominaisarvoja. Asiasta lisää esimerkiksi kirjoissa [30, s. 197–] ja [21, s. 153–].

Epähomogeenisten Markovin ketjujen heikko ergodisuus on yleensä hankalaa todistaa suoraan määritelmästä. Seuraava lemma yhdessä Lemman 3.4.12 kanssa antavat käytännöllisen keinon heikon ergodisuuden toteamiseksi kontraktiokertoimen avulla.

LEMMA 3.4.10. *Olkoon $(P_n)_{n \geq 1} \subset \mathcal{M}_{E \times E}$ epähomogeenisen Markovin ketjun siirtymämatriisit. Tällöin ketju on heikosti ergodinen jos ja vain jos $c(P_i \cdots P_n) \xrightarrow{n \rightarrow \infty} 0$.*

TODISTUS. Väite on suora seuraus kontraktiokertoimen määritelmästä 3.4.1 ja Lauseesta 3.3.4. \square

Tässä kohtaa on hyvä hetki pohdiskella esimerkin avulla, mitä heikko ergodisuus oikeastaan tarkoittaa.

ESIMERKKI 3.4.11. Olkoon $(P_n)_{n \geq 1} \subset \mathcal{M}_{E \times E}$ epähomogeenisen Markovin ketjun $(X_n)_{n \geq 0}$ siirtymämatriisit. Oletetaan lisäksi, että P_2 on vakiorivimatriisi. Tällöin Lauseen 1.4.4 kohtien (ii) ja (iii) nojalla voidaan osoittaa, että myös tulo $P_1 P_2 \cdots P_n$ on vakiorivimatriisi kaikilla $n \geq 2$. Lauseen 1.4.4 kohdan (i) nojalla tämä tarkoittaa, että ketjun jakauma $P_\mu(X_n = \cdot) = \mu P_1 \cdots P_n$ hetkellä $n \geq 2$ ei riipu lainkaan alkujakaumasta μ . Todetaan, että hetkellä $n = 2$ tapahtuu niin sanottu *muistin menetys*. Koska lisäksi vakiorivisyyden nojalla $c(P_1 \cdots P_n) = 0$ kaikilla $n \geq 2$, niin erityisesti $c(P_1 \cdots P_n) \xrightarrow{n \rightarrow \infty} 0$. On kuitenkin huomattava, että *mikään ei takaa* luvun $c(P_i \cdots P_n)$ suppenemista nolnaan, kun $i > 2$. Heikon ergodisuuden idea muistuttaakin muotoa (3.4) olevien äärettömien tulojen nolnaan hajaantumista, sillä myös

niiden yhteydessä vaaditaan, että nollaan painuminen tapahtuu *rajalla*: äärellinen määrä vakiorivimatriiseja äärettömässä tulossa $P_1 P_2 \cdots$ ei riitä takaamaan heikkoa ergodisuutta. Heikon ergodisuuden yhteydessä puhutaankin tavanomaisesti *asymptoottisesta muistin menetyksestä*.

Seuraava varsin hyödylliseksi osoittautuva tulos antaa keinon heikon ergodisuuden osoittamiseksi nollaan hajaantuvien äärettömien tulojen avulla.

LEMMA 3.4.12. *Olkoon $(X_n)_{n \geq 0}$ epähomogeeninen Markovin ketju, jolla on siirtymämatriisit $(P_n)_{n \geq 1} \subset \mathcal{M}_{E \times E}$. Prosessi $(X_n)_{n \geq 0}$ on tällöin heikosti ergodinen jos ja vain jos tulo $P_1 P_2 \cdots$ voidaan esittää blokkien $(P_1 P_2 \cdots P_{T_1})$, $(P_{T_1+1} P_{T_1+2} \cdots P_{T_2})$, \dots tulona siten, että*

$$(3.12) \quad \sum_{k=1}^{\infty} (1 - c(Q_k)) = \infty,$$

missä $Q_k = P_{T_{k-1}+1} \cdots P_{T_k}$ kaikilla $k = 1, 2, \dots$ ja $T_0 = 0$.

TODISTUS. Todistus on peräisin kirjasta [14, s. 151]. Osoitetaan ainoastaan implikaatio "(3.12) \implies (X_n) on heikosti ergodinen", sillä käänteistä implikaatiota ei tässä tutkielmassa tarvita.

Olkoon $i \geq 1$ ja olkoon lisäksi $n > i$ niin suuri, että $T_{K(n)} > T_{M(i)+1}$, missä

$$K(n) = \max\{k \geq 1 : T_k < n\} \quad \text{ja} \quad M(i) = \min\{k \geq 1 : T_k > i\}.$$

Tällöin Lausetta 3.4.4 soveltamalla saadaan

$$(3.13) \quad c(P_i \cdots P_n) \leq c(P_i \cdots P_{T_{M(i)}}) \left[\prod_{k=M(i)+1}^{K(n)} c(Q_k) \right] c(P_{T_{K(n)}+1} \cdots P_n) \leq \prod_{k=M(i)+1}^{K(n)} c(Q_k).$$

Ehto $T_{K(n)} > T_{M(i)+1}$ takaa, että epäyhtälön (3.13) tulon termejä on vähintään kaksi. Epäyhtälöä soveltamalla voidaan edelleen arvioida

$$(3.14) \quad 0 \leq \lim_{n \rightarrow \infty} c(P_i \cdots P_n) \leq \limsup_{n \rightarrow \infty} c(P_i \cdots P_n) \leq \limsup_{n \rightarrow \infty} \prod_{k=M(i)+1}^{K(n)} c(Q_k).$$

Koska oletuksesta $\sum_{k=1}^{\infty} (1 - c(Q_k)) = \infty$ yhdessä Lauseen 3.2.5 kanssa seuraa, että $\prod_{k=i}^{\infty} c(Q_k) = 0$ kaikilla $i \geq 1$, niin erityisesti

$$\limsup_{n \rightarrow \infty} \prod_{k=M(i)+1}^{K(n)} c(Q_k) = \lim_{n \rightarrow \infty} \prod_{k=M(i)+1}^{K(n)} c(Q_k) = 0, \quad \text{sillä} \quad K(n) \xrightarrow{n \rightarrow \infty} \infty.$$

Siten $c(P_i \cdots P_n) \xrightarrow{n \rightarrow \infty} 0$ epäyhtälön (3.14) nojalla. \square

3.5. Vahva ergodisuus

Vahva ergodisuus vastaa *asymptoottista muistin menetystä* ja *suppenemista tasapainojakaumaan*. Kirjassa [14, s. 137] heikon ja vahvan ergodisuuden eroa on havainnollistettu sisällöllisesti suurin piirtein seuraavalla tavalla: epähomogeeninen Markovin ketju $(X_n)_{n \geq 0}$ on heikosti ergodinen, mikäli jakaumat $\mathbb{P}_\mu(X_n = \cdot)$ ja $\mathbb{P}_\nu(X_n = \cdot)$ ovat "lähellä toisiaan" kaikilla alkujakaumilla μ ja ν , kun n on riittävän suuri. Ketjun

vahvaan ergodisuuteen vaaditaan kuitenkin lisäksi se, että peräkkäisiä aika-askelia vastaavien marginaalijakaumien $\mathbb{P}_\mu(X_n = \cdot)$ ja $\mathbb{P}_\mu(X_{n+1} = \cdot)$ täytyy olla ‐lähel‐ lä toisiaan‐ suurilla n (alkujakaumasta μ riippumatta). Nimittäin ketjulle, joka on ainoastaan heikosti ergodinen, nämä peräkkäiset jakaumat saattavat poiketa toisistaan paljonkin riippumatta siitä, kuinka suureksi n kasvatetaan: tämä käy hyvin ilmi Esimerkistä 3.5.1 [14, Example V.1.3.].

ESIMERKKI 3.5.1. Olkoon $|E| = 2$ ja olkoon $(P_n)_{n \geq 1} \subset \mathcal{M}_{E \times E}$ epähomogeenisen Markovin ketjun siirtymämatriisit, joille

$$P_{2n-1} = \begin{bmatrix} 1 - \frac{1}{2n-1} & \frac{1}{2n-1} \\ 1 - \frac{1}{2n-1} & \frac{1}{2n-1} \end{bmatrix}, \quad P_{2n} = \begin{bmatrix} \frac{1}{2n} & 1 - \frac{1}{2n} \\ \frac{1}{2n} & 1 - \frac{1}{2n} \end{bmatrix} \quad \text{kaikilla } n = 1, 2, \dots$$

Koska matriisien $P_n, n \geq 1$, rivit ovat samat, Lausetta 1.4.4 toistuvasti soveltamalla nähdään, että kaikille alkujakaumille $\mu \in \mathcal{D}_E$ pätee

$$\mu P_i \cdots P_m = p_m \quad \text{kaikilla } m \geq i \geq 1,$$

missä jakauma p_m vastaa matriisin P_m riviä. Jos m on pariton, niin $\mu P_i \cdots P_m = (1 - \frac{1}{m}, \frac{1}{m})$. Parittomalle m sen sijaan $\mu P_i \cdots P_m = (\frac{1}{m}, 1 - \frac{1}{m})$. On siten selvää, ettei ketju suppene kohti mitään tasapainojakaumaa $\pi \in \mathcal{D}_E$, joten se ei ole vahvasti ergodinen. Ketju on kuitenkin heikosti ergodinen, sillä jälleen Lausetta 1.4.4 soveltamalla nähdään, että

$$c(P_i \cdots P_m) = c(P_m) = 0 \quad \text{kaikilla } m > i \geq 1.$$

Kirjassa [14, s. 157–180] on myös esitelty lukuisia riittäviä ehtoja vahvan ergodisuuden osoittamiseksi monen tyyppisissä tilanteissa. Ongelmana on, ettei ole tiedossa mitään yleispäteviä, joka tilanteeseen sopivia riittäviä ja välttämättömiä ehtoja vahvan ergodisuuden näyttämiseksi. Menemättä sen syvemmälle matriisiteorian maailmaan esitämme *tämän* tutkielman tarpeisiin vastaavan epähomogeenisten Markovin ketjujen suppenemista koskevan tuloksen, jota voidaan oikeutetusti pitää tämän luvun päätuloksena.

LAUSE 3.5.2 (R. L. Dobrushin). *Olkoon $(P_n)_{n \geq 1} \subset \mathcal{M}_{E \times E}$ jono siirtymämatriiseja, joille $c(P_i \cdots P_n) \xrightarrow{n \rightarrow \infty} 0$ kaikilla $i \geq 1$. Olkoon lisäksi $(\mu_n) \subset \mathcal{D}_E$ jono jakaumia, joille on voimassa*

$$(3.15) \quad \begin{aligned} & \bullet \quad \mu_n P_n = \mu_n \quad \text{kaikilla } n \geq 1, \\ & \bullet \quad \sum_{n=1}^{\infty} \|\mu_n - \mu_{n+1}\| < \infty. \end{aligned}$$

Tällöin on olemassa rajajakauma $\mu_n \xrightarrow{TV} \mu_\infty \in \mathcal{D}_E$ siten, että

$$\nu P_i \cdots P_n \xrightarrow{TV} \mu_\infty \quad \text{kaikilla } i \geq 1 \text{ ja } \nu \in \mathcal{D}_E.$$

Siirtymämatriiseja $(P_n)_{n \geq 1}$ vastaava Markovin ketju on siis vahvasti ergodinen.

HUOMAUTUS 3.5.3. Markovin ketjujen kielellä väite voidaan ilmaista yhtäpitävässä muodossa

$$\mathbb{P}(X_n = y \mid X_m = x) \xrightarrow{n \rightarrow \infty} \mu_\infty(y) \quad \text{kaikilla } x, y \in E \text{ ja } m \geq 0,$$

missä $(X_n)_{n \geq 0}$ on matriiseja $(P_n)_{n \geq 1}$ vastaava epähomogeeninen Markovin ketju (Lem-
ma 1.3.1 ja Lause 3.3.4).

TODISTUS. Kiinnitetään $\varepsilon > 0$. Koska $\sum_{j=1}^{\infty} \|\mu_j - \mu_{j+1}\| < \infty$, Lauseen 1.3.7
nojalla on olemassa rajajakauma $\mu_{\infty} \in \mathcal{D}_E$, jolle $\|\mu_n - \mu_{\infty}\| \xrightarrow{n \rightarrow \infty} 0$, joten voidaan
valita luku $m \leq 3$ niin suureksi, että

$$2 \sup_{j \geq m} \|\mu_{\infty} - \mu_j\| + \sum_{j=m}^{\infty} \|\mu_j - \mu_{j+1}\| < \frac{\varepsilon}{2}.$$

Aloitetaan osoittamalla, että

$$(3.16) \quad \sum_{j=1}^k (\mu_{m+j-1} - \mu_{m+j}) P_{m+j} \cdots P_{m+k} = \mu_m P_{m+1} \cdots P_{m+k} - \mu_{m+k}$$

kaikilla $k \geq 1$. Jos $k = 1$, yhtälön 3.16 vasen puoli yksinkertaistuu muotoon

$$\mu_m P_{m+1} - \mu_{m+1} P_{m+1} = \mu_{m+1} P_{m+1} - \mu_{m+1}$$

ja asia selvä. Jos puolestaan $k \geq 2$, voidaan kirjoittaa

$$\begin{aligned} \sum_{j=1}^k (\mu_{m+j-1} - \mu_{m+j}) P_{m+j} \cdots P_{m+k} &= \sum_{j=1}^k \mu_{m+j-1} P_{m+j} \cdots P_{m+k} - \sum_{j=1}^k \mu_{m+j} P_{m+j} \cdots P_{m+k} \\ &= \mu_m P_{m+1} \cdots P_{m+k} + \sum_{j=2}^k \mu_{m+j-1} P_{m+j} \cdots P_{m+k} - \sum_{j=1}^{k-1} \mu_{m+j} P_{m+j} \cdots P_{m+k} - \mu_{m+k} P_{m+k} \\ &= \mu_m P_{m+1} \cdots P_{m+k} - \mu_{m+k}. \end{aligned}$$

Edellä käytettiin tietoa

$$\begin{aligned} \sum_{j=2}^k \mu_{m+j-1} P_{m+j} \cdots P_{m+k} - \sum_{j=1}^{k-1} \mu_{m+j} P_{m+j} \cdots P_{m+k} \\ = \sum_{j=1}^{k-1} \mu_{m+j} P_{m+j+1} \cdots P_{m+k} - \sum_{j=1}^{k-1} \mu_{m+j} P_{m+j} \cdots P_{m+k} = 0, \end{aligned}$$

sillä invarianssin nojalla

$$(\mu_{m+j} P_{m+j}) P_{m+j+1} \cdots P_{m+k} = \mu_{m+j} P_{m+j+1} \cdots P_{m+k}, \quad \text{kun } 1 \leq j \leq k-1.$$

Yhtälö (3.16) on siis voimassa kaikilla $k \geq 1$. Sitä soveltamalla saadaan edelleen
kaikilla $k \geq 1$

$$(3.17) \quad \begin{aligned} \mu_{\infty} P_m \cdots P_{m+k} - \mu_{\infty} &= (\mu_{\infty} - \mu_m) P_m \cdots P_{m+k} + \mu_m P_{m+1} \cdots P_{m+k} - \mu_{\infty} \\ &= (\mu_{\infty} - \mu_m) P_m \cdots P_{m+k} \\ &\quad + \sum_{j=1}^k (\mu_{m+j-1} - \mu_{m+j}) P_{m+j} \cdots P_{m+k} + \mu_{m+k} - \mu_{\infty}. \end{aligned}$$

Käyttämällä nyt Lausetta 3.4.4, Seurausta 3.4.5 ja yhtälöä (3.17) saadaan arvio

$$\begin{aligned}
& \left\| \mu_\infty P_m \cdots P_{m+k} - \mu_\infty \right\| \\
& \leq \left\| (\mu_\infty - \mu_m) P_m \cdots P_{m+k} + \sum_{j=1}^k (\mu_{m+j-1} - \mu_{m+j}) P_{m+j} \cdots P_{m+k} \right\| + \left\| \mu_{m+k} - \mu_\infty \right\| \\
& \leq \left\| (\mu_\infty - \mu_m) P_m \cdots P_{m+k} \right\| + \left\| \sum_{j=1}^k (\mu_{m+j-1} - \mu_{m+j}) P_{m+j} \cdots P_{m+k} \right\| + \left\| \mu_{m+k} - \mu_\infty \right\| \\
& \leq c(P_m \cdots P_{m+k}) \left\| \mu_\infty - \mu_m \right\| + \sum_{j=1}^k c(P_{m+j} \cdots P_{m+k}) \left\| \mu_{m+j-1} - \mu_{m+j} \right\| + \left\| \mu_{m+k} - \mu_\infty \right\| \\
& \leq \left\| \mu_\infty - \mu_m \right\| + \sum_{j=1}^k \left\| \mu_{m+j-1} - \mu_{m+j} \right\| + \left\| \mu_{m+k} - \mu_\infty \right\| \\
& \leq 2 \sup_{j \geq m} \left\| \mu_\infty - \mu_j \right\| + \sum_{j=m}^{\infty} \left\| \mu_j - \mu_{j+1} \right\| \quad \text{kaikilla } k \geq 1.
\end{aligned}$$

Luvun m valinnan nojalla kaikille kokonaisluvuille $n > m$ ja $1 \leq i < m - 1$ on siten

$$\begin{aligned}
\left\| \nu P_i \cdots P_n - \mu_\infty \right\| &= \left\| (\nu P_i \cdots P_{m-1} - \mu_\infty) P_m \cdots P_n + \mu_\infty P_m \cdots P_n - \mu_\infty \right\| \\
&\leq \left\| (\nu P_i \cdots P_{m-1} - \mu_\infty) P_m \cdots P_n \right\| + \left\| \mu_\infty P_m \cdots P_n - \mu_\infty \right\| \\
&\leq c(P_m \cdots P_n) \left\| \nu P_i \cdots P_{m-1} - \mu_\infty \right\| + \left\| \mu_\infty P_m \cdots P_n - \mu_\infty \right\| \\
&\leq 2c(P_m \cdots P_n) + 2 \sup_{j \geq m} \left\| \mu_\infty - \mu_j \right\| + \sum_{j=m}^{\infty} \left\| \mu_j - \mu_{j+1} \right\| \\
&\leq 2c(P_m \cdots P_n) + \frac{\varepsilon}{2}.
\end{aligned}$$

Valitsemalla n niin suureksi, että $c(P_m \cdots P_n) \leq \frac{\varepsilon}{4}$, saadaan väite todistettua. \square

HUOMAUTUS 3.5.4. Edellä oleva todistus antaa teoriassa keinon tutkia Lauseen 3.5.2 oletukset täyttävien epähomogeenisten Markovin ketjujen sekoittumisaikaa, sillä sen nojalla kaikilla $n \geq 1$ on voimassa epäyhtälö

$$(3.18) \quad \sup_{\nu \in \mathcal{D}_E} \left\| \nu P_1 \cdots P_n - \mu_\infty \right\| \leq 2 \prod_{k=1}^n c(P_k) + 2 \sup_{k \geq 1} \left\| \mu_\infty - \mu_k \right\| + \sum_{k=1}^n \left\| \mu_k - \mu_{k+1} \right\|.$$

Käytännön tilanteissa voi tosin olla äärimmäisen vaikeaa arvioida tarkasti epäyhtälön (3.18) oikean puolen termejä, jos jonojen $(\mu_n)_{n \geq 1}$ ja $(c(P_n))_{n \geq 1}$ käyttäytymistä ei täysin tunneta.

Gibbs-otanta

Olkoon $\mathcal{G} = (S, \mathcal{N})$, suuntaamaton verkko, $|S| = N < \infty$. Kuten Luvussa 2, olkoon konfiguraatioavaruus E muotoa $E = \prod_{s \in S} \Lambda_s$, missä faasiavaruudet Λ_s ovat äärellisiä reaaliakselin \mathbb{R} osajoukkoja. Olkoon lisäksi $X = (X_s)_{s \in S}$ E -arvoinen Gibbsin kenttä todennäköisyysavaruudella (Ω, \mathbb{P}) jonkin naapuripotentialin $\mathcal{V}_{\mathcal{N}}$ suhteen.¹ Voidaan ajatella, että kukin konfiguraatioavaruuden E alkio edustaa mustavalkokuvaa, jonka pikseleitä vastaavat solmut $s \in S$ ja pikseleihin liittyviä harmaan eri sävyjä puolestaan vastaavat eri faasit $\lambda \in \Lambda_s, s \in S$. Tavoitteena on hankkia otos Gibbsin kentän X jakaumasta

$$\mathbb{P}(X = \cdot) = \pi_T = \frac{1}{Z_T} \exp\left(-\frac{U(\cdot)}{T}\right) = \frac{\exp\left(-\frac{U(\cdot)}{T}\right)}{\sum_{y \in E} \exp\left(-\frac{U(y)}{T}\right)},$$

missä $T > 0$ on *lämpötilaparametri* tai *kontrolliparametri*. Ongelmaksi muodostuu se, että partitiofunktion Z_T laskeminen on usein käytännössä mahdotonta, jos E on suuri. Jos esimerkiksi S vastaa 256×256 -neliöruudukkoa ja solmuille yhteinen faasiavaruus Λ 256 harmaan eri sävyä, erilaisten kuvien lukumäärä $|E| = |\Lambda|^N \approx 10^{157,826}$ on valtava [30, s.16]. Jakauman π_T todennäköisyyksiin ei siten päästä käsiksi suoraan.

Eräs tapa kiertää tämä ongelma on käyttää MCMC-menetelmiä (Markov Chain Monte Carlo). Muodostetaan Markovin ketju (X_0, X_1, \dots) tila-avaruudella E , jossa yksi siirtymä vastaa nykyisen konfiguraation muuttumista täsmälleen yhdessä solmussa.² Tarvitaan siis jokin ennalta määrätty (joko deterministinen tai satunnainen) solmujen $s \in S$ päivitysmenetelmä eli *päivitysjono* $(\sigma_n)_{n \geq 1} \subset S$, joka kertoo, minkä solmun σ_n arvoa ollaan kulloinkin muuttamassa. Tämän niin kutsutun *Gibbs-otannan* yhtä askelta vastaava *algoritmi* (Gibbs sampler) on seuraava:

- olkoon $X_{n-1} = x \in E$.
- Arvotaan uusi faasi $\lambda \in \Lambda_{\sigma_n}$ *ehdollisesta jakaumasta* $\pi_T(\cdot \mid x_{S \setminus \{\sigma_n\}})$ ja asetetaan $X_n = y := (\lambda, x_{S \setminus \{\sigma_n\}})$.

Näin muodostettu stokastinen prosessi $(X_n)_{n \geq 0}$ on selvästi Markovin ketju, sillä prosessin tila hetkellä n riippuu ainoastaan edeltävän hetken $n - 1$ tilasta.

Otantamenetelmän idea perustuu siihen, että vaikka jakauma π_T on saavuttamattomissa, sen lokaalit tunnukset ovat yleensä laskettavissa, sillä (lämpötilakorjatuille) lokaaleille tunnuksille saadaan Lauseen 2.3.9 avulla esitys

¹Palautetaan mieleen Luvun 2 merkintä $\mathcal{V}_{\mathcal{N}} = \{V_C : E \rightarrow \mathbb{R} \mid C \subset \mathcal{C}_{\mathcal{N}}\}$, missä $\mathcal{C}_{\mathcal{N}}$ on naapurustojärjestelmän \mathcal{N} indusoimien klikkien kokoelma.

²Ketjun $(X_n)_{n \geq 0}$ "satunnaismuuttujat" ovat itse asiassa satunnaisalkioita, tarkemmin sanottuna $\prod_{s \in S} \Lambda_s$ -arvoisia *satunnaiskuvauksia*.

$$(4.1) \quad \pi_T(x_s \mid x_{S \setminus \{s\}}) = \frac{\prod_{C \in \mathcal{A}_s} \varphi_C(x)}{\sum_{\lambda \in \Lambda_s} \prod_{C \in \mathcal{A}_s} \varphi_C(\lambda, x_{S \setminus \{s\}})},$$

missä $\mathcal{A}_s = \{C \in \mathcal{C}_N : s \in C\}$ ja $\varphi_C(x) = \exp\left(-\frac{V_C(x)}{T}\right)$.

Laskettavuus seuraa siitä tosiasiasta, että yhtälön (4.1) oikean puolen osamäärässä nimittäjän summattavia termejä on vain $|\Lambda_s|$ kappaletta $\prod_{s \in S} |\Lambda_s|$ kappaleen sijaan. Lisäksi energiafunktioista U tarvitaan kulloinkin ainoastaan ne potentiaali-funktiot V_C , joihin solmu s kuuluu. Mitä vähemmän solmulla s on naapureita, sitä vähemmän on olemassa solmun s sisältäviä klikkejä, ja siten pienille naapurustoille lokaalit tunnukset ovat tavallisesti mahdollista laskea.

4.1. Gibbs-otannan päivitysjonoista ja suppenemisestä

Mikäli päivitysjono $(\sigma_n)_{n \geq 1}$ on valittu sopivalla tavalla, Gibbs-otannan algoritmin muodostamalle tila-avaruuden E homogeeniselle Markovin ketjulle $(X_n)_{n \geq 0}$ on voimassa

$$(4.2) \quad \mathbb{P}(X_n = y \mid X_0 = x) \xrightarrow{n \rightarrow \infty} \pi_T(y) \quad \text{kaikilla } x, y \in E.$$

Lisäksi, jos ketju on epähomogeeninen, niin

$$(4.3) \quad \mathbb{P}(X_n = y \mid X_i = x) \xrightarrow{n \rightarrow \infty} \pi_T(y) \quad \text{kaikilla } i \geq 0 \text{ ja } x, y \in E.$$

Simuloimalla Markovin ketjua riittävän monen iteraation ajan päästään siis lähelle otosta jakaumasta π_T . Riittävän pitkän *burn-in* -jakson jälkeen likimääräisesti siis $X_n \sim \pi_T$. On toki huomattava, että lähtökonfiguraation valinnalla ja avaruuden E koolla voi olla suurtakin vaikutusta suppenemiseen.

Gibbs-otantamenetelmät jaetaan tavallisesti kahteen ryhmään riippuen siitä, miten päivitysstrategia $(\sigma_n)_{n \geq 1}$ valitaan. Tyypillinen tapa on numeroida solmut ja suorittaa päivitykset toistaen tätä ennalta määrättyä järjestystä:

$$\begin{aligned} s_1 &= \sigma_1 = \sigma_{N+1} = \sigma_{2N+1} = \dots, \\ s_2 &= \sigma_2 = \sigma_{N+2} = \sigma_{2N+2} = \dots, \\ &\vdots \\ s_N &= \sigma_N = \sigma_{2N} = \sigma_{3N} = \dots \end{aligned}$$

ja niin edelleen. Kyseistä menetelmää kutsutaan *jaksolliseksi Gibbs-otannaksi*. Jokainen solmu tulee siis päivitettyä täsmälleen kerran yhden kiinnitetyn N aika-askelen pituisen jakson aikana. Tämä menetelmä tuottaa epähomogeenisen Markovin ketjun, jolle suppenemistulos (4.3) on voimassa. Suppenemisen takaamiseksi riittää itse asiassa olettaa vähemmän: deterministiselle päivitysjonolle riittää, että jokainen solmu tulee päivitettyksi äärettömän monta kertaa äärettömän pitkän ajanjakson kuluessa. Tämä tulosta [8, Theorem A] vastaava Lause 4.4.6 luonnollisesti kattaa siten myös jaksollisen Gibbs-otannan suppenemisen.

Toinen tapa on valita kullakin hetkellä päivitettävä solmu satunnaisesti (esimerkiksi tasajakaumasta). Riittää olettaa, että jokainen solmu voidaan valita aidosti positiivisella ajanhetkestä n riippumattomalla todennäköisyydellä. Tämä menetelmä tuottaa homogeenisen Markovin ketjun, jonka suppeneminen (4.2) osoitetaan Lauseessa 4.5.3.

4.2. Menetelmän nimestä ja alkuperästä

Tietävästi ensimmäinen julkaistu teksti koskien Gibbs-otantaa on suomalaisen Pentti Suomelan satunnaiskenttiä käsittelevä väitöskirja vuodelta 1976 [28]. Tavallisesti kunnia menetelmän ja *Gibbs sampler*-nimen keksimisestä langetetaan kuitenkin tavallisesti D. ja S. Gemanille tämän tutkielman polttopisteessä olevasta artikkelistaan [8] (1984). Toinen huomattava Gibbs-otantaa koskeva artikkeli on A. Gelfandin ja A. F. M. Smithin vuoden 1990 artikkeli [7], jonka innoittamana alunperin Gibbsin jakaumaan vahvasti identifioitunutta menetelmää alettiin hyödyntää myös useiden muiden yhteisjakaumien yhteydessä. Tietotekniikan kehityksen myötä Gibbsin otannasta on tullut varsinkin Bayes-tilastotieteilijöiden piirissä tärkeä apuväline esimerkiksi posteriorijakaumien laskemisessa.

4.3. Suppenemistulokset, johdanto

MÄÄRITELMÄ 4.3.1. Olkoon x ja y joukon E konfiguraatioita. Sanotaan, että x ja y ovat s -naapureita, ja merkitään $x \overset{s}{\sim} y$, jos $x_{S \setminus \{s\}} = y_{S \setminus \{s\}}$.

HUOMAUTUS 4.3.2. Relaatiot $\overset{s}{\sim}$, $s \in S$, ovat ekvivalenssirelaatioita joukolla E . Relaatioiden refleksiivisyys, symmetrisyys ja transitiivisuus seuraavat suoraan määritelmästä.

Oletetaan tämän luvun ajan, että lämpötilaparamteri $T > 0$ on jokin kiinteä vakio. Määritellään nyt solmun $s \in S$ päivitystä vastaava Gibbsin algoritmista juontuva siirtymämatriisi $\Gamma_{T,s} \in \mathcal{M}_{E \times E}$:

$$\Gamma_{T,s}(x, y) = \begin{cases} \pi_T(y_s \mid x_{S \setminus \{s\}}), & \text{jos } x \overset{s}{\sim} y \\ 0, & \text{muuten} \end{cases}.$$

LAUSE 4.3.3. Siirtymämatriisi $\Gamma_{T,s}$ ja Gibbsin jakauma π_T toteuttavat yksityiskohtaiset tasapainoyhtälöt, eli

$$\pi_T(x)\Gamma_{T,s}(x, y) = \pi_T(y)\Gamma_{T,s}(y, x) \quad \text{kaikilla } x, y \in E.$$

TODISTUS. Tapaus $x = y$ on selvä. Lisäksi, jos x ja y poikkeavat toisistaan useammassa kuin yhdessä solmussa, tasapainoyhtälön molemmat puolet antavat nollan. Ainoa epätriviaali tapaus on siten tilanne, jossa $x \neq y$ ja $x \overset{s}{\sim} y$. Tällöin

$$\begin{aligned} \pi_T(x)\Gamma_{T,s}(x, y) &= \pi_T(x)\pi_T(y_s \mid x_{S \setminus \{s\}}) = \pi_T(x) \frac{\pi_T(y_s, x_{S \setminus \{s\}})}{\sum_{\lambda \in \Lambda_s} \pi_T(\lambda, x_{S \setminus \{s\}})} \\ &= \pi_T(x_s, y_{S \setminus \{s\}}) \frac{\pi_T(y_s, y_{S \setminus \{s\}})}{\sum_{\lambda \in \Lambda_s} \pi_T(\lambda, y_{S \setminus \{s\}})} = \pi_T(y_s, y_{S \setminus \{s\}}) \frac{\pi_T(x_s, y_{S \setminus \{s\}})}{\sum_{\lambda \in \Lambda_s} \pi_T(\lambda, y_{S \setminus \{s\}})} \\ &= \pi_T(y)\pi_T(x_s \mid y_{S \setminus \{s\}}) = \pi_T(y)\Gamma_{T,s}(y, x). \end{aligned}$$

□

4.4. Deterministinen päivitysjono

Määritellään kutakin solmua $s \in S$ vastaavat, energiefunktion U maksimaalista lokaalia värähtelyä s -naapureiden joukossa kuvaavat luvut

$$\Delta_s := \max_{x,y \in E} \{|U(x) - U(y)| : x \sim y\}; \text{ olkoon lisäksi } \Delta := \max_{s \in S} \Delta_s \geq 0.^3$$

LEMMA 4.4.1. $\Delta = 0$ jos ja vain jos U on vakiofunktio.

TODISTUS. Jos $U(x) = U(y)$ kaikilla $x, y \in E$, niin triviaalisti myös $\Delta = 0$.

Olkoot sitten $\Delta = 0$ ja $x, y \in E$. Numeroidaan $S = \{s_1, \dots, s_N\}$ ja määritellään konfiguraatiot $z^{(0)}, z^{(1)}, \dots, z^{(N)}$ asettamalla

$$\begin{aligned} z^{(0)} &= x, & z^{(1)} &= (y_{s_1}, x_{S \setminus \{s_1\}}), & z^{(2)} &= (y_{\{s_1, s_2\}}, x_{S \setminus \{s_1, s_2\}}), & \dots \\ z^{(N-1)} &= (y_{S \setminus \{s_N\}}, x_{s_N}) \text{ ja } z^{(N)} &= y. \end{aligned}$$

Koska oletuksen nojalla $\Delta = 0$, ja siten tietysti myös $\Delta_{s_k} = 0$ kaikilla $k = 1, 2, \dots, N$, ja koska edelleen $z^{(i-1)} \stackrel{s_i}{\sim} z^{(i)}$ kaikilla $i = 1, 2, \dots, N$, saadaan

$$U(x) = U(z^{(0)}) = U(z^{(1)}) = \dots = U(z^{(N)}) = U(y).$$

□

HUOMAUTUS 4.4.2. Lemman 4.4.1 nojalla tilanteessa $\Delta = 0$ jokaisella konfiguraatiolla $x \in E$ on sama energia, mistä puolestaan seuraa, että π_T on konfiguraatioavaruuden E tasajakauma. Koska tämä tilanne ei ole kovinkaan kiinnostava, *oletetaan tästä lähtien, että $\Delta > 0$* . Todetaan vielä, että $\Delta < \infty$, sillä joukon E äärellisyyden nojalla $\max_{x \in E} |U(x)| < \infty$.

LEMMA 4.4.3. *Olkoon $\delta_T := \min_{s \in S} \min_{x \in E} \pi_T(x_s | x_{S \setminus \{s\}})$. Tällöin $\delta_T \in (0, 1)$. On lisäksi voimassa epäyhtälö*

$$\delta_T \geq \frac{1}{|E|} \exp\left(-\frac{\Delta}{T}\right).$$

TODISTUS. Kaikille $x \in E$ ja $s \in S$ on voimassa

$$\pi_T(x_s | x_{S \setminus \{s\}}) = \frac{\pi_T(x)}{\sum_{\lambda \in \Lambda_s} \pi_T(\lambda, x_{S \setminus \{s\}})} = \frac{\pi_T(x)}{\pi_T(x) + \sum_{\lambda \in \Lambda_s \setminus \{x_s\}} \pi_T(\lambda, x_{S \setminus \{s\}})} < \frac{\pi_T(x)}{\pi_T(x)} = 1,$$

joten myös $\delta_T < 1$. Lisäksi, koska luvun δ_T määritelmässä minimi otetaan yli äärellisen joukon, voidaan valita $x \in E$ ja $s \in S$ siten, että $\delta_T = \pi_T(x_s | x_{S \setminus \{s\}})$. Olkoon $y_s = \arg \min_{\lambda \in \Lambda_s} U(\lambda, x_{S \setminus \{s\}})$, jolloin $0 \leq U(x) - U(y_s, x_{S \setminus \{s\}}) \leq \Delta_s \leq \Delta$. Saadaan

$$\begin{aligned} \delta_T = \pi_T(x_s | x_{S \setminus \{s\}}) &= \frac{\exp\left(-\frac{U(x)}{T}\right)}{\sum_{\lambda \in \Lambda_s} \exp\left(-\frac{U(\lambda, x_{S \setminus \{s\}})}{T}\right)} \geq \frac{\exp\left(-\frac{U(x)}{T}\right)}{|\Lambda_s| \exp\left(-\frac{U(y_s, x_{S \setminus \{s\}})}{T}\right)} \\ &= \frac{1}{|\Lambda_s|} \exp\left(-\frac{U(x) - U(y_s, x_{S \setminus \{s\}})}{T}\right) \geq \frac{1}{|\Lambda_s|} \exp\left(-\frac{\Delta_s}{T}\right) \geq \frac{1}{|E|} \exp\left(-\frac{\Delta}{T}\right) > 0. \end{aligned}$$

□

³Gemanien artikkelin [8] todistuksissa käytetään em. luvun Δ tilalla maksimaalista *globaalia* värähtelyä vastaavaa lukua $\max_{x,y \in E} \{|U(x) - U(y)| : x, y \in E\}$, vaikkakin samaisessa artikkelissa todetaan suppenemistulosten toteutuvan myös (mahdollisesti pienemmälle) luvulle Δ .

Olkoon sitten $(\sigma_n)_{n \geq 1} \subset S$ sellainen deterministinen jono, jolle $\#\{n \geq 1 : \sigma_n = s\} = \infty$ kaikilla $s \in S$. Määritellään epähomogeeninen Markovin ketju, jonka siirtymämatriiseiksi $(P_n)_{n \geq 1}$ asetetaan

$$(4.4) \quad P_n(x, y) := \Gamma_{T, \sigma_n}(x, y) = \begin{cases} \pi_T(x_{\sigma_n} \mid x_{S \setminus \{\sigma_n\}}), & \text{jos } x \stackrel{\sigma_n}{\sim} y \\ 0, & \text{muuten.} \end{cases}$$

Määritellään lisäksi aidosti kasvava jono ajanhetkiä $T_k, k = 0, 1, 2, \dots$, asettamalla

$$T_0 = 0, \quad T_k = \min \left\{ n > T_{k-1} : \{\sigma_{T_{k-1}+1}, \dots, \sigma_n\} = S \right\} \quad \text{kaikilla } k \geq 1.$$

Oletus päivitysjonosta $(\sigma_n)_{n \geq 1}$ takaa, että kaikille $m \geq 1$ on olemassa luku $n > m$ siten, että $\{\sigma_m, \sigma_{m+1}, \dots, \sigma_n\} = S$. Jono $(T_k)_{k \geq 0}$ on siis hyvin määritelty, ja lisäksi $T_k \xrightarrow{k \rightarrow \infty} \infty$. Erityisesti jono $(K(n))_{n \geq 1}$, missä $K(n) = \max\{k \geq 1 : T_k < n\}$, kasvaa rajatta, kun $n \rightarrow \infty$.

Määritellään sitten kaikilla $k = 1, 2, \dots$ yhtä täyttä päivityskierrosta vastaavat siirtymämatriisitblokit asettamalla

$$Q_k := P_{T_{k-1}+1} \cdots P_{T_k}.$$

Siirtymämatriiseja $(P_n)_{n \geq 1}$ vastaavan epähomogeenisen Markovin ketjun $(X_n)_{n \geq 0}$ avulla ilmaistuna on siten

$$Q_k(x, y) = \mathbb{P}(X_{T_k} = y \mid X_{T_{k-1}} = x) \quad \text{kaikilla } x, y \in E, k \geq 1.$$

Ajanhetkien T_k määritelmästä seuraa, että Q_k on aidosti positiivinen kaikilla $k \geq 1$. Tämä ei kuitenkaan itsessään vielä riitä ketjun heikon ergodisuuden näyttämiseen. Ketjun heikko ergodisuus saadaan kuitenkin näytettyä, sillä matriisin Q_k todennäköisyyksille on olemassa indeksistä k riippumaton aidosti positiivinen alaraja.

LEMMA 4.4.4. *Kaikilla $k \geq 1$ on voimassa epäyhtälö*

$$\min_{x, y \in E} Q_k(x, y) \geq (\delta_T)^N.$$

TODISTUS. Olkoon $\{s_1, s_2, \dots, s_N\}$ jokin joukon S numerointi. Kuten todettua, jokainen solmu s_i päivitetään vähintään kerran jokaisen ajanjakson $(T_{k-1}, T_k]$ aikana. Voidaan siten määritellä kaikilla $k \geq 1$ ajanhetket

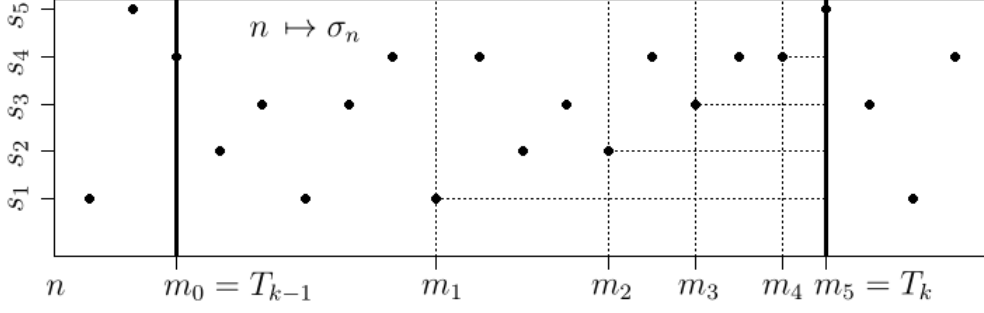
$$m_i(k) := \max\{n \leq T_k : \sigma_n = s_i\},$$

missä $m_i(k)$ on solmun s_i viimeinen päivityshetki aikavälillä $(T_{k-1}, T_k]$. Olkoon lisäksi $m_0(k) := T_{k-1}$. Kiinnitetään nyt $k \geq 1$, jolloin voidaan merkitä lyhyemmin $m_i(k) = m_i$. Numeroidaan lisäksi solmut (mahdollisesti uudelleen) siten, että on voimassa järjestys

$$T_{k-1} = m_0 < m_1 < \dots < m_{N-1} < m_N = T_k \quad (\text{näin voidaan tietysti tehdä}).$$

Ylläolevaa tilannetta on pyritty havainnollistamaan esimerkin avulla Kuvassa 4.1. Kiinnitetään sitten konfiguraatiot $x, y \in E$ ja määritellään joukot

$$E^{(0)} := E, \quad E^{(i)} := \{z \in E : z_{s_k} = y_{s_k}, k = 1, 2, \dots, i\} \quad \text{kaikilla } i = 1, \dots, N.$$



KUVA 4.1. Esimerkkikuva erään päivitysjonon $(\sigma_n)_{n \geq 1}$ yhtä täyttä päivityskierrosta vastaavasta graafista; $|S| = 5$.

Välitön seuraus on, että $E^{(0)} \supset E^{(1)} \supset \dots \supset E^{(N-1)} \supset E^{(N)} = \{y\}$. Näitä merkintöjä käyttämällä voidaan nyt kirjoittaa

$$\begin{aligned} Q_k(x, y) &= P_{T_{k-1}+1} \cdots P_{T_k}(x, y) = P_{m_0+1} \cdots P_{m_N}(x, y) \\ &= \sum_{\substack{z_i \in E^{(i)}; \\ 1 \leq i \leq N}} P_{m_0+1} \cdots P_{m_1}(x, z_1) P_{m_1+1} \cdots P_{m_2}(z_1, z_2) \cdots P_{m_{N-1}+1} \cdots P_{m_N}(z_{N-1}, z_N) \\ &= \sum_{\substack{z_i \in E^{(i)}; \\ 1 \leq i \leq N}} R_0(x, z_1) R_1(z_1, z_2) \cdots R_{N-1}(z_{N-1}, z_N), \end{aligned}$$

missä matriisituloille $P_{m_i+1} \cdots P_{m_{i+1}}$ on käytetty lyhennysmerkintää R_i .

Todistetaan tässä välissä edellä olevia merkintöjä käyttäen pieni todistuksen sisäinen aputuloks.

ALIVÄITE: Kaikilla $i = 0, 1, \dots, N-1$ ja $w \in E^{(i)}$ on voimassa epäyhtälö

$$\sum_{v \in E^{(i+1)}} R_i(w, v) \geq \delta_T.$$

ALIVÄITTEEN TODISTUS: Kiinnitetään $0 \leq i \leq N-1$ ja $w \in E^{(i)}$. Koska $P_{m_{i+1}} = \Gamma_{T, \sigma_{m_{i+1}}} = \Gamma_{T, s_{i+1}}$ ja $\Gamma_{T, s_{i+1}}(x, y) > 0$ jos vain jos $x \stackrel{s_{i+1}}{\sim} y$, niin

$$\sum_{v \in E^{(i+1)}} P_{m_{i+1}}(w, v) = \Gamma_{T, s_{i+1}}(w, \tilde{v}), \text{ missä } \tilde{v} = (y_{s_{i+1}}, w_{S \setminus s_{i+1}}) \in E^{(i+1)}.$$

Siten

$$(4.5) \quad \sum_{v \in E^{(i+1)}} P_{m_{i+1}}(w, v) = \Gamma_{T, s_{i+1}}(w, \tilde{v}) = \pi_T(y_{s_{i+1}} \mid w_{S \setminus s_{i+1}}) \geq \delta_T.$$

Edelleen, koska oletusten nojalla hetken m_i mennessä ketjun arvot solmuissa s_1, \dots, s_i on lukittu arvoihin y_{s_1}, \dots, y_{s_i} , todennäköisyys siirtyä tilaan $z \in E \setminus E^{(i)}$ aikavälillä $(m_i, T_k]$ on nolla. Siten

$$(4.6) \quad \sum_{z \in E^{(i)}} P_{m_{i+1}} \cdots P_{m_{i+1}-1}(w, z) = \sum_{z \in E} P_{m_{i+1}} \cdots P_{m_{i+1}-1}(w, z) = 1.$$

Tuloksia (4.5) ja (4.6) soveltamalla saadaan nyt

$$\begin{aligned}
\sum_{v \in E^{(i+1)}} P_{m_i+1} \cdots P_{m_i+1}(w, v) &= \sum_{v \in E^{(i+1)}} \sum_{z \in E} P_{m_i+1} \cdots P_{m_i+1-1}(w, z) P_{m_i+1}(z, v) \\
&= \sum_{z \in E} P_{m_i+1} \cdots P_{m_i+1-1}(w, z) \left(\sum_{v \in E^{(i+1)}} P_{m_i+1}(z, v) \right) \\
&= \sum_{z \in E^{(i)}} P_{m_i+1} \cdots P_{m_i+1-1}(w, z) \left(\sum_{v \in E^{(i+1)}} P_{m_i+1}(z, v) \right) \\
&\geq \delta_T \sum_{z \in E^{(i)}} P_{m_i+1} \cdots P_{m_i+1-1}(w, z) = \delta_T.
\end{aligned}$$

Siten $\sum_{v \in E^{(i+1)}} R_i(w, v) = \sum_{v \in E^{(i+1)}} P_{(m_i)+1} \cdots P_{m_i+1}(w, v) \geq \delta_T$. $\square_{\text{ALIVÄITE}}$

Koska konfiguraatiot $x, y \in E$ sekä $k \geq 1$ valittiin mielivaltaisesti, itse väite voidaan nyt todistaa soveltamalla N kertaa aputulosta, kun huomataan kirjoittaa

$$\begin{aligned}
Q_k(x, y) &= \sum_{\substack{z_i \in E^{(i)}; \\ 1 \leq i \leq N}} R_0(x, z_1) R_1(z_1, z_2) \cdots R_{N-1}(z_{N-1}, z_N) \\
&= \sum_{z_1 \in E^{(1)}} R_0(x, z_1) \left(\sum_{z_2 \in E^{(2)}} R_1(z_1, z_2) \left(\cdots \left(\sum_{z_N \in E^{(N)}} R_{N-1}(z_{N-1}, z_N) \right) \cdots \right) \right) \\
&\geq \delta_T \underbrace{(\delta_T \cdots (\delta_T) \cdots)}_{N-1 \text{ kpl}} = (\delta_T)^N.
\end{aligned}$$

\square

SEURAUUS 4.4.5. Kaavalla (4.4) määritellyt siirtymämatriisit $(P_n)_{n \geq 1} \subset \mathcal{M}_{E \times E}$ toteuttavat ehdon $c(P_i \cdots P_n) \xrightarrow{n \rightarrow \infty} 0$ kaikilla $i \geq 1$.

TODISTUS. Lemmojen 4.4.3, 4.4.4 ja 3.4.6 nojalla

$$c(Q_k) \leq 1 - \min_{x, y \in E} Q_k(x, y) \leq 1 - (\delta_T)^N \leq 1 - \frac{1}{|E|^N} \exp\left(-\frac{N\Delta}{T}\right) \quad \text{kaikilla } k \geq 1.$$

Merkitään $\alpha := |E|^{-N} \exp\left(-\frac{N\Delta}{T}\right) \in (0, 1)$, jolloin

$$\sum_{k=1}^{\infty} (1 - c(Q_k)) \geq \sum_{k=1}^{\infty} \alpha = \infty,$$

sillä luku α ei riipu indeksistä k . Koska $P_1 P_2 \cdots = Q_1 Q_2 \cdots$, väite seuraa käyttämällä Lemmoja 3.4.10 ja 3.4.12. \square

Voidaan nyt kirjoittaa deterministisen päivitysjonon tapausta vastaava suppenemistulos lauseen muodossa.

LAUSE 4.4.6. Olkoon $E = \prod_{s \in S} \Lambda_s$ äärellinen konfiguraatioavaruus ja $(\sigma_n)_{n \geq 1} \subset S$ deterministinen päivitysjono, jolle $\#\{n \geq 1 : \sigma_n = s\} = \infty$ kaikilla $s \in S$. Olkoon lisäksi siirtymämatriisit $(P_n)_{n \geq 1} \subset \mathcal{M}_{E \times E}$ määritelty kaavalla (4.4). Tällöin kaikilla alkujakaumilla $\nu \in \mathcal{D}_E$

$$\nu P_i \cdots P_n \xrightarrow{TV} \pi_T \quad \text{kaikilla } i \geq 1.$$

TODISTUS. Lauseiden 3.1.14 ja 4.3.3 nojalla π_T on invariantti jakauma jokaiselle $P_n, n \geq 1$. Kontraktiokertoimen ominaisuuksia ja Seurausta 4.4.5 soveltamalla voidaan nyt arvioida alkujakaumalle $\nu \in \mathcal{D}_E$

$$\left\| \nu P_i \cdots P_n - \pi_T \right\| = \left\| \nu P_i \cdots P_n - \pi_T P_i \cdots P_n \right\| \leq \|\nu - \pi_T\| c(P_i \cdots P_n).$$

Väite seuraa, sillä $c(P_i \cdots P_n) \xrightarrow{n \rightarrow \infty} 0$ Seurauksen 4.4.5 nojalla. \square

4.5. Satunnainen päivitysjono

Olkoon nyt $\psi \in \mathcal{D}_S$ jokin aidosti positiivinen jakauma solmujen joukolla S . Muodostetaan homogeeninen Markovin ketju, jonka siirtymämatriisi vastaa yhden solmun satunnaista päivitystä:

$$(4.7) \quad P(x, y) := \sum_{s \in S} \psi(s) \Gamma_{T,s}(x, y), \text{ toisin sanoen}$$

$$P(x, y) = \begin{cases} \sum_{s \in S} \psi(s) \pi_T(x_s | x_{S \setminus \{s\}}), & \text{jos } x = y \\ \psi(s) \pi_T(y_s | x_{S \setminus \{s\}}), & \text{jos } x \stackrel{s}{\sim} y, x \neq y \\ 0, & \text{muuten.} \end{cases}$$

On selvää, että $P(x, y) \geq 0$. Koska lisäksi

$$\begin{aligned} \sum_{y \in E} P(x, y) &= \sum_{y \in E} \left(\sum_{s \in S} \psi(s) \Gamma_{T,s}(x, y) \right) = \sum_{s \in S} \psi(s) \left(\sum_{y \in E} \Gamma_{T,s}(x, y) \right) \\ &= \sum_{s \in S} \psi(s) = 1, \quad P \text{ on stokastinen } |E| \times |E| \text{-matriisi.} \end{aligned}$$

LEMMA 4.5.1. P on primitiivinen siirtymämatriisi.

TODISTUS. Osoitetaan, että $P^N(x, y) > 0$ kaikilla $x, y \in E$, missä $N = |S|$. Olkoon $x, y \in E$. Olkoon sitten $S = \{s_1, s_2, \dots, s_N\}$ joukon S numerointi ja konfiguraatiot $x = z^{(0)}, z^{(1)}, \dots, z^{(N)} = y$ konfiguraatiot kuten Lemman 4.4.1 todistuksessa. Koska tällöin $z^{(i-1)} \stackrel{s_i}{\sim} z^{(i)}$ kaikilla $i = 1, 2, \dots, N$, niin edelleen

$$\begin{aligned} P^N(x, y) &\geq \prod_{i=1}^N P(z^{(i-1)}, z^{(i)}) \geq \prod_{i=1}^N \psi(s_i) \Gamma_{T,s_i}(z^{(i-1)}, z^{(i)}) \\ &= \prod_{i=1}^N \psi(s_i) \pi_T(z_{s_i}^{(i)} | z_{S \setminus \{s_i\}}^{(i-1)}) > 0, \end{aligned}$$

sillä ψ on aidosti positiivinen jakauma. \square

LEMMA 4.5.2. *Jakauma π_T on siirtymämatriisin P yksikäsitteinen invariantti jakauma.*

TODISTUS. Osoitetaan, että π_T on invariantti jakauma siirtymämatriisille P tarkastamalla yksityiskohtaiset tasapainoyhtälöt. Olkoon $x, y \in E$. Käyttämällä Lausetta 4.3.3 saadaan

$$\begin{aligned} \pi_T(x)P(x, y) &= \pi_T(x) \sum_{s \in S} \psi(s) \Gamma_{T,s}(x, y) = \sum_{s \in S} \psi(s) \pi_T(x) \Gamma_{T,s}(x, y) \\ &= \sum_{s \in S} \psi(s) \pi_T(y) \Gamma_{T,s}(y, x) = \pi_T(y) \sum_{s \in S} \psi(s) \Gamma_{T,s}(y, x) \\ &= \pi_T(y)P(y, x). \end{aligned}$$

Lisäksi, koska P on Lemman 4.5.1 nojalla primitiivinen, Lauseesta 3.1.14 seuraa, että π_T on yksikäsitteinen invariantti jakauma. \square

LAUSE 4.5.3. *Olkoon $E = \prod_{s \in S} \Lambda_s$ äärellinen konfiguraatioavaruus, $\psi \in \mathcal{D}_S$ aidosti positiivinen jakauma ja olkoon siirtymämatriisi $P \subset \mathcal{M}_{E \times E}$ määritelty kaavalla (4.7). Tällöin kaikilla alkujakaumilla $\nu \in \mathcal{D}_E$*

$$\nu P^n \xrightarrow{TV} \pi_T.$$

TODISTUS. Koska π_T on siirtymämatriisin primitiivisen P (yksikäsitteinen) invariantti jakauma Lemman 4.5.2 nojalla, väite seuraa Lauseesta 3.4.8. \square

Simuloidun jäähtyksen teoria ja Gibbs-jäähtyksen suppeneminen

5.1. Taustoja ja algoritmin esittelyä

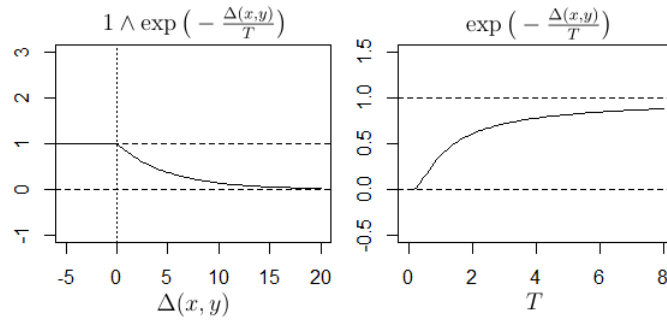
Esimerkki kombinatorisesta ongelmasta matemaattisesti formuloituna on pari (E, U) , missä E on äärellinen *konfiguraatioavaruus*¹ ja $U : E \rightarrow \mathbb{R}$ niin sanottu *energiafunktio* (tai *sakkofunktio*), joka pyritään minimoimaan. Tavoitteena on toisin sanoen löytää energiafunktion U globaali minimikohta eli konfiguraatio $x_0 \in E$, jolle $U(x_0) \leq U(x)$ kaikilla $x \in E$. Minimointi tapahtuu tavallisesti tekemällä siirtymiä joukon E alkioden välillä ja samalla asteittain vähentämällä energiafunktion arvoa. Tavallisten iteratiivisten algoritmien, jotka pyrkivät pienentämään energiaa jokaisella askeleella, ongelmana on niiden terminoituminen ensimmäiseen *lokaaliin* minimikohtaan eli konfiguraatioon $x \in E$, jonka jokaiselle naapurille $y \in E$ on voimassa $U(x) \leq U(y)$. Simuloidussa jäähtyksessä tämä ongelma pyritään kiertämään sallimalla siirtymät matalamman energian konfiguraatioista korkeamman energian konfiguraatioihin tietyllä todennäköisyydellä.

*Simuloitu jäähtyys*² on (diskreetti) kombinatorinen optimointimenetelmä, jolla on pyritty ratkaisemaan erityisesti luokkaan *NP-vaikea* kuuluvia kombinatorisia ongelmia. Sen kehittivät S. Kirkpatrick, C. Gelatt ja M. P. Vecchi [18] (1983) sekä riippumattomasti V. Cerný [4] (1985). Molemmissa artikkeleissa tätä menetelmää sovellettiin kuuluisaan kauppatkustajan ongelmaan, jonka ideana on löytää lyhin mahdollinen reitti ennalta määrätyn määrän kaupunkia läpi kulkemiseen siten, että jokaisessa kaupungissa vierailaan täsmälleen yhden kerran.

Idea simuloidun jäähtyksen takana piilee yhteydessä, joka havaittiin kombinatorisen ongelman ratkaisun ja kiinteän aineen hehkuttamisen (engl. *anneal*) välillä. Statistisessa fysiikassa tämä kuvaa prosessia, jossa kiinteää kappaletta kuumennetaan siihen asti, että kaikki kappaleen hiukkaset järjestäytyvät satunnaisesti nestemäiseen olomuotoon, jonka jälkeen ainetta ryhdytään hitaasti jäähtyttämään. Mikäli alkulämpötila on riittävän korkea ja mikäli jäähtyys toteutetaan riittävän hitaasti, hiukkaset järjestäytyvät aineelle ominaista hilarakennetta vastaavaan matalimman energian tilaan. Jos jäähtyys on sen sijaan liian nopeaa, eikä kappale ehdi termodynaamista tasapainoa jokaista lämpötilan arvoa kohden, aineeseen voi ”jäätää” epävakaita amorfisita rakenteita matalaenergisestä kiteisestä hilarakenteesta sijaan. [19, s. 7–8].

¹Yleisesti simuloidusta jäähtyksestä puhuttaessa konfiguraatioavaruudeksi kelpuutetaan mikä tahansa äärellinen joukko (vrt. Luku 2).

²Termin *simulated annealing* lisäksi muita käytettyjä englanninkielisiä nimityksiä ovat mm. *Monte Carlo annealing*, *statistical cooling*, *probabilistic hill climbing*, *stochastic relaxation* ja *probabilistic exchange algorithm* [19, s. 5]. Suomenkielisistä käytössä olevista nimityksistä mainittakoot *simuloitu hehkutus*, *simuloitu melloitus* ja *simuloitu annelointi*.



KUVA 5.1. (a) Hyväksymisen todennäköisyys energian muutoksen $\Delta(x, y) = U(y) - U(x)$ funktiona, kun lämpötila $T > 0$ on kiinnitetty (kuvassa $T = 5$). (b) Hyväksymisen todennäköisyys lämpötilan $T > 0$ funktiona siirryttäessä korkeamman energian tilaan (kuvassa $\Delta(x, y) = 1$).

Simuloidun jäähtyksen alkuperäisessä muodossa siirtymät konfiguraatioiden välillä suoritettiin käyttämällä MCMC-menetelmiin kuuluvaa *Metropolis-algoritmia*, jota käsitellään tarkemmin Luvussa 6. Metropolis-algoritmin perusajatus on muodostaa (tietyt suppenemisominaisuudet omaava) Markovin ketju, jossa ketjun uudeksi tilaksi arvottu konfiguraatio, *ehdotus*, hyväksytään tietyllä todennäköisyydellä. Simuloidussa jäähtyksessä tätä ideaa hyödynnetään hyväksymällä siirtymä matalamman energian konfiguraatioista $x \in E$ korkeamman energian konfiguraatioon $y \in E$ todennäköisyydellä $\exp(-\frac{U(y)-U(x)}{T(n)})$ lämpötilaparametrin $T(n)$ riippuessa ajanhetkestä $n \geq 1$. Edellä oleva eksponenttitermi syntyy, kun lasketaan osamäärä $\pi_{T(n)}(y)/\pi_{T(n)}(x)$, missä $\pi_{T(n)}$ on Gibbsin jakauma joukolla E . Eräs syy Gibbsin jakauman käyttöön piilee siinä, että lämpötilan laskiessa minimienergiakonfiguraatioiden todennäköisyydet kasvavat, ja rajalle mentäessä pätee

$$\pi_{T(n)} \xrightarrow{T(n) \rightarrow 0} \pi_0,$$

missä π_0 on tasajakauma joukon E minimienergiakonfiguraatioiden osajoukolla. Jos nyt *jäähtytysaikataulu* $(T(n))_{n \geq 1}$ vähenee kohti nollaa riittävän hitaasti, voidaan osoittaa, että simuloitua jäähtytystä vastaava epähomogeeninen Markovin ketju supenee kohti minimienergiatilojen joukkoa. Kuvassa 5.1 on esitetty erikseen sekä energian muutoksen että lämpötilaparametrin vaikutus hyväksymisen todennäköisyyden suuruuteen.

Ensimmäinen simuloidun jäähtyksen suppenemista koskeva teoreettinen tulos on peräisin S. ja D. Gemanin jo Luvussa 4 mainitusta artikkelista [8]. Siinä simuloidun jäähtyksen perustana on käytetty Metropolis-algoritmin sijaan Gibbs-otantaa. Käytämmekin tästä lähtien tilanteesta riippuen epävirallisia nimityksiä *Gibbs-jäähtytys* ja *Metropolis-jäähtytys* sen mukaan, kumpi siirtymämenetelmä kulloinkin on kyseessä.

Ennen Gibbs-jäähtyksen perehtymistä todistetaan kaksi tarvittavaa Gibbsin jakauman lämpötilaparametriin T liittyvää tulosta. Otetaan niitä ja myöhempiä tässä luvussa esiintyviä tuloksia varten käyttöön seuraavat merkinnät:

$$U_{\min} := \min_{x \in E} U(x) \quad \text{ja} \quad E_0 := \{x \in E : U(x) = U_{\min}\}.$$

LAUSE 5.1.1. *Olkoon E äärellinen joukko. Tällöin kuvauksille $T \mapsto \pi_T(x)$, missä π_T on Gibbsin jakauma joukolla E ,*

$$\pi_T(x) \xrightarrow{T \rightarrow 0} \pi_0(x) := \frac{\mathbf{1}_{E_0}(x)}{|E_0|} \quad \text{ja} \quad \pi_T(x) \xrightarrow{T \rightarrow \infty} \frac{1}{|E|} \quad \text{kaikilla } x \in E.$$

TODISTUS. Tehdään pieni trikki, joka yksinkertaistaa laskuja. Olkoon $H : E \rightarrow \mathbb{R}$ kuvaus, jolle $H(x) = U(x) - U_{\min}$. Tällöin $H(x) \geq 0$ kaikilla $x \in E$, ja $H(x) = 0$ täsmälleen silloin, kun $x \in E_0$. Lisäksi koska

$$\pi_T(x) = \frac{\exp\left(-\frac{U(x)}{T}\right)}{\sum_{y \in E} \exp\left(-\frac{U(y)}{T}\right)} = \frac{\exp\left(\frac{U_{\min}}{T}\right) \exp\left(-\frac{U(x)}{T}\right)}{\exp\left(\frac{U_{\min}}{T}\right) \sum_{y \in E} \exp\left(-\frac{U(y)}{T}\right)} = \frac{\exp\left(-\frac{H(x)}{T}\right)}{\sum_{y \in E} \exp\left(-\frac{H(y)}{T}\right)},$$

kaikki kuvaukseen $T \rightarrow \pi_T(x)$ (kiinnitetyle $x \in E$) liittyvät suppenemistulokset, jotka saadaan käyttämällä funktion U tilalla funktiota H , pätevät tietysti myös funktion U tapauksessa.

Olkoon aluksi $x \in E_0$. Tällöin

$$\begin{aligned} \pi_T(x) &= \frac{\exp\left(-\frac{H(x)}{T}\right)}{\sum_{y \in E} \exp\left(-\frac{H(y)}{T}\right)} = \frac{\exp\left(-\frac{H(x)}{T}\right)}{\sum_{\substack{y \in E \\ H(y)=0}} \exp\left(-\frac{H(y)}{T}\right) + \sum_{\substack{y \in E \\ H(y)>0}} \exp\left(-\frac{H(y)}{T}\right)} \\ &= \frac{1}{|E_0| + \sum_{\substack{y \in E \\ H(y)>0}} \exp\left(-\frac{H(y)}{T}\right)} \xrightarrow{T \rightarrow 0} \frac{1}{|E_0|}, \end{aligned}$$

sillä $\exp\left(-\frac{H(y)}{T}\right) \xrightarrow{T \rightarrow 0} 0$, kun $H(y) > 0$.

Jos $E \setminus E_0 = \emptyset$, asia selvä. Oletetaan sitten, että $E \setminus E_0 \neq \emptyset$, ja olkoon $x \in E \setminus E_0$. Tällöin

$$\pi_T(x) = \frac{\exp\left(-\frac{H(x)}{T}\right)}{\sum_{y \in E} \exp\left(-\frac{H(y)}{T}\right)} = \frac{\exp\left(-\frac{H(x)}{T}\right)}{|E_0| + \sum_{\substack{y \in E \\ H(y)>0}} \exp\left(-\frac{H(y)}{T}\right)} \xrightarrow{T \rightarrow 0} 0,$$

sillä rajankäyntiä edeltävän lausekkeen yläkerta suppenee kohti nollaa, ja alakerta kohti lukua $|E_0|$, kun $T \rightarrow 0$.

Osoitetaan vielä, että $\pi_T(x) \xrightarrow{T \rightarrow \infty} 1/|E|$ kaikilla $x \in E$. Jos $x \in E_0$, niin $\exp\left(-\frac{H(x)}{T}\right) = 1$ kaikilla $T > 0$, ja jos taas $x \in E \setminus E_0$, niin $\exp\left(-\frac{H(x)}{T}\right) \xrightarrow{T \rightarrow \infty} 1$. Tämä riittää todistamaan väitteen. \square

LEMMA 5.1.2. *Olkoon kaikilla $T \in (0, \infty)$ $\pi_T \in \mathcal{D}_E$ Gibbsin jakauma äärellisellä joukolla E . Tällöin kuvaus $T \mapsto \pi_T(x)$ on*

- (a) *aidosti vähenevä, kun $x \in E_0$*
- (b) *aidosti kasvava välillä $(0, \varepsilon(x))$ riittävän pienelle $\varepsilon(x) > 0$, kun $x \in E \setminus E_0$.*

TODISTUS. Olkoon $T > 0$. Merkitään jälleen $H(x) = U(x) - U_{\min}$. Pidetään $x \in E$ kiinnitettyinä ja derivoidaan $\pi_T(x)$ muuttujan T suhteen:

$$\begin{aligned} \frac{d}{dT} \pi_T(x) &= \frac{d}{dT} \left(\frac{\exp\left(-\frac{H(x)}{T}\right)}{\sum_{y \in E} \exp\left(-\frac{H(y)}{T}\right)} \right) \\ &= \frac{\frac{H(x)}{T^2} \exp\left(-\frac{H(x)}{T}\right) \sum_{y \in E} \exp\left(-\frac{H(y)}{T}\right) - \exp\left(-\frac{H(x)}{T}\right) \sum_{y \in E} \frac{H(y)}{T^2} \exp\left(-\frac{H(y)}{T}\right)}{\left(\sum_{y \in E} \exp\left(-\frac{H(y)}{T}\right)\right)^2} \\ &= \frac{H(x) \sum_{y \in E} \exp\left(-\frac{H(y)}{T}\right) - \sum_{y \in E} H(y) \exp\left(-\frac{H(y)}{T}\right)}{T^2 \exp\left(\frac{H(x)}{T}\right) \left(\sum_{y \in E} \exp\left(-\frac{H(y)}{T}\right)\right)^2}. \end{aligned}$$

Jos $x \in E_0$, niin $H(x) = 0$ ja $\exp\left(-\frac{H(x)}{T}\right) = 1$, joten

$$\frac{d}{dT} \pi_T(x) = -\frac{\sum_{y \in E} \exp\left(-\frac{H(y)}{T}\right)}{T^2 \left(\sum_{y \in E} \exp\left(-\frac{H(y)}{T}\right)\right)^2} < 0,$$

ja $\pi_T(x)$ on siten aidosti vähenevä kaikilla $T > 0$.

Jos $E \setminus E_0 = \emptyset$, todistus on valmis. Oletetaan, että $E \setminus E_0 \neq \emptyset$, ja olkoon $x \in E \setminus E_0$. Merkitään $\tilde{H}_{\max} = \max\{U(y) : y \in E \setminus E_0\}$ ja $\tilde{H}_{\min} = \min\{U(y) : y \in E \setminus E_0\}$, jolloin $\tilde{H}_{\max}, \tilde{H}_{\min} > 0$. Valitaan $T > 0$ niin pieneksi, että

$$(5.1) \quad e^{-\tilde{H}_{\min}/T} < \frac{H(x)|E_0|}{\tilde{H}_{\max}|E \setminus E_0|}.$$

Tällöin

$$\begin{aligned} &H(x) \sum_{y \in E} \exp\left(-\frac{H(y)}{T}\right) - \sum_{y \in E} H(y) \exp\left(-\frac{H(y)}{T}\right) \\ &= H(x) \left(|E_0| + \sum_{y \in E \setminus E_0} \exp\left(-\frac{H(y)}{T}\right) \right) - \sum_{y \in E \setminus E_0} H(y) \exp\left(-\frac{H(y)}{T}\right) \\ &\geq H(x)|E_0| - \tilde{H}_{\max}|E \setminus E_0| \exp\left(-\frac{\tilde{H}_{\min}}{T}\right) \\ &> H(x)|E_0| - \tilde{H}_{\max}|E \setminus E_0| \left(\frac{H(x)|E_0|}{\tilde{H}_{\max}|E \setminus E_0|} \right) = 0. \end{aligned}$$

Derivaatta on siis aidosti positiivinen ainakin ehdon (5.1) ollessa voimassa:

$$\frac{d}{dT} \pi_T(x) = \frac{H(x) \sum_{y \in E} \exp\left(-\frac{H(y)}{T}\right) - \sum_{y \in E} H(y) \exp\left(-\frac{H(y)}{T}\right)}{T^2 \exp\left(\frac{H(x)}{T}\right) \left(\sum_{y \in E} \exp\left(-\frac{H(y)}{T}\right)\right)^2} > 0,$$

sillä selvästi $g(x, T) := T^2 \exp\left(\frac{H(x)}{T}\right) \left(\sum_{y \in E} \exp\left(-\frac{H(y)}{T}\right)\right)^2 > 0$ kaikilla $x \in E \setminus E_0$ ja $T > 0$. Siten $\pi_T(x)$ on aidosti kasvava riittävän pienille $T > 0$. \square

SEURAUUS 5.1.3. *Olkoon $(T(n))_{n \geq 1} \subset (0, \infty)$ vähenevä jono, jolle $T(n) \xrightarrow{n \rightarrow \infty} 0$. Tällöin Gibbsin jakaumien jono $(\pi_{T(1)}, \pi_{T(2)}, \dots) \subset \mathcal{D}_E$ toteuttaa ehdon*

$$\sum_{n=1}^{\infty} \|\pi_{T(n)} - \pi_{T(n+1)}\| < \infty.$$

TODISTUS. Merkitään $\mu_n = \pi_{T(n)}$ kaikilla $n \geq 1$. Koska Lemman 5.1.2 nojalla jono $(\mu_n)_{n \geq 1}$ on nyt joko (aidosti) kasvava tai (aidosti) vähenevä kun $T(n)$ on riittävän pieni – ja n siten yhtäpitävästi riittävän suuri – niin Lemman 1.3.9 nojalla

$$\sum_{n=1}^{\infty} \|\pi_{T(n)} - \pi_{T(n+1)}\| = \sum_{n=1}^{\infty} \|\mu_n - \mu_{n+1}\| < \infty.$$

□

5.2. Gibbs-jäähdytyksen algoritmi

Kiinnitetään jälleen tämän luvun ajaksi suuntaamaton verkko $\mathcal{G} = (S, \mathcal{N})$, konfiguraatioavaruus $E = \prod_{s \in S} \Lambda_s$, naapuripotentiali $\mathcal{V}_{\mathcal{N}}$ ja E -arvoinen Gibbsin kenttä X kuten Luvun 4 alussa. Yhtä askelta vastaava Gibbs-jäähdytyksen algoritmi on seuraava:

- olkoon $X_{n-1} = x \in E$.
- Poimitaan uusi faasi $\lambda \in \Lambda_{\sigma_n}$ ehdollisesta jakaumasta $\pi_{T(n)}(\cdot \mid x_{S \setminus \{\sigma_n\}})$ ja asetetaan $X_n = y := (\lambda, x_{S \setminus \{\sigma_n\}})$.

Gibbs-jäähdytyksen algoritmi eroaa siis Gibbs-otannan algoritmista ainoastaan siinä, että lämpötilaparametri $T = T(n)$ riippuu nyt ajanhetkestä. Olkoon jälleen

$$\Delta_s := \max_{x, y \in E} \{|U(x) - U(y)| : x \stackrel{s}{\sim} y\} \text{ kaikilla } s \in S \text{ ja } \Delta := \max_{s \in S} \Delta_s.$$

Huomataan, että koska Lemman 4.4.3 nojalla

$$\delta_T = \min_{\substack{s \in S \\ x \in E}} \pi_T(x_s \mid x_{S \setminus \{s\}}) \geq \frac{1}{|E|} \exp\left(-\frac{\Delta}{T}\right), \quad \text{ja koska } \frac{1}{|E|} \exp\left(-\frac{\Delta}{T}\right) > 0$$

kaikilla $T > 0, \Delta < \infty$, aidosti positiivisten Gibbsin jakaumien $\{\pi_T : T > 0\}$ ehdollisiin jakaumiin $\pi_T(\cdot \mid x_{S \setminus \{s\}})$ perustuva Gibbs-jäähdytyksen algoritmi kykenee siten tekemään myös simuloitua jäähdytyksen kannalta välttämättömiä, energiaa *kasvatavia* siirtymiä.

Kiinnitetään seuraavaksi niin sanottu *jäähdytysaikataulu* $(T(1), T(2), \dots)$: olkoon $(T(n))_{n \geq 1} \subset (0, \infty)$ vähenevä jono, jolle

$$(5.2) \quad \begin{aligned} & \bullet \quad T(n) \xrightarrow{n \rightarrow \infty} 0 \quad \text{ja} \\ & \bullet \quad T(n) \geq \frac{N\Delta}{\log n} \text{ kaikilla } n \geq n_0 \text{ jostain luvusta } n_0 \geq 2 \text{ lähtien.} \end{aligned}$$

5.3. Deterministinen päivitysjono

Olkoon $(\sigma_n)_{n \geq 1} \subset S$ sellainen deterministinen päivitysjono, jolle on olemassa kokonaisluku $\tau \geq N$ siten, että

$$(5.3) \quad \{\sigma_{(k-1)\tau+1}, \sigma_{(k-1)\tau+2}, \dots, \sigma_{k\tau}\} = S \quad \text{kaikilla } k = 1, 2, \dots$$

Tämä oletus on vahvempi kuin Luvun 4 deterministiseltä päivitysjonolta vaadittu ehto $\#\{n \geq 1 : \sigma_n = s\} = \infty$ kaikilla $s \in S$. Voidaan lisäksi osoittaa, että se on yhtäpitävä Gemanien tuloksen [8, Theorem B] oletuksen kanssa, jossa vaaditaan, että on olemassa luku $\tilde{\tau} \geq N$ siten, että $\{\sigma_{i+1}, \sigma_{i+2}, \dots, \sigma_{i+\tilde{\tau}}\} = S$ kaikilla $i \geq 1$: tämä ehto selvästi implikoi ehdon (5.3), ja toisaalta toista implikaatiota varten riittää valita esimerkiksi $\tilde{\tau} = 2\tau$. Yhtä kaikki, ehto (5.3) kattaa jaksollisen päivityksen tapauksen, joka on käytännön kannalta periaatteessa ainoa kiinnostava deterministinen päivitysmenetelmä.

Määritellään Gibbs-jäähdytystä vastaava jono siirtymämatriiseja $(P_n)_{n \geq 1} \subset \mathcal{M}_{E \times E}$ asettamalla

$$(5.4) \quad P_n(x, y) := \Gamma_{T(n), \sigma_n}(x, y) = \begin{cases} \pi_{T(n)}(y_{\sigma_n} \mid x_{S \setminus \{\sigma_n\}}), & x \stackrel{\sigma_n}{\sim} y \\ 0, & \text{muuten.} \end{cases}$$

Olkoon sitten $\{s_1, s_2, \dots, s_N\}$ jokin solmujen joukon S numerointi. Määritellään ajanhetket $T_k := k\tau$ kaikille $k = 0, 1, \dots$, jolloin päivitysjonon $(\sigma_n)_{n \geq 1}$ oletus (5.3) voidaan kirjoittaa myös muodossa $\{\sigma_{T_{k-1}+1}, \dots, \sigma_{T_k}\} = S$ kaikilla $k \geq 1$. Jaetaan jälleen matriisitulo $P_1 P_2 \dots$ yhtä täyttä päivitystä vastaaviin blokkeihin Q_1, Q_2, \dots asettamalla

$$Q_k := P_{T_{k-1}+1} \dots P_{T_k} \quad \text{kaikilla } k \geq 1.$$

LEMMA 5.3.1. *On olemassa luku $C \in (0, 1)$ siten, että*

$$\min_{x, y} Q_k(x, y) \geq \frac{C}{|E|^{Nk\tau}} \quad \text{kaikilla } k \geq 1.$$

TODISTUS. Todistus menee suurilta osin samaan tapaan kuin Lemman 4.4.4 todistus. Kiinnitetään $k \geq 1$ ja määritellään jälleen kullekin solmulle viimeiset päivityshetket aikavälillä $(T_{k-1}, T_k]$ asettamalla

$$m_0 = m_0(k) := T_{k-1} \quad \text{ja} \quad m_j = m_j(k) := \max \{n \leq T_k : \sigma_n = s_j\}$$

kaikilla $j = 1, \dots, N$ siten, että $T_{k-1} = m_0 < m_1 < \dots < m_N = T_k$. Tarvittaessa siis numeroidaan solmujen joukko S uudestaan siten, että edellä oleva järjestys on voimassa. Asetetaan jälleen (kuten Lemman 4.4.4 todistuksessa)

$$E^{(0)} := E, \quad E^{(j)} := \{z \in E : z_{s_k} = y_{s_k}, k = 1, 2, \dots, j\} \quad \text{kaikilla } j = 1, \dots, N,$$

jolloin $E = E^{(0)} \supset E^{(1)} \supset \dots \supset E^{(N-1)} \supset E^{(N)} = \{y\}$. Kirjoittamalla

$$\begin{aligned} Q_k(x, y) &= P_{T_{k-1}+1} \dots P_{T_k}(x, y) = P_{m_0+1} \dots P_{m_N}(x, y) \\ &= \sum_{\substack{z_j \in E^{(j)} \\ 1 \leq j \leq N}} P_{m_0+1} \dots P_{m_1}(x, z_1) P_{m_1+1} \dots P_{m_2}(z_1, z_2) \dots P_{m_{N-1}+1} \dots P_{m_N}(z_{N-1}, z_N) \\ &= \sum_{\substack{z_j \in E^{(j)} \\ 1 \leq j \leq N}} R_0(x, z_1) R_1(z_1, z_2) \dots R_{N-1}(z_{N-1}, z_N) \end{aligned}$$

ja etenemällä kuten Lemman 4.4.4 aputuloksen todistuksessa, voidaan osoittaa, että

$$\sum_{v \in E^{(j+1)}} R_j(w, v) \geq \delta_{T(m_{j+1})} = \min_{\substack{s \in S \\ x \in E}} \pi_{T(m_{j+1})}(x_s \mid x_{S \setminus \{s\}}), \quad \text{kun } w \in E^{(j)}.$$

Siten

$$Q_k(x, y) = \sum_{\substack{z_j \in E^{(j)} \\ 1 \leq j \leq N}} R_0(x, z_1) R_1(z_1, z_2) \cdots R_{N-1}(z_{N-1}, z_N) \geq \prod_{j=1}^N \delta_{T(m_j)}.$$

Lemman 4.4.3 nojalla $\delta_T \geq \frac{1}{|E|} \exp\left(-\frac{\Delta}{T}\right)$ kaikilla $T > 0$, joten edelleen

$$(5.5) \quad Q_k(x, y) \geq \prod_{j=1}^N \delta_{T(m_j)} \geq \prod_{j=1}^N \frac{1}{|E|} \exp\left(-\frac{\Delta}{T(m_j)}\right).$$

Koska $m_j \leq T_k = k\tau$ kaikilla $0 \leq j \leq N$ ja siten $T(m_j) \geq T(k\tau)$, ja koska lisäksi funktio $T \mapsto \exp\left(-\frac{\Delta}{T}\right)$ on kasvava, niin

$$(5.6) \quad \prod_{j=1}^N \frac{1}{|E|} \exp\left(-\frac{\Delta}{T(m_j)}\right) \geq |E|^{-N} \exp\left(-\frac{N\Delta}{T(k\tau)}\right).$$

Ollaan siten todistettu, että epäyhtälö (5.6) on voimassa kaikilla $k \geq 1$. Jos nyt $k\tau \geq n_0$, niin jäädytysaikataulua koskevasta oletuksesta (5.2) seuraa, että

$$(5.7) \quad T(k\tau) \geq \frac{N\Delta}{\log(k\tau)} \iff \log(k\tau) \geq \frac{N\Delta}{T(k\tau)},$$

jolloin epäyhtälöt (5.5), (5.6) ja (5.7) yhdistämällä saadaan

$$Q_k(x, y) \geq |E|^{-N} \exp\left(-\frac{N\Delta}{T(k\tau)}\right) \geq \frac{1}{|E|^N k\tau}.$$

Entä jos $k\tau < n_0$? Merkitään

$$C := \min_{\substack{x, y \in E \\ k \geq 1}} \{Q_k(x, y) : k\tau < n_0\}.$$

Rajoitus $k\tau < n_0$ takaa, että minimi otetaan yli äärellisen joukon. Yhdessä epäyhtälön (5.5) kanssa tämä implikoi, että $C \in (0, 1)$. Siten mielivaltaiselle $k \geq 1$ ja kaikille $x, y \in E$ saadaan voimaan epäyhtälö

$$Q_k(x, y) \geq \min\left\{C, \frac{1}{|E|^N k\tau}\right\} \geq \frac{C}{|E|^N k\tau}.$$

□

SEURAUUS 5.3.2.

$$c(P_i \cdots P_n) \xrightarrow{n \rightarrow \infty} 0 \quad \text{kaikilla } i \geq 1.$$

TODISTUS. Olkoon C kuten Lemman 5.3.1 todistuksessa. Lemmojen 3.4.6 ja 5.3.1 nojalla tällöin

$$c(Q_k) \leq 1 - \min_{x, y \in E} Q_k(x, y) \leq 1 - \frac{C}{|E|^N k\tau} \quad \text{kaikilla } k \geq 1.$$

Väite seuraa käyttämällä Lemmaa 3.4.12, sillä

$$\sum_{k=1}^{\infty} (1 - c(Q_k)) \geq \sum_{k=1}^{\infty} \frac{C}{|E|^N k\tau} = \frac{C}{|E|^N \tau} \sum_{k=1}^{\infty} \frac{1}{k} = \infty.$$

□

Kootaan nyt kaikki edellinen deterministisen päivitysjonon Gibbs-jäähtyksen suppenemislauseeksi.

LAUSE 5.3.3 (Gibbs-jäähtyminen, deterministinen päivitysjono). *Olkoon $E = \prod_{s \in S} \Lambda_s$ äärellinen konfiguraatioavaruus ja olkoon $(P_n)_{n \geq 1} \subset \mathcal{M}_{E \times E}$ epähomogeenisen Markovin ketjun siirtymämatriisit määritelty kaavalla*

$$P_n(x, y) = \begin{cases} \pi_{T(n)}(y_{\sigma_n} \mid x_{S \setminus \{\sigma_n\}}), & \text{jos } x \stackrel{\sigma_n}{\sim} y \\ 0, & \text{muuten} \end{cases},$$

missä deterministiselle päivitysjonolle $(\sigma_n)_{n \geq 1} \subset S$ ja jäähtytysaikataululle $(T(n))_{n \geq 1}$ on voimassa seuraavat oletukset:

- on olemassa luku $\tau \geq 1$ siten, että $\{\sigma_{i+1}, \sigma_{i+2}, \dots, \sigma_{i+\tau}\} = S$ kaikilla $i \geq 1$.
- $(T(n))_{n \geq 1} \subset (0, \infty)$ on vähenevä jono, jolle $T(n) \xrightarrow{n \rightarrow \infty} 0$.
- $T(n) \geq \frac{N\Delta}{\log n}$ kaikilla $n \geq n_0$ jostain $n_0 \geq 2$ lähtien.

Tällöin ketju on vahvasti ergodinen, toisin sanoen

$$\nu P_i \cdots P_n \xrightarrow{TV} \pi_0 \quad \text{kaikilla } \nu \in \mathcal{D}_E \text{ ja } i \geq 1,$$

missä π_0 on tasajakauma tila-avaruuden E minienergiatilojen osajoukossa E_0 .

TODISTUS. Todistus perustuu Dobrushinin lauseeseen (Lause 3.5.2). Todetaan aluksi, että kaikilla $n \geq 1$ siirtymämatriisi P_n on kääntyvä jakauman $\pi_{T(n)}$ suhteen; tämä seuraa suoraan Lauseesta 4.3.3. Lause 5.1.1, Lemma 5.1.2 ja Seuraus 5.1.3 yhdessä takaavat, että invarianttien jakaumien $(\pi_{T(n)})_{n \geq 1}$ summautumisehto (3.15) toteutuu, ja että $\lim_{n \rightarrow \infty} \pi_{T(n)} = \pi_0$. Koska lisäksi $c(P_i \cdots P_n) \xrightarrow{n \rightarrow \infty} 0$ kaikilla $i \geq 1$ Seurauksen 5.3.2 nojalla, Lauseen 3.5.2 oletukset ovat siten voimassa, ja siten kaikilla $\nu \in \mathcal{D}_E$ ja $i \geq 1$ pätee

$$\nu P_i \cdots P_n \xrightarrow{TV} \pi_0.$$

□

5.4. Satunnainen päivitysjono

Olkoon sitten $\psi \in \mathcal{D}_S$ aidosti positiivinen jakauma solmujen joukolla S , ja yhden solmun päivitystä vastaava siirtymämatriisi muotoa

$$\tilde{P}_n(x, y) := \sum_{s \in S} \psi(s) \Gamma_{T(n), s}(x, y).$$

Olkoon lisäksi varsinaisen jäähtytysketjun siirtymämatriisi muotoa

$$P_n(x, y) := \tilde{P}_n^N(x, y) = \underbrace{\tilde{P}_n \cdots \tilde{P}_n}_{N \text{ kpl}}(x, y).$$

Merkitään $\gamma := \min_{s \in S} \psi(s) > 0$, joka on siis pienin jakaumaan ψ liittyvä todennäköisyys.

LEMMA 5.4.1.

$$c(P_i \cdots P_n) \xrightarrow{n \rightarrow \infty} 0 \text{ kaikilla } i \geq 1.$$

TODISTUS. Olkoon $n \geq 1$. Kiinnitetään x ja $y \in E$ ja määritellään konfiguraatiot $z^{(1)}, \dots, z^{(N-1)} \in E$ jollekin solmujen joukon numeroinnille $S = \{s_1, \dots, s_N\}$ kuten Lemman 4.5.1 todistuksessa, jolloin saadaan epäyhtälö

$$(5.8) \quad P_n(x, y) = \tilde{P}_n^N(x, y) \geq \prod_{i=1}^N \psi(s_i) \pi_{T(n)}(z_{s_i}^{(i)} \mid z_{S \setminus \{s_i\}}^{(i-1)}).$$

Epäyhtälöitä $\delta_{T(n)} \geq \frac{1}{|E|} \exp\left(-\frac{\Delta}{T(n)}\right)$ ja (5.8) soveltamalla saadaan siten

$$\begin{aligned} \min_{\substack{x, y \in E \\ n \geq n_0}} P_n(x, y) &= \min_{\substack{x, y \in E \\ n \geq n_0}} \tilde{P}_n^N(x, y) \geq \prod_{i=1}^N \gamma \delta_{T(n)} = (\gamma \delta_{T(n)})^N \\ &\geq \frac{\gamma^N}{|E|^N} \exp\left(-\frac{N\Delta}{T(n)}\right) \geq \frac{\gamma^N}{n|E|^N}, \end{aligned}$$

sillä oletuksen (5.2) nojalla $T(n) \geq \frac{N\Delta}{\log n}$, kun $n \geq n_0$. Olkoon sitten

$$C := \min_{\substack{x, y \in E \\ 1 \leq n < n_0}} P_n(x, y) = \min_{\substack{x, y \in E \\ 1 \leq n < n_0}} \tilde{P}_n^N(x, y) \in (0, 1),$$

jolloin kaikille $n \geq 1$ saadaan arvio

$$\min_{x, y \in E} P_n(x, y) \geq \min \left\{ C, \frac{\gamma^N}{n|E|^N} \right\} \geq \frac{\gamma^N C}{n|E|^N}.$$

Käyttämällä nyt Lemmaa 3.4.6 saadaan kontraktiokertoimille alaraja

$$(5.9) \quad c(P_n) \leq 1 - \min_{x, y \in E} P_n(x, y) \leq 1 - \frac{\gamma^N C}{n|E|^N}.$$

Väite seuraa jälleen käyttämällä Lemmaa 3.4.12, sillä valitsemalla blokeiksi $Q_k := P_k$ kaikilla $k \geq 1$ ja käyttämällä epäyhtälöä (5.9) saadaan

$$\sum_{k=1}^{\infty} (1 - c(Q_k)) = \sum_{k=1}^{\infty} (1 - c(P_k)) \geq \frac{\gamma^N C}{|E|^N} \sum_{k=1}^{\infty} \frac{1}{k} = \infty.$$

□

LAUSE 5.4.2. *Olkoon $E = \prod_{s \in S} \Lambda_s$ äärellinen konfiguraatioavaruus ja olkoot epähomogeenisen Markovin ketjun siirtymämatriisit $(P_n)_{n \geq 1} \subset \mathcal{M}_{E \times E}$ määritelty kaavalla $P_n(x, y) = \tilde{P}_n^N(x, y)$, missä*

$$\tilde{P}_n^N(x, y) = \sum_{s \in S} \psi(s) \Gamma_{T(n), s}(x, y) \begin{cases} \sum_{s \in S} \psi(s) \pi_{T(n)}(x_s \mid x_{S \setminus \{s\}}), & x = y \\ \psi(s) \pi_{T(n)}(y_s \mid x_{S \setminus \{s\}}), & x \neq y, x \stackrel{s}{\sim} y \\ 0, & \text{muuten} \end{cases}.$$

Oletetaan lisäksi, että $\min_{s \in S} \psi(s) > 0$ ja että jäähdytysaikataululle $(T(n))_{n \geq 1} \subset (0, \infty)$ on voimassa seuraavat oletukset:

- $(T(n))_{n \geq 1}$ on vähenevä jono, jolle $T(n) \xrightarrow{n \rightarrow \infty} 0$.
- $T(n) \geq \frac{N\Delta}{\log n}$ kaikilla $n \geq n_0$ jostain $n_0 \geq 2$ lähtien.

Tällöin ketju on vahvasti ergodinen, toisin sanoen

$$\nu P_i \cdots P_n \xrightarrow{TV} \pi_0 \quad \text{kaikilla } \nu \in \mathcal{D}_E \text{ ja } i \geq 1,$$

missä π_0 on tasajakauma tila-avaruuden E minienergiatilojen osajoukossa E_0 .

TODISTUS. Osoitetaan aluksi, että $\pi_{T(n)} P_n = \pi_{T(n)}$ kaikilla $n \geq 1$. Olkoon $n \geq 1$. Lauseen nojalla siirtymämatriisi \tilde{P}_n ja Gibbsin jakauma $\pi_{T(n)}$ toteuttavat yksityiskohdaiset tasapainoyhtälöt, joten Lemman 4.5.2 nojalla $\pi_{T(n)}$ on siten siirtymämatriisin \tilde{P}_n invariantti jakauma, joten myös

$$\begin{aligned} \pi_{T(n)} P_n &= \pi_{T(n)} \underbrace{\tilde{P}_n \cdots \tilde{P}_n}_{N \text{ kpl}} = \left(\pi_{T(n)} \tilde{P}_n \right) \underbrace{\tilde{P}_n \cdots \tilde{P}_n}_{N-1 \text{ kpl}} = \pi_{T(n)} \underbrace{\tilde{P}_n \cdots \tilde{P}_n}_{N-1 \text{ kpl}} = \dots \\ &= \pi_{T(n)} \tilde{P}_n = \pi_{T(n)}. \end{aligned}$$

Koska edelleen $\sum_{n=1}^{\infty} \|\pi_{T(n)} - \pi_{T(n+1)}\| < \infty$ ja $c(P_i \cdots P_n) \xrightarrow{n \rightarrow \infty} 0$ kaikilla $i \geq 1$ Seurauksen 5.1.3 ja Lemman 5.4.1 nojalla, Lauseen 3.5.2 oletukset ovat nyt voimassa. Siten kaikilla alkujakaumilla $\nu \in \mathcal{D}_E$ ja kaikilla $i \geq 1$

$$\nu P_i \cdots P_n \xrightarrow{TV} \pi_0.$$

□

Metropolis-dynamiikalla varustettu simuloitu jäähtytys

Luvussa 7 simuloitua jäähtytystä pyritään soveltamaan tilanteessa, jossa tulomuotoiselle konfiguraatioavaruudelle E on annettu rajoitteita tiettyjä rajoitteita. Tavoitteena on minimoida koko avaruudella E määritelty energiafunktio tietyn *sallittuista* konfiguraatioista muodostuvan ei-tulomuotoisen aidon osajoukon $F \subset E$ suhteen. Luvussa 5 esiteltyä Gibbs-jäähtytystä ei sellaisenaan voida tällaisessa tilanteessa soveltaa. On kuitenkin mahdollista yleistää Gibbs-jäähtytystä koskemaan *ajanhetkestä riippuvia* energiafunktioita [30, Chapter 7]. Sopivasti rakennetun energiafunktion avulla on nimittäin mahdollista saada yleistetty Gibbs-jäähtytyksen ketju suppenemaan joukon F minimienergiakonfiguraatioiden tasajakaumaan. Yksi tapa määrittellä tällainen energiafunktio on asettaa

$$U_n(x) = T(n)^{-1} \left(U(x) - \min_{y \in \tilde{E}} U(y) + \lambda(n)V(x) \right), \quad \text{missä}$$

$(T(n))_{n \geq 1}$ on sopivasti valittu jäähtytysaikataulu, $U : E \rightarrow \mathbb{R}$ on E -arvoisen Gibb-sin kentän energiafunktio, $V(x) = \alpha(x)\mathbf{1}_{E \setminus F}(x)$ on joukon F ulkopuolisten alkioiden energiaa korostava funktio joillekin vakioille $\alpha(x) > 0$ ja $\lambda(n) \xrightarrow{n \rightarrow \infty} \infty$. Menetelmän toimiminen edellyttää, että jäähtytysketjun (X_n) tulee voida aluksi vieraillla myös osajoukon F ulkopuolella ennen kuin ei-toivotut siirtymät kielletylle alueelle $E \setminus F$ saadaan estettyä asteittain kasvavien ”potentiaalimuurien” avulla. Mikäli kuitenkin halutaan, että $\mathbb{P}(X_n \in F) = 1$ kaikilla ajanhetkillä $n \geq 1$, lienee syytä siirtyä käyttämään yleisemmille konfiguraatioavaruuksille soveltuvaa *Metropolis-algoritmiin* perustuvaa simuloitua jäähtytystä.

6.1. Metropolis-algoritmi

Olkoon $E \neq \emptyset$ mielivaltainen äärellinen joukko eli *konfiguraatioavaruus* ja olkoon $\pi \in \mathcal{D}_E$ jokin jakauma, joka tunnetaan normalisoivaa ainoastaan vakiota vaille, ts.

$$\pi(x) = \frac{\tilde{\pi}(x)}{C} = \frac{\tilde{\pi}(x)}{\sum_{y \in E} \tilde{\pi}(y)} \quad \text{kaikilla } x \in E,$$

missä funktio $\tilde{\pi} : E \rightarrow [0, \infty)$ tunnetaan. Halutaan otos jakaumasta π . Kuten Gibbs-otannan yhteydessä todettiin, eräs tapa on konstruoida Markovin ketju (X_0, X_1, \dots) , jonka tasapainojakauma on π . Näin edetään myös Metropolis-algoritmin tapauksessa. Olkoon $\Psi \in \mathcal{M}_{E \times E}$ *symmetrinen* siirtymämatriisi, jonka rivejä vastaavia jakaumia $\Psi(x, \cdot)$ kutsutaan *ehdotusjakaumiksi*.

Metropolis-algoritmi (Metropolis *et al*, 1953 [22]) on seuraava:

- olkoon $X_{n-1} = x \in E$.
- Arvotaan ehdotus $y \in E$ ketjun uudeksi arvoksi jakaumasta $\Psi(x, \cdot)$.
- Lasketaan *hyväksymisen* todennäköisyys

$$\alpha(x, y) := 1 \wedge \frac{\pi(y)}{\pi(x)}.$$

- Jos $\alpha(x, y) = 1$, hyväksytään ehdotus ja asetetaan $X_n = y$.
- Jos puolestaan $\alpha(x, y) < 1$, arvotaan luku u välin $[0, 1]$ tasajakaumasta ja
 - hyväksytään ehdotus, jos $u \leq \alpha(x, y)$
 - hylätään ehdotus ja asetetaan $X_n = X_{n-1}$, jos $u > \alpha(x, y)$.

Algoritmia soveltamalla voidaan muodostaa homogeenista Markovin ketjua vastaava siirtymämatriisi $P \in \mathcal{M}_{E \times E}$,

$$(6.1) \quad P(x, y) = \begin{cases} \Psi(x, y)\alpha(x, y), & x \neq y \\ 1 - \sum_{z \in E \setminus \{x\}} \Psi(x, z)\alpha(x, z), & x = y \end{cases}.$$

Jos P on primitiivinen (tai yhtäpitävästi pelkistymätön ja jaksoton), niin Seurauksen 3.4.7 nojalla

$$\nu P^n \xrightarrow{TV} \pi \quad \text{kaikilla alkujakaumilla } \nu \in \mathcal{D}_E,$$

sillä seuraavan Lemman 6.1.1 nojalla siirtymämatriisin P yksikäsitteinen invariantti jakauma on π . Todetaan vielä, että kuten Gibbs-otannan tapauksessa, Metropolis-otannassa riittää tuntea jakauma π normalisoivaa tekijää vaille, sillä laskettaessa suhteita $\frac{\pi(y)}{\pi(x)}$ vakio C supistuu pois.

LEMMA 6.1.1. *Metropolis-algoritmin siirtymämatriisi P (6.1) on kääntyvä jakauman π suhteen, ts.*

$$\pi(x)P(x, y) = \pi(y)P(y, x) \quad \text{kaikilla } x, y \in E.$$

TODISTUS. Jos $x = y$, niin asia selvä. Olkoon $x \neq y$. Jos $\pi(x) = \pi(y)$, niin $\alpha(x, y) = \alpha(y, x) = 1$, jolloin ehdotusmatriisin Ψ symmetrisyys takaa, että

$$\pi(x)P(x, y) = \pi(x)\Psi(x, y)\alpha(x, y) = \pi(y)\Psi(y, x)\alpha(y, x) = \pi(y)P(y, x).$$

Voidaan sitten yleisyyttä loukkaamatta olettaa, että $\pi(y) < \pi(x)$, jolloin $\alpha(x, y) = \frac{\pi(y)}{\pi(x)}$ ja $\alpha(y, x) = 1$. Nyt

$$\begin{aligned} \pi(x)P(x, y) &= \pi(x)\Psi(x, y)\alpha(x, y) = \pi(x)\Psi(y, x)\frac{\pi(y)}{\pi(x)} \\ &= \pi(y)\Psi(y, x) = \pi(y)\Psi(y, x)\alpha(y, x) = \pi(y)P(y, x). \end{aligned}$$

□

HUOMAUTUS 6.1.2. Lemman 6.1.1 tuloksen paikkansapitävyys ei tietenkään ole sattumaa, sillä Metropolis-algoritmi on rakennettu varta vasten siten, että matriisista P tulee kääntyvä jakauman π suhteen (ks. esimerkiksi [21, s. 38]).

HUOMAUTUS 6.1.3. Metropolis-algoritmi voidaan yleistää tilanteisiin, joissa ehdotusmatriisilta Ψ ei vaadita symmetrisyyttä. Tämän niin kutsutun Metropolis–Hastings -algoritmin (W.K. Hastings, 1970 [12]) siirtymämatriisi voidaan kirjoittaa muodossa

$$(6.2) \quad P(x, y) = \begin{cases} \Psi(x, y)\alpha(x, y), & x \neq y \\ 1 - \sum_{z \in E \setminus \{x\}} \Psi(x, z)\alpha(x, z), & x = y \end{cases},$$

missä

$$(6.3) \quad \alpha(x, y) = \begin{cases} 1 \wedge \frac{\pi(y)\Psi(y, x)}{\pi(x)\Psi(x, y)}, & \pi(x)\Psi(x, y) > 0 \\ 1, & \pi(x)\Psi(x, y) = 0 \end{cases}.$$

Voidaan osoittaa, että myös tällöin P on kääntyvä jakauman π suhteen.

Todetaan vielä seuraava kiintoisa seikka: *Gibbs-otannan* algoritmia voidaan pitää Metropolis–Hastings -algoritmin erikoistapauksena (jossa hyväksymisen todennäköisyys on identtisesti yksi), kuten seuraava aiheesta hieman sivuraiteille menevä mutta yleissivistävä lemma osoittaa.

LEMMA 6.1.4. *Olkoon E äärellinen tulomuotoinen konfiguraatioavaruus ja olkoon π johonkin naapuripotentialiin $\mathcal{V}_{\mathcal{N}}$ liittyvän E -arvoisen Gibbsin kentän jakauma. Jos asetetaan $\Psi(x, y) := \pi(y_s | x_{S \setminus \{s\}}) \mathbb{1}\{x \stackrel{s}{\sim} y\}$ kaikilla $x, y \in E$, on Metropolis–Hastings -algoritmin siirtymämatriisi (6.2) tällöin muotoa*

$$P(x, y) = \Gamma_s(x, y) = \begin{cases} \pi(y_s | x_{S \setminus \{s\}}), & x \stackrel{s}{\sim} y \\ 0, & \text{muuten} \end{cases}.$$

TODISTUS. Olkoot $x, y \in E$, $s \in S$ ja $x \stackrel{s}{\sim} y$. Tällöin

$$\begin{aligned} \frac{\pi(y)\Psi(y, x)}{\pi(x)\Psi(x, y)} &= \frac{\pi(y)}{\pi(x)} \cdot \frac{\pi(x_s, y_{S \setminus \{s\}})}{\pi(y_s, x_{S \setminus \{s\}})} \cdot \frac{\sum_{\lambda \in \Lambda_s} \pi(\lambda, x_{S \setminus \{s\}})}{\sum_{\lambda \in \Lambda_s} \pi(\lambda, y_{S \setminus \{s\}})} \\ &= \frac{\pi(y)}{\pi(x)} \cdot \frac{\pi(x)}{\pi(y)} \cdot \frac{\sum_{\lambda \in \Lambda_s} \pi(\lambda, x_{S \setminus \{s\}})}{\sum_{\lambda \in \Lambda_s} \pi(\lambda, x_{S \setminus \{s\}})} = 1. \end{aligned}$$

Siten $\alpha(x, y) = 1$, kun $x \stackrel{s}{\sim} y$. Jos puolestaan $x_{S \setminus \{s\}} \neq y_{S \setminus \{s\}}$, niin $\mathbb{1}\{x \stackrel{s}{\sim} y\} = 0$, ja siten myös $\pi(x)\Psi(x, y) = 0$. On siis $\alpha(x, y) = 1$ kaikilla $x, y \in E$. Nyt jos $x \neq y$, niin yhtälön (6.2) matriisi P saa muodon

$$P(x, y) = \Psi(x, y)\alpha(x, y) = \pi(y_s | x_{S \setminus \{s\}}) \mathbb{1}\{x \stackrel{s}{\sim} y\}.$$

Koska lisäksi jokaiselle $x \in E$

$$\begin{aligned} P(x, x) &= 1 - \sum_{z \in E \setminus \{x\}} \Psi(x, z) = 1 - \sum_{\lambda \in \Lambda \setminus \{x_s\}} \pi(\lambda | x_{S \setminus \{s\}}) \\ &= 1 - \left(1 - \pi(x_s | x_{S \setminus \{s\}})\right) = \pi(x_s | x_{S \setminus \{s\}}), \end{aligned}$$

yhtäsuuruus $P(x, y) = \Gamma_s(x, y)$ pätee kaikilla $x, y \in E$. \square

6.2. Metropolis-algoritmin ja Metropolis-jäähdytyksen suppeneminen

Olkoon konfiguraatioavaruus E tästä lähtien jokin kiinnitetty äärellinen epätyhjä joukko. Päädytään tarkastelemaan tilannetta, jossa yhtälön (6.1) siirtymämatriisiin

hyväksymisen todennäköisyydessä tasapainojakauma π korvataan kullakin hetkellä $n \geq 1$ Gibbsin jakaumalla

$$\pi_{T(n)} = \frac{1}{Z_{T(n)}} \exp\left(-\frac{U(\cdot)}{T(n)}\right)$$

jonkin sopivasti valitun jäähdytysaikataulun $(T(n))_{n \geq 1} \subset (0, \infty)$ ja funktion $U : E \rightarrow \mathbb{R}$ suhteen. On hyvä huomata, että koska E ei ole nyt välttämättä tulomuotoa, jakaumiin $(\pi_{T(n)})_{n \geq 1}$ kytköksissä olevaa Gibbsin kenttää ei myöskään välttämättä ole olemassa. Metropolis-algoritmin etuna onkin, että funktioksi U voidaan valita mikä tahansa reaaliarvoinen funktio.

Määritellään Metropolis-jäähdytystä vastaavan epähomogeenisen Markovin ketjun siirtymämatriisit $(P_n)_{n \geq 1} \subset \mathcal{M}_{E \times E}$ asettamalla

$$P_n(x, y) = \begin{cases} \Psi(x, y) \left(1 \wedge \frac{\pi_{T(n)}(y)}{\pi_{T(n)}(x)}\right), & x \neq y, \\ 1 - \sum_{z \in E \setminus \{x\}} P_n(x, z), & x = y, \end{cases},$$

missä oletetaan, että symmetrinen matriisi $\Psi \in \mathcal{M}_{E \times E}$ on lisäksi *pelkistymätön*. Koska

$$\frac{\pi_{T(n)}(y)}{\pi_{T(n)}(x)} = \frac{Z_{T(n)}}{Z_{T(n)}} \cdot \exp\left(-\frac{U(y)-U(x)}{T(n)}\right) = \exp\left(-\frac{U(y)-U(x)}{T(n)}\right),$$

voidaan siirtymämatriisi P_n kirjoittaa myös eksplisiittisemmässä muodossa

$$(6.4) \quad P_n(x, y) = \begin{cases} \Psi(x, y) \left(1 \wedge \exp\left(-\frac{U(y)-U(x)}{T(n)}\right)\right), & x \neq y \\ 1 - \sum_{z \in E \setminus \{x\}} P_n(x, z), & x = y \end{cases}.$$

Tavoitteena on todistaa yhtälön (6.4) siirtymämatriiseja $(P_n)_{n \geq 1}$ vastaavalle Markovin ketjulle

- suppeneminen tasapainojakaumaan π_T , kun $T = T(n) = \text{vakio}$ (Metropolis-algoritmi), ja
- jäähdytyslause, kun jäähdytysaikataulu $(T(n))_{n \geq 1}$ toteuttaa tietyt ehdot (Metropolis-jäähdytys).

Aloitetaan määrittelemällä liuta vakioita. Olkoon kaikille $x, y \in E$ luvut

$$l(x, y) := \min\{k \geq 1 : \Psi^k(x, y) > 0\}.$$

Ehdotusmatriisin Ψ pelkistymättömyys takaa, että $l(x, y) < \infty$ kaikille $x, y \in E$. Olkoon lisäksi

$$\tau := \max\{l(x, y) : x, y \in E\} \quad \text{ja} \\ \vartheta := \min\{\Psi(x, y) : x, y \in E, \Psi(x, y) > 0\}.$$

Erityisesti siten jokainen siirtymä $x \rightarrow y$ matriisin Ψ suhteen on mahdollista tehdä vähintään τ askeleella. Korvataan lopuksi Gibbs-jäähdytyksessä esiintynyt luku Δ asettamalla

$$\Delta := \max\{U(y) - U(x) : \Psi(x, y) > 0, x, y \in E\},$$

joka tosin edelleen kuvaa maksimaalista lokaalia värähtelyä. *Oletetaan myös jälleen, että $\Delta > 0$.*¹ Määritellään vielä kunkin konfiguraation $x \in E$ naapurien joukko asettamalla

$$\mathcal{N}(x) := \{y \in E \mid y \neq x, \Psi(x, y) > 0\}.$$

LEMMA 6.2.1. *Konfiguraatioille $x \in E$ ja $y \in \mathcal{N}(x)$ on voimassa*

$$(6.5) \quad P_n(x, y) \geq \vartheta \exp\left(-\frac{\Delta}{T(n)}\right) > 0 \quad \text{kaikilla } n \geq 1.$$

On lisäksi olemassa konfiguraatio $\tilde{x} \in E$ ja luku $\delta > 0$ siten, että

$$(6.6) \quad P_n(\tilde{x}, \tilde{x}) \geq \vartheta \left(1 - \exp\left(-\frac{\delta}{T(n)}\right)\right) > 0 \quad \text{kaikilla } n \geq 1.$$

TODISTUS. Jokaiselle $x \in E$, $y \in \mathcal{N}(x)$ ja $n \geq 1$ pätee alaraja

$$P_n(x, y) \geq \Psi(x, y) \left(1 \wedge \exp\left(-\frac{\Delta}{T(n)}\right)\right) \geq \vartheta \exp\left(-\frac{\Delta}{T(n)}\right) > 0,$$

joten epäyhtälö (6.5) on tosi.

Koska U ei ole vakiofunktio ja koska Ψ on pelkistymätön, on olemassa konfiguraatiot $\tilde{x} \in E_0$ ja $z \in \mathcal{N}(\tilde{x})$, joille $\delta := U(z) - U(\tilde{x}) = U(z) - U_{\min} > 0$. Nyt

$$\begin{aligned} \sum_{y \in \mathcal{N}(\tilde{x})} P_n(\tilde{x}, y) &= \sum_{y \in \mathcal{N}(\tilde{x})} \Psi(\tilde{x}, y) \left(1 \wedge \exp\left(-\frac{U(y) - U(\tilde{x})}{T(n)}\right)\right) \\ &\leq \Psi(\tilde{x}, z) \exp\left(-\frac{\delta}{T(n)}\right) + \sum_{y \in \mathcal{N}(\tilde{x}) \setminus \{z\}} \Psi(\tilde{x}, y) \\ &\leq \Psi(\tilde{x}, z) \exp\left(-\frac{\delta}{T(n)}\right) + \left(1 - \Psi(\tilde{x}, \tilde{x}) - \Psi(\tilde{x}, z)\right) \\ &\leq 1 - \Psi(\tilde{x}, z) \left(1 - \exp\left(-\frac{\delta}{T(n)}\right)\right) \\ (6.7) \quad &\leq 1 - \vartheta \left(1 - \exp\left(-\frac{\delta}{T(n)}\right)\right). \end{aligned}$$

Epäyhtälöä (6.7) soveltamalla voidaan nyt arvioida

$$\begin{aligned} P_n(\tilde{x}, \tilde{x}) &= 1 - \sum_{y \in E \setminus \{\tilde{x}\}} P_n(\tilde{x}, y) = 1 - \sum_{y \in \mathcal{N}(\tilde{x})} P_n(\tilde{x}, y) \\ &\geq 1 - 1 + \vartheta \left(1 - \exp\left(-\frac{\delta}{T(n)}\right)\right) \\ &= \vartheta \left(1 - \exp\left(-\frac{\delta}{T(n)}\right)\right) > 0, \end{aligned}$$

joten myös epäyhtälö (6.6) on tosi. □

Voidaan nyt todistaa kiinteää lämpötilaparametria vastaava suppenemistulos.

LAUSE 6.2.2 (Metropolis-algoritmin suppeneminen). *Olkoon $T(n) = T > 0$ kaikilla $n \geq 1$ ja $P = P_n \in \mathcal{M}_{E \times E}$ kuten yhtälössä (6.4). Tällöin*

$$\nu P^n \xrightarrow{TV} \pi_T \quad \text{kaikilla } \nu \in \mathcal{D}_E.$$

¹Koska Ψ on pelkistymätön, voidaan helposti osoittaa (samaa tapaan kuten Lemmassa 4.4.1), että $\Delta = 0$ ja vain jos $U(x) = U(y)$ kaikilla $x, y \in E$. Oletus $\Delta > 0$ tehtiin siitä syystä, että jos U on vakiofunktio, suppenemistulokset 6.2.2 ja 6.2.5 eivät ole voimassa. (ks. [30, s. 184–185]).

TODISTUS. Ehdotusmatriisiin Ψ pelkistymättömyys takaa, että myös matriisi P on pelkistymätön. Lisäksi Lemman 6.2.1 nojalla on olemassa konfiguraatio $\tilde{x} \in E$ siten, että $P(\tilde{x}, \tilde{x}) > 0$, joten P on myös jaksoton ([15, Exercise 4.2]). P on siten primitiivinen Lauseen 3.1.14 nojalla. Koska lisäksi π_T on siirtymämatriisin P yksikäsitteinen invariantti jakauma (Lemma 6.1.1), Seuraus 3.4.8 antaa väitteen. \square

Olkoon sitten $(T(n))_{n \geq 1} \subset (0, \infty)$ jäähdytysaikataulu, jolle on voimassa seuraavat oletukset:

$$(6.8) \quad \begin{aligned} & \bullet (T(n))_{n \geq 1} \text{ on vähenevä ja } T(n) \xrightarrow{n \rightarrow \infty} 0 \\ & \bullet T(n) \geq \frac{\tau \Delta}{\log n} \text{ kaikilla } n \geq n_0 \text{ jostain } n_0 \geq 2 \text{ lähtien.} \end{aligned}$$

Merkitään lisäksi jälleen kaikilla $k \geq 1$

$$Q_k = P_{T_{k-1}+1} \cdots P_{T_k}, \text{ missä nyt } T_k := k\tau, k = 0, 1, \dots$$

LEMMA 6.2.3. *On olemassa konfiguraatio $\tilde{x} \in E$ ja vakio $C \in (0, 1)$ siten, että*

$$\min_{x \in E} Q_k(\tilde{x}, x) \geq \frac{C\vartheta^\tau}{k\tau} \text{ kaikilla } k \geq 1.$$

TODISTUS. Olkoon $\tilde{x} \in E$ kuten Lemman 6.2.1 todistuksessa. Huomataan aluksi, että koska $\exp\left(-\frac{\delta}{T(n)}\right) \xrightarrow{n \rightarrow \infty} 0$, epäyhtälön (6.6) nojalla on olemassa luku $n_1 \geq 1$ siten, että

$$(6.9) \quad P_n(\tilde{x}, \tilde{x}) \geq \vartheta \left(1 - \exp\left(-\frac{\delta}{T(n)}\right)\right) \geq \vartheta \exp\left(-\frac{\Delta}{T(n)}\right), \text{ kun } n \geq n_1.$$

Riittävän suurilla $n \geq 1$ todennäköisyydelle $P_n(\tilde{x}, \tilde{x})$ on siis voimassa sama alaraja, kuin todennäköisyyksille $P_n(x, y)$, missä $y \in \mathcal{N}(x)$ (epäyhtälö (6.5)).

Todetaan, että konfiguraatio \tilde{x} voidaan saavuttaa lähtien mistä tahansa konfiguraatiosta $x \in E$ tasan $l(x, \tilde{x}) \leq \tau$ askeleella². Epäyhtälön (6.6) nojalla on tämän jälkeen mahdollista "levätä" konfiguraatiossa \tilde{x} tasan $\tau - l(x, \tilde{x})$ ajanhetkeä. Jokaiselle $x \in E$ ja $k \geq 1$ on siten olemassa jono $x = z^{(0)}, z^{(1)}, \dots, z^{(\tau)} = \tilde{x}$ joukon E konfiguraatioita siten, että $P_{T_{k-1}+j}(z^{(j-1)}, z^{(j)}) > 0$ kaikilla $j = 1, 2, \dots, \tau$. Ollaan siis saatu osoitettua, että kaikille $x \in E$ ja $k \geq 1$ sekä sopivasti valitulle polulle on voimassa

$$(6.10) \quad Q_k(x, \tilde{x}) \geq \prod_{j=1}^{\tau} P_{T_{k-1}+j}(z^{(j-1)}, z^{(j)}) > 0.$$

Jos nyt (a) $T_{k-1} + 1 \geq n_1$ ja (b) $T_k \geq n_0$, käyttämällä epäyhtälöitä (6.5), (6.9) ja (6.10) sekä jäähdytysaikatauluun liittyvää oletusta (6.8) saadaan

$$\begin{aligned} Q_k(x, \tilde{x}) & \geq \prod_{j=1}^{\tau} P_{T_{k-1}+j}(z^{(j-1)}, z^{(j)}) \stackrel{(a)}{\geq} \prod_{j=1}^{\tau} \vartheta \exp\left(-\frac{\Delta}{T([k-1]\tau + j)}\right) \\ & \geq \vartheta^\tau \exp\left(-\frac{\tau \Delta}{T(k\tau)}\right) \stackrel{(b)}{\geq} \frac{\vartheta^\tau}{k\tau}, \end{aligned}$$

$$\text{sillä } T([k-1]\tau + j) \geq T(k\tau) \geq \frac{\tau \Delta}{\log(k\tau)} \text{ kaikilla } j = 1, \dots, \tau.$$

²Helposti nähdään, että jos $\Psi^{l(x, \tilde{x})}(x, \tilde{x}) > 0$, niin myös $P_i \cdots P_{i+l(x, \tilde{x})}(x, \tilde{x}) > 0$ kaikilla $i \geq 1$.

Ollaan siten saatu arvio $\min_{x \in E} Q_k(x, \tilde{x}) \geq \vartheta^\tau (k\tau)^{-1}$ edellyttäen, että ehdot (a) ja (b) ovat voimassa. Muussa tapauksessa määritellään

$$C := \min_{\substack{x \in E \\ k \geq 1}} \{Q_k(x, \tilde{x}) : \text{ehto (a) tai (b) ei toteudu}\}.$$

Koska minimi otetaan yli äärellisen joukon ja epäyhtälön (6.10) nojalla $Q_k(x, \tilde{x}) > 0$ kaikilla $x \in E$ ja $k \geq 1$, niin $C \in (0, 1)$. Siten kaikille $k \geq 1$ pätee

$$\min_{x \in E} Q_k(x, \tilde{x}) \geq \min \left\{ C, \frac{\vartheta^\tau}{k\tau} \right\} \geq \frac{C\vartheta^\tau}{k\tau},$$

mikä todistaa väitteen. \square

SEURAUUS 6.2.4.

$$c(P_i \cdots P_n) \xrightarrow{n \rightarrow \infty} 0 \text{ kaikilla } i \geq 1.$$

TODISTUS. Luvuille $c(Q_k)$ saadaan yläraja käyttämällä Lausetta 1.3.5 ja Lemmaa 6.2.3:

$$\begin{aligned} c(Q_k) &= \frac{1}{2} \max_{x, y \in E} \|Q_k(x, \cdot) - Q_k(y, \cdot)\| = \max_{x, y \in E} \left(1 - \sum_{z \in E} Q_k(x, z) \wedge Q_k(y, z) \right) \\ &\leq 1 - \min_{x, y \in E} \sum_{z \in E} Q_k(x, z) \wedge Q_k(y, z) \leq 1 - \sum_{z \in E} \min_{x \in E} Q_k(x, z) \\ &\leq 1 - \min_{x \in E} Q_k(x, \tilde{x}) \leq 1 - \frac{C\vartheta^\tau}{k\tau}. \end{aligned}$$

Koska siten

$$\sum_{k=1}^{\infty} (1 - c(Q_k)) \geq \frac{C\vartheta^\tau}{\tau} \sum_{k=1}^{\infty} \frac{1}{k} = \infty,$$

väite saadaan käyttämällä Lemmaa 3.4.12. \square

LAUSE 6.2.5 (Metropolis-jäähdytys). *Olkoon E äärellinen konfiguraatioavaruus ja $(P_n)_{n \geq 1} \subset \mathcal{M}_{E \times E}$ kaavan (6.4) mukaiset epähomogeenisen Markovin ketjun siirtymämatriisit. Oletetaan lisäksi, että energiefunktio $U : E \rightarrow \mathbb{R}$ ei ole vakiofunktio, ja että jäähdytysaikataululle $(T(n))_{n \geq 1} \subset \mathbb{R} \setminus \{0\}$ on voimassa seuraavat oletukset:*

- $(T(n))_{n \geq 1}$ on vähenevä ja $T(n) \xrightarrow{n \rightarrow \infty} 0$
- $T(n) \geq \frac{\tau \Delta}{\log n}$ kaikilla $n \geq n_0$ jostain $n_0 \geq 2$ lähtien.

Tällöin ketju on vahvasti ergodinen;

$$\nu P_i \cdots P_n \xrightarrow{TV} \pi_0 \quad \text{kaikilla } \nu \in \mathcal{D}_E \text{ ja } i \geq 1,$$

missä π_0 on tasajakauma joukon E minienergiakonfiguraatioiden osajoukossa E_0 .

TODISTUS. Käytetään jälleen Dobrushinin lausetta 3.5.2 suppenemisen todistamiseksi. Lauseen 3.5.2 oletukset ovat voimassa, sillä

- $c(P_i \cdots P_n) \xrightarrow{n \rightarrow \infty} 0$ kaikilla $i \geq 1$ Seurauksen 6.2.4 nojalla
- $\pi_{T(n)} P_n = \pi_{T(n)}$ kaikille $n \geq 1$ Lemman 6.1.1 nojalla

- $\sum_{n=1}^{\infty} \left| \left| \pi_{T(n)} - \pi_{T(n+1)} \right| \right| < \infty$ Seurauksen 5.1.3 nojalla.

Koska lisäksi $\pi_0 = \lim_{n \rightarrow \infty} \pi_{T(n)}$ Lauseen 5.1.1 nojalla, saadaan väite todistettua. \square

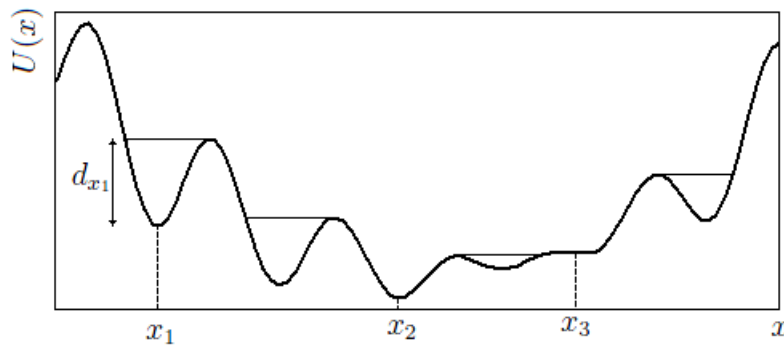
6.3. Optimaalinen vakio jäähdytysaikataululle

Sekä tämän luvun että Luvun 5 jäähdytyslauseissa ei juurikaan kiinnitetty huomiota siihen, kuinka pieneksi jäähdytysaikataulun oletukseen

$$T(n) \geq \frac{c}{\log n} \quad (\text{jostain lähtien})$$

liittyvä vakio $c > 0$ voidaan valita suppenemisen takaamiseksi. Mitä suurempi vakio c on, sitä suuremmalla todennäköisyydellä ketjun on mahdollista tehdä energiaa kasvattavia siirtymiä. Liian pieni c puolestaan mahdollistaa ketjun juuttumisen lokaaliin minimikohtaan. Kuten todettua, ensimmäisen arvion suppenemisen takaavalle vakiolle c löysivät S. ja D. Geman [8]. Kyseiselle vakiolle $N\Delta$, missä N on solmujen lukumäärä ja Δ maksimaalinen kahden konfiguraation energian erotus, jäähdytys on käytännön kannalta usein aivan liian hidasta. Esimerkeksi tilanteessa, jossa $N = 20000$ ja $\Delta = 1$, tarvitaan e^{40000} iteraatiota lämpötilan $T = 0,5$ saavuttamiseksi (kuten samaisessa artikkelissa [8] todetaan). Artikkelin julkaisun jälkimainingeissa pyrittiinkin kuumeisesti löytämään yhä parempia ja terävämpiä arvoja jäähdytysvakion arvoksi monenkaltaisissa tilanteissa, kunnes lopulta B. Hajek osoitti vuoden 1986 artikkelissaan [11], että vakion c suuruudelle on olemassa suppenemisen takaamiseksi riittävä ja välttämättömätön ehto. Todistus on tehty tarkastelemalla jatkuva-aikaista vastinetta diskreettiaikaiselle Markovin ketjulle. Koska tuloksen ja siihen liittyvien käsitteiden avulla voidaan melko suoraviivaisella tavalla havainnollistaa simuloidun jäähdytyksen perusajatusta, esitellään ne vielä lyhyesti tämän luvun päätteeksi.

Olkoon $\Psi \in \mathcal{M}_{E \times E}$ symmetrinen ja pelkistymätön ehdotusmatriisi. Sanotaan, että konfiguraatiot x ja $y \in E$ *kommunikovat tasolla* h , jos joko $x = y$ ja $U(x) \leq h$ tai on olemassa konfiguraatioita $x \neq y$ yhdistävä polku $x = z_0, z_1, \dots, z_m = y$ siten, että $\Psi(z_{i-1}, z_i) > 0$ kaikille $i = 1, 2, \dots, m$, ja lisäksi $U(z_i) \leq h$ kaikilla $i = 0, 1, \dots, m$.



KUVA 6.1. Esimerkki jatkuvaksi laajennetusta jonkin energiefunktion U graafista. Kuvassa $x_1 \in E_{loc}$ (aito lokaali minimikohta), $x_2 \in E_0$ (globaali minimikohta) ja $x_3 \notin E_{loc}$ (epäaito lokaali minimikohta). Lisäksi nähdään, että $d_{x_1} = \sup\{d_x : x \in E_{loc} \setminus E_0\}$.

Sanotaan edelleen, että $x \in E$ on *aito lokaali minimikohta*, mikäli se ei kommunikoi minkään sellaisen tilan $y \in E$ kanssa tasolla $U(x)$, jolle $U(y) < U(x)$. Tämä tarkoittaa, että jos x kuitenkin kommunikoi tilan y kanssa (jollain tasolla), jokaiselta niistä yhdistävää polkua kohden on oltava vähintään yksi polun alkio $z \in E$, jolle $U(z) > U(x)$.

Määritellään sitten tilan $x \in E$ *syvyys* d_x : olkoon d_x pienin luku $d \geq 0$, jolle on olemassa $y \in E$ siten, että $U(y) < U(x)$ sekä x ja y kommunikoivat tasolla $U(x) + d$. Lisäksi globaaleille minimikohtille $x \in E_0$ asetetaan $d_x := \infty$.

LAUSE 6.3.1. *Olkoon $(X_n)_{n \geq 0}$ yhtälön (6.4) siirtymämatriiseja $(P_n)_{n \geq 1} \subset \mathcal{M}_{E \times E}$ vastaava epähomogeeninen Markovin ketju. Tällöin kaikilla alkujakaumilla $\nu \in \mathcal{D}_E$*

$$\mathbb{P}_\nu(X_n \in E_0) = \sum_{x \in E_0} \nu P_1 \cdots P_n(x) \xrightarrow{n \rightarrow \infty} 1 \quad \text{jos ja vain jos}$$

$$\sum_{n=1}^{\infty} \exp\left(-\frac{D}{T(n)}\right) = \infty,$$

missä $D = \sup\{d_x : x \in E_{loc} \setminus E_0\}$ (ks. Kuva 6.1) ja E_{loc} on joukon E aitojen lokaalien minimikohtien osajoukko.

HUOMAUTUS 6.3.2. Luvusta 5 muistetaan, että termi $\exp\left(-\frac{D}{T(n)}\right)$ voidaan tulkita sellaisen (hetkellä n) tapahtuvan siirtymän hyväksymisen todennäköisyydeksi, missä energia kasvaa määrällä $c > 0$. Mitä suurempi Lauseessa 6.3.1 esiintyvä D on, sitä herkemmin energiaa kasvattavia siirtymiä hyväksytään. Jos D on puolestaan hyvin pieni, hyväksytään käytännössä ainoastaan energiaa vähentäviä siirtymiä, lokaaliin minimikohtaan juuttumisen riski kasvaa.

HUOMAUTUS 6.3.3. Jos jäähdytysaikataulu on muotoa $T(n) = \frac{c}{\log(n+1)}$ jollakin $c > 0$, Lauseen 6.3.1 nojalla ketju suppenee minimienergiatilojen joukkoon jos ja vain jos $c \geq D$. Näin on, sillä

$$\sum_{n=1}^{\infty} \exp\left(-\frac{D}{T(n)}\right) = \sum_{n=2}^{\infty} \exp\left(-\frac{D}{c} \log n\right) = \sum_{n=2}^{\infty} n^{-\frac{D}{c}} = \infty \iff c \geq D.$$

Tenttialiongelma ja simuloitu jäähditys käytännössä

7.1. Tenttialiongelman sanallinen kuvailu

Tarkastellaan suorakulmion muotoista $M \times N$ -ruudukkoa, jossa jokainen ruutu vastaa tenttialin yhtä istumapaikkaa. Oletetaan, että lähestyvään tenttitilaisuuteen on ilmoittautunut tenttijöitä L :ää eri tenttiä varten, joita merkitään symboleilla $\lambda_1, \dots, \lambda_L$. Tiedetään lisäksi, että kuhunkin tenttiin λ_i on ilmoittautunut m_i tenttijää siten, että kaikki tentin tekijät mahtumat saliin, ts. $m_1 + \dots + m_L \leq MN$. Symbolilla λ_0 merkittävien tyhjien paikkojen lukumäärä on tällöin $MN - \sum_{i=1}^L m_i \geq 0$. Halutaan sijoittaa eri tenttijät tenttialiin vilpin estämiseksi ja työrauhan maksimimiseksi mahdollisimman hyvin siten, että istumajärjestyksen x hyvyyttä mittaava kokonaissakko (energia) $H(x)$ on mahdollisimman pieni. Kokonaissakkoa mittaava sakkofunktio/energiafunktio H muodostuu seuraavien ehtojen perusteella:

- I jos kaksi saman tentin tekijää ovat vierekkäin, sakko kasvaa 20 yksikköä
- II jos kaksi eri tentin tekijää ovat vierekkäin, sakko kasvaa 7 yksikköä
- III jos kaksi saman tentin tekijää ovat peräkkäin, sakko vähenee 1 yksikön
- IV jos kaksi eri tentin tekijää ovat peräkkäin, sakko kasvaa 1 yksikön
- V jos kaksi saman tentin tekijää ovat vinottain, sakko kasvaa 5 yksikköä.

Tavoitteena on nyt

- muodostaa ongelmalle Metropolis-jäähdityksen soveltamista varten sopiva matemaattinen malli
- osoittaa, että laadittu malli toteuttaa Lauseen 6.2.5 ehdot
- pyrkiä minimoimaan kokonaissakko H simuloidun jäähdityksen avulla.

7.2. Tenttialiongelmaan liittyvän matemaattisen mallin muodostaminen

Lähdetään liikkeelle mahdollisimman yleisellä muotoilulla ja todistetaan, että Lause 6.2.5 on voimassa tilannetta mallintavalle, sopivasti määritellylle Metropolis-jäähdityksen ketjulle. Tämän jälkeen kiinnitetään lukuarvoja, tehdään simulointikoikeita eri parametrien arvoilla ja tulkitaan lopuksi saatuja tuloksia.

Luonteva valinta $M \times N$ -paikkaista tenttialia vastaavaksi solmujen joukoksi on

$$S = S_{M,N} := \{(s_1, s_2) \in \mathbb{Z}^2 \mid s_1 \in \{1, 2, \dots, M\}, s_2 \in \{1, 2, \dots, N\}\} \subset \mathbb{Z}^2,$$

jossa kukin $s = (s_1, s_2) \in S$ vastaa yhtä istumapaikkaa. Indeksillä s_1 kertoo, millä rivillä ja s_2 millä sarakkeella kyseinen paikka sijaitsee. Olkoon sitten $\Lambda = \{\lambda_0, \lambda_1, \dots, \lambda_L\} \subset \mathbb{R}$ faasiavaruutta vastaava joukko, $1 \leq |L| < \infty$, missä

- arvo λ_0 vastaa tyhjää paikkaa
- arvot $\lambda_i, i = 1, 2, \dots, L$ vastaavat eri tenttejä.

Kiinnitetään seuraavaksi luvut $m_0, m_1, \dots, m_L \in \mathbb{N}$ siten, että $\sum_{k=0}^L m_k = MN$. Kuten todettua, luvut m_k ilmaisevat tenttien (tai tyhjien paikkojen) lukumääriä.

Konfiguraatioavaruudeksi eli istumajärjestysten joukoksi E asetetaan tuloavaruuden Λ^S osajoukko

$$E := \{x : S \rightarrow \Lambda \mid \#\{s \in S : x_s = \lambda_k\} = m_k \text{ kaikilla } k = 0, 1, \dots, L\} \subset \Lambda^S.$$

Vaikka ei-tulomuotoisen konfiguraatioavaruuden E yhteydessä esimerkiksi faaseista ja satunnaiskentistä puhuminen on jokseenkin mieleetöntä, energiafunktion rakentamista varten tietyt Luvun 2 käsitteet osoittautuvat kuitenkin hyödyllisiksi. Valitaan joukolle S naapurustojärjestelmä $\mathcal{N} = (\mathcal{N}_s)_{s \in S}$ määrittelemällä kaikille $s \in S$ naapurustot

$$\mathcal{N}_s := \{t \in S : 0 < \|s - t\|_{L^2} \leq \sqrt{2}\}.$$

Solmun $s \in S$ naapureihin kuuluvat siis kaikki yhden pysty-, vaaka-, tai viistoaskeleen päässä olevat solmut. Joukon \mathcal{N}_s alkioiden eli solmun s naapurien lukumäärä on siten kolme, viisi tai kahdeksan solmun s sijainnista riippuen. Naapurustojärjestelmästä \mathcal{N} johdettuun klikkikokoelmaan $\mathcal{C}_{\mathcal{N}}$ kuuluu nyt korkeintaan neljän solmun klikkejä. Halutaan pitää malli yksinkertaisena rajoittamalla vuorovaikutusten määrää, joten muodostetaan naapuripotentialista $\mathcal{V} = (V_C)_{C \in \mathcal{C}_{\mathcal{N}}}$ paripotentiali.

Tarkastellaan kahden solmun klikkien osakokoelmaa $\mathcal{C}_2 = \{C \in \mathcal{C} : |C| = 2\}$. Jaetaan se edelleen kolmeen erilliseen osakokoelmaan $\mathcal{C}_2^{\text{vaaka}}$, $\mathcal{C}_2^{\text{pysty}}$ ja $\mathcal{C}_2^{\text{vino}}$ siten, että

$$\mathcal{C}_2^{\text{vaaka}} := \left\{ \{s, t\} \in \mathcal{C}_2 : \|s - t\|_{L^2} = |s_1 - t_1| = 1 \right\} \text{ (vierekkäin olevat paikat)}$$

$$\mathcal{C}_2^{\text{pysty}} := \left\{ \{s, t\} \in \mathcal{C}_2 : \|s - t\|_{L^2} = |s_2 - t_2| = 1 \right\} \text{ (päällekkäin olevat paikat)}$$

$$\mathcal{C}_2^{\text{vino}} := \left\{ \{s, t\} \in \mathcal{C}_2 : \|s - t\|_{L^2} = \sqrt{2} \right\} \text{ (viistossa toisiinsa nähden olevat paikat)}.$$

Voidaan oitis todeta, että jokainen $C \in \mathcal{C}_2$ kuuluu täsmälleen yhteen näistä osakokoelmista. Notation yksinkertaistamiseksi määritellään sitten kullekin kaksiolle $\{s, t\} \subset S$ indikaattorifunktiot $f_{\{s,t\}}, g_{\{s,t\}} : E \rightarrow \mathbb{R}$;

$$f_{\{s,t\}}(x) = \mathbb{1}\{x_s = x_t \in \Lambda \setminus \{\lambda_0\}\}$$

$$g_{\{s,t\}}(x) = \mathbb{1}\{x_s \neq x_t \text{ ja } x_s, x_t \in \Lambda \setminus \{\lambda_0\}\}$$

Funktio $f_{\{s,t\}}$ ilmaisee, onko paikoissa s ja t samat tentit (1 = kyllä, 0 = ei). Funktio $g_{\{s,t\}}$ puolestaan kertoo, ovatko tentit paikoissa s ja t toisistaan eriävät (1 = kyllä, 0 = ei). Erityisesti siis molemmat funktiot antavat nollan, jos jompi kumpi paikoista on tyhjä, ts. $s = \lambda_0$ tai $t = \lambda_0$.

Määritellään nyt naapuripotentiali $\mathcal{V} := (V_C)_{C \in \mathcal{C}_{\mathcal{N}}}$ asettamalla kaikille $x \in E$

$$V_C(x) := \begin{cases} a_{\text{vaaka}} f_{\{s,t\}}(x) + b_{\text{vaaka}} g_{\{s,t\}}(x), & C = \{s, t\} \in \mathcal{C}_2^{\text{vaaka}} \\ a_{\text{pysty}} f_{\{s,t\}}(x) + b_{\text{pysty}} g_{\{s,t\}}(x), & C = \{s, t\} \in \mathcal{C}_2^{\text{pysty}} \\ a_{\text{vino}} f_{\{s,t\}}(x), & C = \{s, t\} \in \mathcal{C}_2^{\text{vino}} \\ 0, & |C| \neq 2 \end{cases},$$

missä parametrivektorin $\theta := (a_{\text{vaaka}}, b_{\text{vaaka}}, a_{\text{pysty}}, b_{\text{pysty}}, a_{\text{vino}}) \in \mathbb{R}^5$ koordinaatit ovat ennalta määrättyjä vakioita (ks. Taulukko 1). Potentialin \mathcal{V} määräämäksi energiafunktiksi U_{θ} valitulle θ saadaan

$$(7.1) \quad U_{\theta}(x) = \sum_{C \in \mathcal{C}_2^{\text{vaaka}}} V_C(x) + \sum_{C \in \mathcal{C}_2^{\text{pysty}}} V_C(x) + \sum_{C \in \mathcal{C}_2^{\text{vino}}} V_C(x).$$

Ongelmanasettelussa esitelty sakkofunktio H vastaa nyt energiefunktiota U_θ , missä $\theta = (20, 7, -1, 1, 5)$.

TAULUKKO 1. Energiaparametrien tulkinnat.

<i>muutos</i>	<i>energian muutoksen syy</i>
a_{vaaka}	kaksi saman tentin tekijää ovat vierekkäin
b_{vaaka}	kaksi eri tentin tekijää ovat vierekkäin
a_{pysty}	kaksi saman tentin tekijää ovat peräkkäin
b_{pysty}	kaksi eri tentin tekijää ovat peräkkäin
a_{vino}	kaksi saman tentin tekijää ovat vinottain

7.3. Metropolis-jäähdytysketjun matemaattinen kuvailu

Ryhdytään seuraavaksi rakentamaan simuloitua jäähdytystä vastaavaa Markovin ketjua. Aloitetaan määrittelemällä kutakin järjestettyä solmuparia $(\alpha, \beta) \in S^2$ kohden yhtä vaihtoa vastaava kuvaus asettamalla $\varphi_{\alpha, \beta} : E \rightarrow E$; $\varphi_{\alpha, \beta}(x) = (\varphi_{\alpha, \beta}(x))_{s \in S}$, missä

$$(\varphi_{\alpha, \beta}(x))_s := \begin{cases} x_\beta, & \text{jos } s = \alpha \\ x_\alpha, & \text{jos } s = \beta \\ x_s, & \text{muuten.} \end{cases}$$

Kuvaus $x \mapsto \varphi_{\alpha, \beta}(x)$ siis vaihtaa konfiguraation x arvot keskenään solmuissa α ja β ja jättää muut solmut koskemattomiksi. Käytetään ehdotusmatriisin muodostamisessa seuraavaa yksinkertaista algoritmia:

- olkoon $Y_{n-1} = x \in E$ jokin alkukonfiguraatio,
- poimitaan riipputtomat kaksi solmua $\alpha, \beta \in S$ joukon S tasajakaumasta
- asetetaan $Y_n = y := \varphi_{\alpha, \beta}(x)$.

Lopputuloksena on yhtä vaihtoa vastaava siirtymämatriisi $\Psi = [\Psi(x, y)]$, jonka alkiot voidaan kirjoittaa muodossa

$$(7.2) \quad \Psi(x, y) = \frac{\#\{(\alpha, \beta) \in S^2 \mid \varphi_{\alpha, \beta}(x) = y\}}{\#\{(\alpha, \beta) \in S^2\}} = \frac{\sum_{(\alpha, \beta)} \mathbf{1}\{\varphi_{\alpha, \beta}(x) = y\}}{|S|^2}$$

kaikilla $x, y \in E$. Voidaan kirjoittaa myös eksplisiittisemmin

$$\Psi(x, y) = \begin{cases} \frac{\sum_{k=0}^L m_k^2}{(MN)^2}, & \text{jos } x = y \\ \frac{2}{(MN)^2}, & \text{jos } \exists (\alpha, \beta) \in S^2 \text{ s.e. } x \neq \varphi_{\alpha, \beta}(x) = y \\ 0, & \text{muuten.} \end{cases}$$

Perustelut:

- jos $x = y$, niin $\varphi_{\alpha, \beta}(x) = y$ täsmälleen silloin, kun transpositio ei tee mitään, toisin sanoen $x_\alpha = x_\beta$. Tulosäännon nojalla tällaisia pareja (α, β) on $m_0^2 + m_1^2 + \dots + m_L^2$ kappaletta.
- Jos x ja y poikkeavat täsmälleen kahdessa solmussa α ja β , ainoat soveliaat parit ovat (α, β) ja (β, α) , joita on siis kaksi kappaletta.
- Jos x ja y poikkeavat useammassa kuin kahdessa solmussa, mikään transpositio ei tuota siirtymää $x \rightarrow y$.

LEMMA 7.3.1. Ψ on symmetrinen joukon $\mathcal{M}_{E \times E}$ stokastinen siirtymämatriisi.

TODISTUS. Osoitetaan ensin, että $\Psi \in \mathcal{M}_{E \times E}$. Esityksestä (7.2) nähdään heti, että $\Psi \geq 0$. Olkoon sitten $x \in E$. Tällöin

$$\begin{aligned} \sum_{y \in E} \Psi(x, y) &= \frac{1}{|S|^2} \sum_{y \in E} \sum_{(\alpha, \beta) \in S^2} \mathbb{1}\{\varphi_{\alpha, \beta}(x) = y\} \\ &= \frac{1}{|S|^2} \sum_{(\alpha, \beta) \in S^2} \sum_{y \in E} \mathbb{1}\{\varphi_{\alpha, \beta}(x) = y\} \\ &= \frac{1}{|S|^2} \sum_{(\alpha, \beta) \in S^2} \mathbb{1}\{\varphi_{\alpha, \beta}(x) \in E\} = \frac{|S|^2}{|S|^2} = 1. \end{aligned}$$

Siten $\Psi \in \mathcal{M}_{E \times E}$. Todistetaan seuraavaksi ehdotusmatriisin Ψ symmetrisyys. Olkoon $(\alpha, \beta) \in S^2$ ja $x \in E$. Osoitetaan aluksi, että $\varphi_{\alpha, \beta} \circ \varphi_{\alpha, \beta} = \text{id}_E$, missä id_E on identtinen kuvaus joukolla E . Näin on, sillä kuvauksen $\varphi_{\alpha, \beta}$ määritelmän nojalla

$$(\varphi_{\alpha, \beta}(\varphi_{\alpha, \beta}(x)))_s = \begin{cases} (\varphi_{\alpha, \beta}(x))_\beta = x_\alpha, & s = \alpha \\ (\varphi_{\alpha, \beta}(x))_\alpha = x_\beta, & s = \beta \\ (\varphi_{\alpha, \beta}(x))_s = x_s, & s \in S \setminus \{\alpha, \beta\} \end{cases}.$$

Jos siis nyt $\varphi_{\alpha, \beta}(x) = y$, niin yhtäpitävästi $\varphi_{\alpha, \beta}(y) = \varphi_{\alpha, \beta}(\varphi_{\alpha, \beta}(x)) = \text{id}_E(x) = x$. Väitön seurauksena on, että $\{(\alpha, \beta) \in S^2 \mid \varphi_{\alpha, \beta}(x) = y\} = \{(\alpha, \beta) \in S^2 \mid \varphi_{\alpha, \beta}(y) = x\}$. On siis $\Psi(x, y) = \Psi(y, x)$, ja väite on siten todistettu. \square

LEMMA 7.3.2. Ψ on pelkistymätön.

TODISTUS. Matriisin Ψ pelkistymättömyys seuraa osoittamalla, että Ψ on itse asiassa primitiivinen: $r := MN - 1$ on riittävän iso luku, jolle Ψ^r on aidosti positiivinen.

Olkoon x ja $y \in E$. Numeroidaan solmut $s \in S$ luonnollisilla luvuilla $1, 2, \dots, r + 1$, jolloin siis $S = \{1, 2, \dots, r + 1\}$. Konstruoidaan sitten polku $z^{(0)}, z^{(1)}, \dots, z^{(r)}$ joukon E konfiguraatioita, joille $\Psi(z^{(i-1)}, z^{(i)}) > 0$ kaikilla $i = 1, \dots, r$. Määritellään nämä konfiguraatiot rekursiivisesti asettamalla

$$z^{(0)} := x \quad \text{ja} \quad z^{(i)} := \varphi_{\alpha_i, \beta_i}(z^{(i-1)}),$$

missä $\alpha_i = i$ ja $\beta_i = \min \{s \in \{i, i + 1, \dots, r + 1\} : z_s^{(i-1)} = y_i\}$ kaikilla $i = 1, \dots, r$. Mitä käytännössä tehdään on se, että jokaisella aika-askeleella etsitään sen hetki-tilan $z^{(i-1)}$ rajoittumajoukosta $\{z_s^{(i-1)} : i \leq s \leq r + 1\}$ järjestyksessä ensimmäinen sellainen solmu \tilde{s} , jossa $z^{(i-1)}$ vastaa tavoitekonfiguraation y arvoa solmussa i , ja uusi tila $z^{(i)}$ muodostetaan vaihtamalla konfiguraation $z^{(i-1)}$ arvot keskenään näissä solmuissa (jos $\tilde{s} = i$, muutosta ei tapahdu). Lukijan harrastuneisuuden varaan jätetään osoittaa, että jono $(z^{(i)})_{i=0}^r \subset E$ on tällöin hyvin määritelty, $\Psi(z^{(i-1)}, z^{(i)}) > 0$ kaikilla $i = 1, \dots, r$ ja $z^{(r)} = y$. Siirtymämatriisin Ψ primitiivisyys – ja siten myös pelkistymättömyys – seuraa nyt toteamalla, että

$$\Psi^r(x, y) \geq \prod_{i=1}^r \Psi(z^{(i-1)}, z^{(i)}) > 0.$$

\square

Voidaan nyt määritellä Metropolis-jäähdytyksen siirtymämatriiseja $(P_n)_{n \geq 1}$ vastaava ketju soveltamalla kaavaa (6.4) kaikille $n \geq 1$, missä Ψ määräytyy yhtälöstä (7.2) ja $\pi_{T(n)}$ vastaa Gibbsin jakaumaa konfiguraatioavaruudella E energiafunktion $U = U_\theta$ (7.1) suhteen kiinnitettylle parametrivektorille $\theta \in \mathbb{R}^6$. Jos nyt

- θ on valittu siten, että U_θ ei ole vakiofunktio¹
- *vähenevälle* jäähdytysaikataululle pätee $T(n) \xrightarrow{n \rightarrow \infty} 0$ ja

$$T(n) \geq \frac{\tau \Delta}{\log n} \text{ kaikilla } n \geq n_0 \text{ jostain } n_0 \geq 2 \text{ lähtien,}^2$$

niin Lauseen 6.2.5 oletukset ovat voimassa, ja siten siirtymämatriiseja (P_n) vastaava epähomogeeninen Markovin ketju (X_0, X_1, \dots) on vahvasti ergodininen. Erityisesti siten kaikille alkukonfiguraatioille $x \in E$

$$\mathbb{P}(X_n = y \mid X_0 = x) \xrightarrow{n \rightarrow \infty} \pi_0(y) = \begin{cases} 1/|E_0|, & y \in E_0 \\ 0, & \text{muuten} \end{cases}.$$

HUOMAUTUS 7.3.3. Lausetta 6.2.5 voidaan siten soveltaa luonnollisesti myös tutkimusongelmaan liittyvän sakkofunktion $H = U_{(20,7,-1,1,5)}$ tapauksessa, kunhan on olemassa alkio $x, y \in E$ siten, että $H(x) \neq H(y)$.

7.4. Empiiriset tarkastelut

Tässä kohdassa käsitellään tenttialiongelman ratkaisemiseen tähtävää Metropolis-jäähdytyksen käytännön soveltamista. Tehdään yksinkertaistus ja merkitään $\lambda_i = i$ kaikilla $i = 1, 2, \dots, L$. Tyhjä paikka vastaa siis lukua 0 ja luvut $1, 2, \dots, L$ eri tenttejä. Tenttialin koon ja koostumuksen valinnassa simulointikokeita varten oli luonnollisesti melko runsaasti valinnanvaraa. Simulointeihin käytettävän ajan kontrolloimiseksi, ja toisaalta selkeyden vuoksi päädyttiin tarkastelemaan ensisijaisesti seuraavanlaista tilannetta:

- 6×8 -ruudukko, jolle $m_0 = 21$, $m_1 = \dots = m_3 = 7$ ja $m_4 = 6$.

Toisistaan poikkeavia istumajärjestyksiä on (näinkin pienen salin) tapauksessa yhteensä

$$\frac{48!}{21! 7! 7! 7! 6!} \approx 2.64 \cdot 10^{27} \text{ kappaletta.}$$

Simuloinneissa käytettiin energiafunktioita $H = U_{(20,7,-1,1,5)}$ ja ehdotusmatriiseja Ψ (7.2) *logaritmisten* jäähdytysaikataulujen $(T(n))_{n \geq 1}$ ollessa logaritmista muotoa

$$T(n) = \frac{c}{\log(n+1)} \text{ kaikilla } n \geq 1 \text{ jollekin } c > 0.$$

Simuloitu jäähdytys voidaan pysäyttää periaateessa kahdella eri tavalla: voidaan joko määrätä iteraatioiden määrä etukäteen tai soveltaa tietynlaista ennalta määrättyä *pysäyttämissääntöä*; voidaan esimerkiksi terminoida algoritmi, kun energioiden muutokset pysyttelevät ennalta sovitun kynnsarvon alapuolella riittävän pitkään. Tämän tutkielman simuloinneissa iteraatioiden määrät kiinnitettiin ennalta.

¹On huomattava, että funktion U_θ vakiona oleminen ei riipu pelkästään vektorista θ , vaan myös konfiguraatioavaruuden E rakenteesta. Esimerkiksi 4×4 -tenttialille, jolle $m_0 = |\lambda_0| = 3$ ja $m_1 = |\lambda_1| = 1$ U_θ on vakiofunktio *kaikilla* $\theta \in \mathbb{R}^5$, kuten helposti voidaan todeta.

²Muistutus: $\tau = \max\{l(x, y) : x, y \in E\}$, missä $l(x, y) = \min\{k \geq 1 : \Psi^k(x, y) > 0\}$ ja $\Delta = \max\{U(x) - U(y) : \Psi(x, y) > 0, x, y \in E\}$.

Koska kaikki käytännössä tapahtuva simuloitu jäädytys on äärellisen ajan simulointia, ja aiemmin todistetut suppenemislauseet takaavat ainoastaan rajalla tapahtuvan minimienergiatilojen joukkoon suppenemisen, joudutaan usein tyytymään approksimaatioihin. Pyritäänkin seuraavaksi simulointikokeiden avulla selvittämään, minkälaisia vaikutuksia energiafunktion U , jäädytysvakion c , iteraatioiden määrän n , ehdotusmatriisiin Ψ ja konfiguraatioavaruuden E valinnalla on havaittaviin likiarvoihin.

0	0	0	0	0	0	0	0	0	1	0	2	0	4	0	0	3	1	0	2	0	3	0	4	0
0	0	0	0	0	0	0	0	0	1	0	2	0	4	0	0	3	1	0	2	0	3	0	4	1
0	0	0	0	0	1	1	1	1	1	0	2	1	0	2	0	3	1	0	2	0	3	0	4	0
1	1	1	1	2	2	2	2	2	0	3	0	1	0	2	0	3	1	0	2	0	3	0	4	2
2	2	2	3	3	3	3	3	3	4	0	4	1	0	2	0	3	1	0	2	0	3	0	4	0
3	3	4	4	4	4	4	4	4	4	0	4	1	0	2	0	3	1	0	2	0	3	0	4	3
x_0									x_1							x_2								

KUVA 7.1. Kuvassa x_0 vastaa yleisesti käytössä ollutta alkuarvoa; x_1 on lokaali minimikohta ja x_2 on globaali minimikohta. Energiat ovat $H(x_0) = 440$, $H(x_1) = 3$ ja $H(x_2) = 1$.

7.4.1. Energiafunktion skaalaaminen ja jäädytysvakion valinta. Olkoon $c > 0$ logaritmiseen jäädytys-aikatauluun $(T(n))_{n \geq 1}$ liittyvä jäädytysvakio. Ongelma: miten annettu sakkofunktio $H = U_{(20,7,-1,1,5)}$ kannattaa skaalata, jotta äärellisen monella Metropolis-jäädytyksen iteraatiolla saataisiin mahdollisimman hyviä tuloksia (iteraatioiden maksimimäärän ollessa kiinteä)? Toisin sanoen, miten skaalauskerroin $\alpha > 0$ kannattaa valita simuloinnissa sovellettavalle energiafunktiolle $U := \alpha H$? Tarkastellaan energiaa määrällä $U(y) - U(x) > 0$ kasvattavan siirtymän $x \rightarrow y$ hyväksymisen todennäköisyyttä hetkellä $n \geq 1$:

$$\begin{aligned} \exp\left(-\frac{U(y) - U(x)}{T(n)}\right) &= \exp\left(-\frac{\alpha(H(y) - H(x)) \log(n+1)}{c}\right) \\ &= \exp\left(-\frac{H(y) - H(x)}{\tilde{T}(n)}\right), \quad \text{missä } \tilde{T}(n) = \frac{c/\alpha}{\log(n+1)}. \end{aligned}$$

Käytännön kannalta on siten yhtäpitävää käyttää simuloinnissa energiafunktio/jäädytysvakio-para $(\alpha H, c)$ tai para $(H, c/\alpha)$. Skaalausvakiolla α ja jäädytysvakiolla c ei siis ole itsenäistä merkitystä – ainoastaan niiden keskinäisellä suhteella. Ei siten tarvitse erikseen tutkia lukujen c ja α vaikutusta simulointien lopputulokseen: riittää tehdä simulointeja jäädytysvakion c eri arvoilla. Kun c korvataan uudella vakiolla c/α , $\alpha < 1$ siis mahdollistaa ketjulle alkuperäistä todennäköisempiä energiaa kasvattavia siirtymiä päinvastoin kuin $\alpha > 1$.

Ennen simulointikokeita tarkastellaan vielä energiaa kasvattavan siirtymän suuruuden konkreettista vaikutusta hyväksymisen todennäköisyyteen logaritmiselle jäädytysaikataululle.

Olkoon $\Delta(x, y) > 0$ energiaa kasvattavaan (hetkellä $n \geq 1$ tapahtuvaan) siirtymään $x \rightarrow y$ liittyvä energian muutos; ei oteta nyt kantaa siihen, mihin energiafunktioon em. muutos liittyy. Metropolis-jäädytyksen algoritmi hyväksyy siirtymän

tällöin todennäköisyydellä

$$\exp\left(-\frac{\Delta(x,y)}{T(n)}\right) = \exp\left(-\frac{\Delta(x,y)}{c} \log(n+1)\right) = (n+1)^{-\Delta(x,y)/c}.$$

Määritellään $p(q, n) := (n+1)^{-q}$, missä $q = \Delta(x, y)/c$, jolloin voidaan kätevästi taulukoida eri suhteisiin $q \in (0, 1]$ ja iteraatiomääriin $n \geq 1$ liittyviä hyväksymistodennäköisyyksiä (Taulukko 2).

TAULUKKO 2. Energiaa kasvattavan siirtymän hyväksymisen todennäköisyydet $p(q, n)$ iteraatioiden määrän n ja suhteen q funktiona.

q	iteraatioiden lukumäärä n							
	100	500	1000	2000	5000	25000	50000	10^5
0.005	0.98	0.97	0.97	0.96	0.96	0.95	0.95	0.94
0.010	0.95	0.94	0.93	0.93	0.92	0.90	0.90	0.89
0.050	0.79	0.73	0.71	0.68	0.65	0.60	0.58	0.56
0.100	0.63	0.54	0.50	0.47	0.43	0.36	0.34	0.32
0.250	0.32	0.21	0.18	0.15	0.12	0.08	0.07	0.06
0.500	0.10	0.04	0.03	0.02	0.01	0.01	0.00	0.00
1.000	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Jos $\Delta(x, y) = c$, siirtymän hyväksymisen todennäköisyys on korkeintaan luokkaa 0.01 ajanhetkestä n riippumatta (luvut on pyöristetty kolmen numeron tarkkuuteen). Jos sen sijaan c on 200-kertainen lukuun $\Delta(x, y)$ nähden ($q = 0.005$), siirtymä hyväksytään lähes varmasti vielä sadantuhannenkin iteraation jälkeen. Käytännön tilanteissa, joissa käytettävissä oleva laskenta-aika on rajoitettu (esimerkiksi $n \leq 10^5$) eikä suppenemista minimienergiatilojen joukkoon voida siten taata, on jäähdytysvakio syytä valita siten, että näiltä ääri-ilmioilta vältytään. Tehdään kuitenkin seuraava havainto: jos $\Delta(x, y) = D$ Lauseen 6.3.1 vakiolle D , algoritmin suppeneminen minimienergiatilojen joukkoon toteutuu kuitenkin *asymptoottisesti* kaikilla Taulukossa 2 listatuilla suhteen q arvoilla, sillä $q \leq 1$ implikoi vaaditun epäyhtälön $c \geq D$.

7.4.2. Simulointikokeita eri jäähdytysvakioiden arvoilla. Tehtiin simulointikokeita eri jäähdytysvakion c arvoilla tilanteissa, joissa käytössä olevien iteraatioiden maksimimäärien n annettiin vaihdella välillä 1000–20000. Alkutilaksi X_0 kussakin simuloinnissa valittiin Kuvan 7.1 konfiguraatio x_0 (jota sovellettiin oletusarvoisesti myös muissa tämän osion 6×8 -tapauksissa). Jokaista paria (c, n) kohden simuloitiin kymmenen polkua, joiden hetken n arvoista $X_n^{(1)}, X_n^{(2)}, \dots, X_n^{(10)} \in E$ taulukoitiin energioiden keskiarvot ja keskihajonnat;

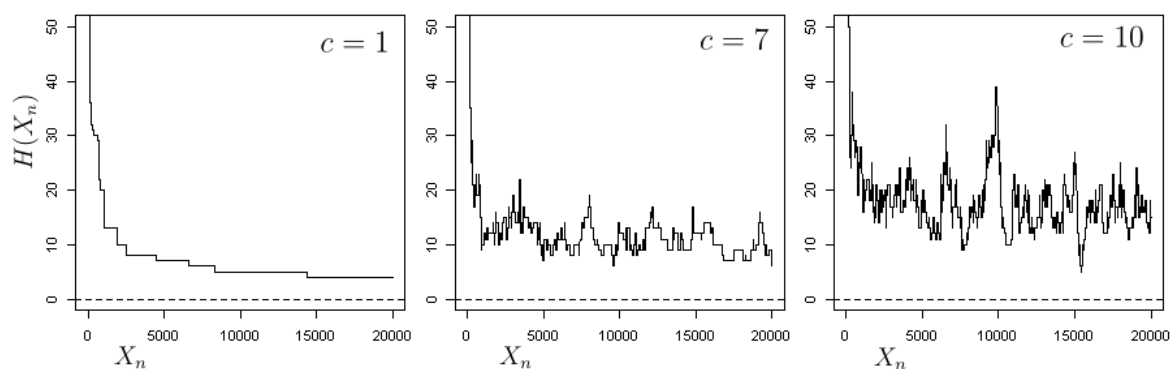
$$\overline{U(X_n)} = \frac{\sum_{i=1}^{10} U(X_n^{(i)})}{10}, \quad s.d.(X_n) = \sqrt{\frac{\sum_{i=1}^{10} \left(U(X_n^{(i)}) - \overline{U(X_n)} \right)^2}{9}}.$$

Tulokset on koottu Taulukkoon 3. Voidaan havaita, että arvoilla $c = 1$ ja $c = 3$ saatiin hyvin samankaltaisia tuloksia. Näillä vakioilla simuloitujen ketjut hyväksyivät käytännössä yksinomaan energiaa vähentäviä siirtymiä. Arvoa $c = 5$ suuremmilla vakioilla erot alkavat näkyä jo selvemmin. Merkillepantavaa on, että jo arvon $c = 8$

TAULUKKO 3. Logaritmisen jäähtytysjonon vakion $c > 0$ vaikutus energiakeskiarvoihin.

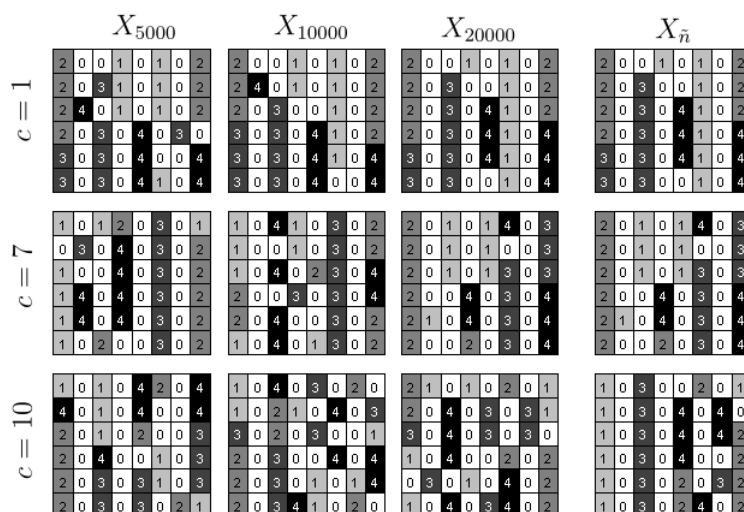
	$n = 1000$	$n = 5000$	$n = 10000$	$n = 20000$
$c = 1$	12.2 ± 4.2	7.3 ± 1.3	4.1 ± 1.4	2.9 ± 1.9
$c = 3$	15.5 ± 5.5	6.0 ± 2.7	3.5 ± 1.4	3.0 ± 1.9
$c = 5$	15.2 ± 3.0	7.7 ± 2.1	6.5 ± 2.5	5.9 ± 2.6
$c = 8$	22.1 ± 5.2	12.7 ± 3.5	14.3 ± 4.8	10.5 ± 3.8
$c = 10$	22.8 ± 4.0	17.9 ± 4.0	15.6 ± 4.3	15.1 ± 4.4
$c = 15$	34.1 ± 9.9	28.1 ± 8.9	22.9 ± 5.1	25.6 ± 9.9

ylittävät jäähtytysvakiot vaikuttavat antavan verraten huonoja tuloksia käytetyillä iteraatiomäärillä. Vakioita $c = 1$, $c = 7$ ja $c = 10$ edustavien esimerkkipolkujen arvoja eri ajanhetkillä on esitetty Kuvassa 7.3. Näitä samoja polkuja vastaavat energiagraafit on esitetty Kuvassa 7.2, josta jäähtytysvakion vaikutus energiaa kasvattaviin hyppyihin käy melko selvästi ilmi.

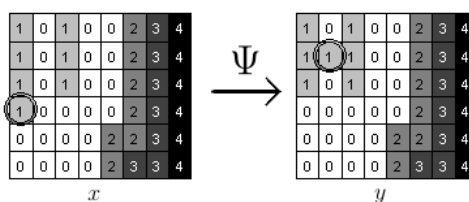
KUVA 7.2. Jäähtytysvakion c arvoja 1, 7 ja 10 vastaavien simuloidulla jäähtytyksellä realisoitujen polkujen energiat aikavälillä $n = 0-20000$.

Muistetaan, että Lauseessa 6.2.5 esiintyvä arvo suppenemisen takaavalle jäähtytysketjulle on $c = \tau\Delta$. Käytetylle mallille voidaan päätellä (ks. Kuva 7.4), että $\Delta = 61$. Siten $\tau\Delta$ on vähintäänkin 61. Viimeistään jäähtytysvakioita $c = 60$ vastaavan ketjun polun energiagraafi (Kuva 7.5, 500000 iteraatiota) osoittaa kouriintuntuvasti, että näin suurista vakioista ei ole välttämättä paljoakaan iloa käytännön kannalta. Koska toisaalta Lauseen 6.3.1 antavaa suppenemisen takaavaa optimaalista vakioita D on hyvin hankala arvioida tarkasti, sillä ainakin periaatteessa pitäisi tuntea jokaisen mahdollisen konfiguraation energiat, hyvien tuloksien saamiseksi jouduttiin turvautumaan yritys ja erehdys -menetelmään.

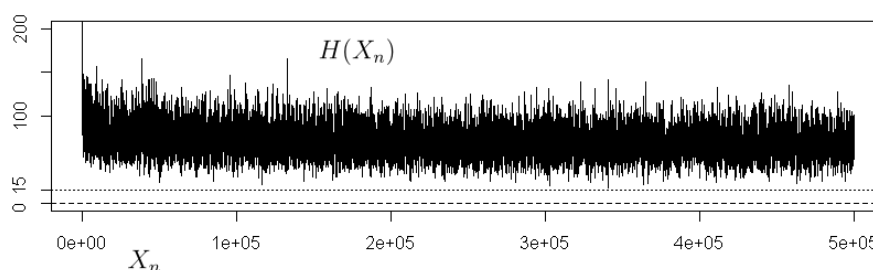
7.4.3. Ehdotusmatriisin ja alkukonfiguraation vaikutus. Kokeiden perusteella havaittiin, että jäähtytysketjun suppenemisnopeus riippuu monista eri tekijöistä. Edellisessä alakohdassa nähtiin, kuinka dramaattinen vaikutus pelkästään logaritmisestä jäähtytysaikataulun vakion valinnalla voi olla lopputulosten kannalta. Myös tenttialin koolla ja koostumuksella havaittiin olevan selviä vaikutuksia algoritmin etenemiseen. Esimerkiksi kokoluokkaa 5×5 oleville tenttisaleille suppeneminen globaaliin



KUVA 7.3. Kuvan 7.2 energiograafeja vastaavien polkujen arvoja eri ajanhetkillä. Energiat järjestyksessä vasemmasta yläkulmasta lähtien: 7, 5, 4, 4; 7, 12, 6, 6; 16, 27, 15, 5. Konfiguraatiot $X_{\tilde{n}}$ edustavat aikavälillä 0 – 20000 saavutettuja pienimmän energian tiloja; $\tilde{n} = 20000$, kun $c = 1, 7$; $\tilde{n} = 15440$, kun $c = 10$.



KUVA 7.4. Maksimaalinen yhtä Ψ -siirtymää vastaava energian muutos havainnollistettuna: $\Delta = H(y) - H(x) = 2 \cdot 20 + 4 \cdot 5 + 1 = 61$.



KUVA 7.5. Viidensadantuhannen iteraation polku, $c = 60$. Energiafunktion graafista nähdään, että polun konfiguraatioiden energiat pysyttelevät melko vakaasti 15 yksikön energiatason yläpuolella.

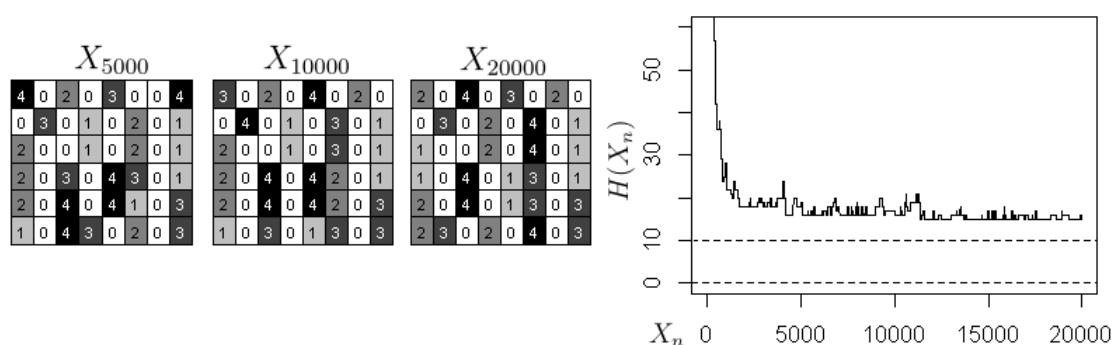
minimienergiatilaan (mikäli lopputulos sellaiseksi voitiin todentaa) saatiin usein toteutettua jo alle viidellä tuhannella iteraatiolla. Sen sijaan 10×10 -kokoisen (epätriviaalin) konfiguraatioavaruuden tilanteessa oli toisinaan vaikeuksia saada hyviä tuloksia edes parilla sadalla tuhannella iteraatioilla (vakioista c riippumatta).

Myös ehdotusmatriisin valinnalla havaittiin olevan silmännähtävää vaikutusta simulointien lopputulosten perusteella. Jo pelkästään Taulukon 3 ja Kuvan 7.2 johdattelemina voitaneen tehdä varovaisia päätelmiä siitä, että käytettäessä siirtymämatriisia Ψ (7.2) energiaa kasvattavien siirtymien tarpeellisuus kelvollisia *likimääräisiä* ratkaisuja etsittäessä joutuu jokseenkin kysenalaiseen valoon.

Päätettiin kokeilla ehdotusmatriisia, jolle energiaa kasvattavien siirtymien merkitys näkyisi selvemmin. Tähän tarkoitukseen valittiin kahdessa *naapurisolmussa* tenttien keskinäisen vaihdoksen tuottava matriisi $\tilde{\Psi}$:

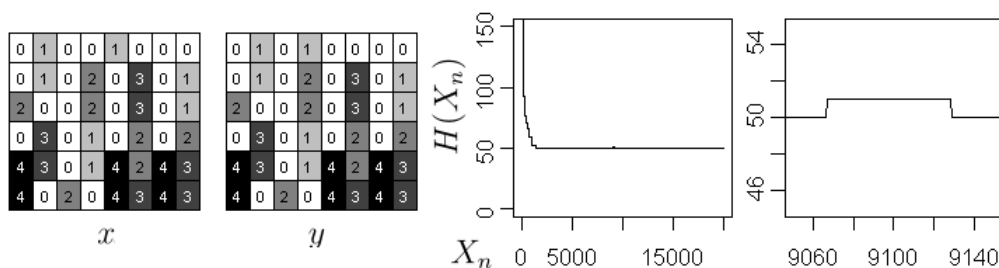
$$\tilde{\Psi}(x, y) = \frac{\#\{(\alpha, \beta) \in S^2 \mid \alpha \sim \beta, \varphi_{(\alpha, \beta)}(x) = y\}}{\#\{(\alpha, \beta) \in S^2 \mid \alpha \sim \beta\}},$$

missä siis $\alpha \sim \beta$ täsmälleen silloin, kun $\|\alpha - \beta\|_{L^2} \in \{1, \sqrt{2}\}$.



KUVA 7.6. Ehdotusmatriisia $\tilde{\Psi}$ soveltamalla simuloitu 20000 iteraation polku; $c = 7$. Energiat järjestyksessä ovat 19, 17 ja 15.

Matriisia $\tilde{\Psi}$ sovellettaessa riski juuttua suurenergiseen lokaaliin minimitilaan vaikutti selvästi kasvavan. Havaittiin myös, että jäähtyisketjuilla tuntui olevan suuria vaikeuksia asettaa saman tentin (Kuva 7.6). Tämä on toisaalta ymmärrettävää, sillä tenttien siirtäminen salin toisesta laidasta toiseen vaati nyt yhden sijaan useita iteraatiota.



KUVA 7.7. Ehdotusmatriisia $\tilde{\Psi}$ soveltamalla simuloitu 20000 iteraation polku "liian pienellä" vakiolla $c = 1$. Tilanteessa on pyritty konkreettisesti havainnollistamaan lokaalin minimikohdan kirous: $X_n = x$ välillä $\{1325, \dots, 20000\}$ lukuunottamatta lyhyttä ajanjaksoa $\{9067, \dots, 9128\}$, jonka aikana puolestaan $X_n = y$. Näiden konfiguraatioiden energiat ovat lähes samat; $H(x) = 50$ ja $H(y) = 51$.

Alkukonfiguraation valinnan merkitys äärellisen ajan simulointien lopputuloksiin vaikutti kokeiden perusteella riippuvan tilanteesta. Mitä pääasiallisesti havaittiin oli se, että alkukonfiguraatio tuhoutui tyypillisesti viimeistään muutaman tuhannen iteraation aikana. Toisaalta esimerkiksi Kuvaa 7.7 katsomalla voidaan päätellä, että lokaaliin minikohtaan juuttumisen kannalta voi myös huonolla alkukonfiguraation valinnalla olla merkitystä.

7.5. Johtopäätökset

Tässä luvussa rakennettiin matemaattinen malli luvun alussa asetetulle tutkimusongelmalle hyödyntämällä pääasiassa kuvankäsittelyyn ja Gibbs-jäähdytykseen liittyviä käsitteitä kuten naapurustot, klikit ja potentiaalit. Vaikka Gibbs-jäähdytyksen soveltaminen ei tullut kyseeseen istumajärjestyksiä edustavan konfiguraatioavaruuden rakenteesta johtuen, pystyttiin näin kuitenkin luontevalla tavalla muodostamaan Metropolis-jäähdytykseen soveltuva, ehtoja I–V vastaava energiafunktio. Siirtymien tekemiseen konfiguraatioiden välillä päädyttiin soveltamaan satunnaisia parittaisia vaihtoja tekevää ehdotusmatriisia Ψ (7.2). Metropolis-jäähdytyksen käyttöönottoa varten tehdyt valmistelut osoittautuivat siis kohtalaisen vaivattomiksi.

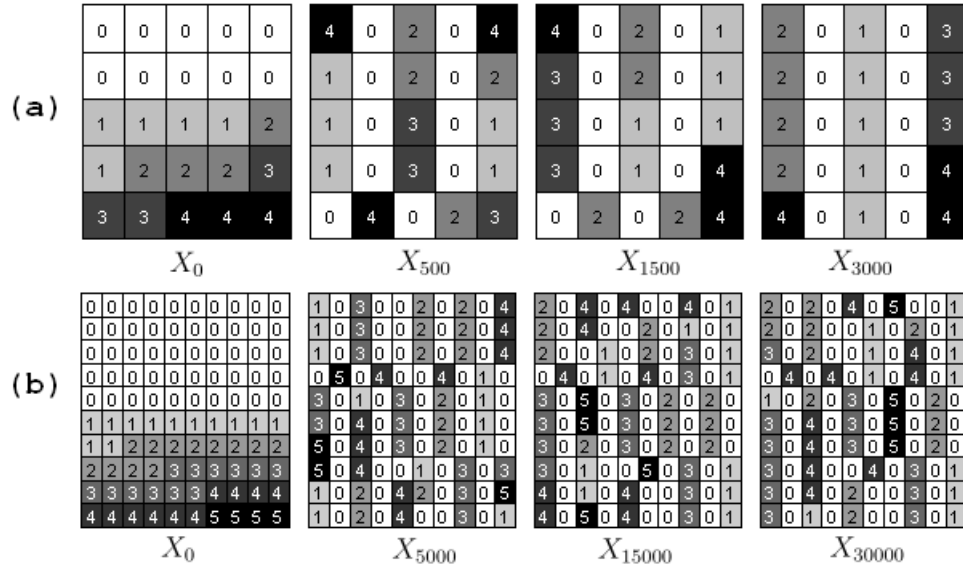
Itse minimointi ja jäähdytyksen soveltaminen käytännössä osoittautui kuitenkin ongelmalliseksi, sillä jo 10×10 -kokoisten tenttisalien kohdalla jouduttiin ongelmiin. Siinä missä pienten, 5×5 -kokoluokkaa olevien salien globaali optimointi onnistui melko vaivattomasti äärellisessä ajassa (Kuva 7.8 (a)), isompien, epätriviaalien salien yhteydessä pelkkää R-ohjelmaa käyttämällä globaaleja minimikohtia kyetty järkeväsä ajassa tavallisesti löytämään.³

Sen sijaan myös suurempien salien tapauksissa onnistuttiin simuloimaan melko hyvältä näyttäviä *likimääräisiä* matalan energian istumajärjestyksiä (Kuva 7.8 (b)). Tällöin jouduttiin kuitenkin käyttämään sellaisia jäähdytysvakioita, joilla asymptoottista suppenemista ei voitu enää taata. Vaikka Lukujen 5 ja 6 jäähdytystä koskevat suppenemislauseet ovat matemaattisessa mielessä ansiokkaita tuloksia, useissa konkreettisisä tilanteissa teoreettisten jäähdytysaikataulujen noudattaminen takaa ainoastaan erittäin hitaan suppenemisen [1, s. 4]. Simuloidun jäähdytyksen nopeuttamiseksi onkin olemassa monenlaisia keinoja, mikäli ollaan valmiita luopumaan asymptoottiseen suppenemiseen liittyvästä takuusta. Koska toisaalta monet näistä menetelmistä ovat monimutkaisia ja niiden kirjo on kohtalaisen suuri, ei niiden tarkempaan analysoimiseen saati soveltamiseen ollut tämän tutkielman puitteissa mahdollisuutta.

Metropolis-jäähdytyksen käyttämiseen tenttisaliongelman yhteydessä liittyi siis sekä hyviä että huonoja puolia. Tilanteissa, joissa tenttialin rakenne oli yksinkertainen (vähän tenttijöitä suhteessa tyhjien paikkojen lukumäärään) saattoi olla jopa helpompaa löytää ratkaisu kynällä ja paperilla simuloinnin sijaan. Toisaalta tilanteissa, joissa salijärjestysten todentaminen matalaenergisiksi pelkällä silmällä oli vaikeaa (paljon tenttijöitä tyhjien paikkojen lukumäärään verrattuna), simuloitu jäähdytys vaikutti tarjoavan joustavan ja melko pienellä vaivalla toteutettavan menetelmän – vähintäänkin likimääräisten ratkaisujen löytämiseksi. Simulointikokeita tehdessä kävi

³Koska silmukoiden ajaminen R-kielillä on sen rakenteesta johtuen hidasta, laskenta-ajan säästämiseksi on tapana kirjoittaa ne erillisiksi, R:n kautta ajettaviksi C/FORTRAN-kielisiksi ohjelmiksi [23].

kuitenkin selväksi, että yksinkertaisimmillaankin menetelmän menestyksekkäs soveltaminen vaatii kärsivällisyyttä ja sopivien parametrien valitsemiseen liittyvää paneutumista.



KUVA 7.8. Kuva (a): 5×5 -sali, 4 tenttiä, 3000 iteraatiota, $c = 4$. Energiat ovat $H(X_0) = 206$, $H(X_{500}) = 7$, $H(X_{1500}) = 3$, $H(X_{3000}) = -8$. Kuva (b): 10×10 -tenttisali, 5 tenttiä, 30000 iteraatiota, $c = 6$. Energiat: $H(X_0) = 936$, $H(X_{5000}) = -11$, $H(X_{15000}) = -19$, $H(X_{30000}) = -22$. Todetaan, että hyvän konfiguraation aikaansaamiseksi kuvan (b) tilanteessa tarvittiin noin kymmenkertainen määrä iteraatioita kuvan (a) tapaukseen verrattuna.

Kirjallisuutta

- [1] R. AZENCOTT, *Simulated Annealing: Parallellization Techniques*. Wiley-Interscience series in discrete mathematics, 1992.
- [2] J. E. BESAG, *On the Statistical Analysis of Dirty Pictures*. Journal of the Statistical Society, Series B, vol. 48, no.8, s. 259–302, 1986.
- [3] P. BRÉMAUD, *Markov Chains, Gibbs Fields, Monte Carlo Simulation and Queues*. Texts in Applied Mathematics, Vol. 31, Springer, 1999.
- [4] V. ČERNÝ, *Thermodynamical Approach to the Traveling Salesman Problem: An Efficient Simulation Algorithm*. Journal of Optimization Theory and Applications, vol. 45, no. 1, 1985.
- [5] R. L. DOBRUSHIN, *Central limit theorem for non-stationary Markov chains, I*. Theory of Probability & Its Applications, Vol. 1, No. 1, s. 65–80, 1956.
- [6] D. GAMERMAN *Markov Chain Monte Carlo: Stochastic simulation for Bayesian inference*. Texts in Statistical Science, Chapman & Hall, 1997.
- [7] A. E. GELFAND, A. F. M. SMITH, *Sampling-Based Approaches to Calculating Marginal Densities*. Journal of the American Statistical Association, 85, s. 398–409, 1990.
- [8] S. GEMAN, D. GEMAN, *Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images*. IEEE Proc. Pattern Analysis and Machine Intelligence, vol. 6, s. 721–741, 1984.
- [9] C. GEISS, S. GEISS, *An Introduction to Probability Theory*. Luentomoniste, Jyväskylän yliopisto, Matematiikan ja tilastotieteen laitos, 2009.
- [10] G. R. GRIMMETT, *A Theorem About Random Fields*. Bull. London Math. Soc, 5, s. 81–81, 1973.
- [11] B. HAJEK, *Cooling Schedules for Optimal Annealing*. Mathematics of Operations Research, 1985, uudistettu versio, 1986.
- [12] W. K. HASTINGS, *Monte Carlo Sampling Methods Using Markov Chains and Their Applications*. Biometrika 57, no. 1, s. 97–109, 1970.
- [13] R. A. HORN, C. R. JOHNSON, *Matrix Analysis*. Cambridge University Press, 1985. 3. painos, 1990.
- [14] D. L. ISAACSON, R. W. MADSEN, *Markov Chains Theory and Applications*. Wiley, 1976, 4. painos, 1976.
- [15] O. HÄGGSTRÖM, *Finite Markov Chains and Algorithmic Applications*. Cambridge University Press, 2002, 5. painos, 2008.
- [16] L. KAHANPÄÄ, *Funktionaalianalyysi, suoraviivaista ajattelua – osa II*, luentomoniste. Jyväskylän yliopisto, Matematiikan ja tilastotieteen laitos, 2006.
- [17] A. KANKAINEN, *Todennäköisyyslaskenta, osa A*, luentomoniste. Jyväskylän yliopisto, Matematiikan ja tilastotieteen laitos, 2004.
- [18] S. KIRKPATRICK, C. D. GELATT, M. P. VECCHI, *Optimization by Simulated Annealing*. Science vol. 220, no. 4598, s. 671–680, 1983.
- [19] P.J.M. VAN LAARHOVEN, E.H.L. AARTS, *Simulated Annealing: Theory and Applications*. D. Reidel Publishing Company, 1987.
- [20] J. LEMON, *Plotrix: a Package in the Red Light District of R*. R-News, vol. 6/4, s. 8–12, 2006.
- [21] D. A. LEVIN, Y. PERES, E. L. WILMER, *Markov Chains and Mixing Times*. American Mathematical Society, 2008.
- [22] N. METROPOLIS, A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER, *Equation of State Calculations by Fast Computing Machines*. Journal of Chemical Physics, vol. 21, no. 6, 1953.

- [23] M. MYLLYMÄKI, *R-ohjelmointi*, luentomoniste. Jyväskylän yliopisto, Matematiikan ja tilastotieteen laitos, 2009.
- [24] A. PENTTINEN *Stokastinen simulointi*, luentomoniste. Jyväskylän yliopisto, Matematiikan ja tilastotieteen laitos, versio 2, 2010.
- [25] V. PURMONEN, *Lineaarinen algebra ja geometria 1*, luentomoniste. Jyväskylän yliopisto, Matematiikan ja tilastotieteen laitos, 2007.
- [26] R CORE TEAM, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, version 2.15.1 (2012-06-22), 2012.
- [27] A. N. SHIRYAEV, *Probability*. Springer, 1984. 2. painos, 1996.
- [28] P. SUOMELA, *Construction of Nearest Neighbour Systems*. Annales Academiae Scientiarum Fennicae, I. Mathematica Dissertationes 10, Suomalainen tiedeakatemia, 1976.
- [29] J. VÄISÄLÄ, *Topologia I*. Limes ry, 4. korjattu painos, 2007.
- [30] G. WINKLER, *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods*. Stochastic Modelling and Applied Probability, Vol. 27, Springer, 2nd Edition, 2003.