kandi työotsikkko

Topias Karjalainen

27. maaliskuuta 2020

Sisältö

1	Johdanto				2	
2	Teoriaa				3	
	2.1	Perusn	määritelmiä		3	
	2.2	Marko	ovin ketjut		4	
			Äärellinen tilajoukko			
		2.2.2	Ääretön jatkuva tilajoukko			
3	Markov Chain Monte Carlo				8	
	3.1	Gibbsi	in otanta-algoritmi		8	
	3.2	Metro	polis–Hastings algoritmi		8	
			Ehdotusjakauman valinnasta			
	3.3	Yleisiä	a käytäntöjä MCMC-menetelmissä		13	
	3.4	Konve	ergenssi ja Diagnostiikka		13	
			Gelmanin \hat{R}			
		3.4.2	ESS			
4	Lop	opusana	at		15	

Johdanto

Tilastotieteissä frekventistinen koulukunta oli pitkään vallitseva koulukunta. Viimeaikoina kuitenkin suosiotaan on kasvattanut Bayesilainen koulukunta. Aiemmin Bayesiläinen päättely ei päässyt leviämään, sillä toisin kuin frekventistinen koulukunta, Bayesiläisyys ei tarjonnut suurinpaan osaan kysymyksiä analyyttisiä ratkaisuja. Vasta tietokoneiden aikakautena Markovin ketju Monte Carlo -menetelmät (MCMC-menetelmät) ovat antaneet mahdollisuuden ratkaista posteriori-jakaumat monimutkaisemmilta malleilta.

Monte Carlo menetelmän kehitteli 50-luvulla Los Alamosissa työskennelleet Nicholas Metropolis, Stanislav Ulam ja yleisnero John von Neumann. Yleinen määritelmä Monte Carlo menetelmälle on toistuva satunnainen arvojen arpominen. Yksinkertainen esimerkki Monte Carlo simuloinnista on esimerkiksi π :n arvon estimointi arpomalla sattumanvaraisesti pisteitä tasosta, ja laskemalla kuinka moni niistä on ympyrän säteen sisällä.

Markovin ketjut ovat stokastisia prosesseja, jotka on nimetty venäläisen matemaatikon Andrey Markov'n mukaan.

MCMC-menetelmiä käytetään nykyään enimmäkseen tilastotieteessä laaja-alaisesti perinteisestä parametriestimoinnista aina vaalitulosten ennustukseen. Niillä on myös sovelluksia biologiassa, fysiikassa ja kielitieteissä.

Teoriaa

2.1 Perusmääritelmiä

Määritellään ensiksi todennäköisyys.

Määritelmä 2.1. σ -algebra. Olkoot Ω mielivaltainen epätyhjä joukko. Sigma-algebra perusjoukolla Ω on sen osajoukkojen joukkoperhe \mathcal{F} , joka toteuttaa ehdot:

- 1. $\emptyset \in \mathcal{F}$
- 2. jos $A \in \mathcal{F}$, $niin A^c \in \mathcal{F}$
- 3. jos jos $A_k \in \mathcal{F}$, kaikilla $k \in K$, missä K on numeroituva joukko, niin $\bigcup_{k \in K} A_k \in \mathcal{F}$

Määritelmä 2.2. Kuvaus P liittää kuhunkin tapahtumaan A todennäköisyyden, joka on luku suljetulla välillä [0,1] ja sille pätee:

- 1. $P(\Omega) = 1$
- 2. Jos Aon tapahtuma, niin sen komplementtitapahtuman A^c todennäköisyys on $\mathbf{P}(A^c)=1-\mathbf{P}(A)$
- 3. Jos $(A_k)_{k\in\mathbb{N}}$ ovat erillisiä tapahtumia, niin

$$\mathbf{P}(\bigcup_{k\in\mathbb{N}}A_k)=\sum_{k\in\mathbb{N}}\mathbf{P}(A_k)$$

Määritelmä 2.3. Kolmikkoa $(\Omega, \mathcal{F}, \mathbf{P})$ kutsutaan todennäköisyysavaruudeksi.

Määritelmä 2.4. Satunnaismuuttuja X on (lähes) mielivaltainen kuvaus $X:\Omega\to S,$ jossa S on tilajoukko.

2.2 Markovin ketjut

2.2.1 Äärellinen tilajoukko

Määritelmä 2.5. Jono $(X_n: n=1,2,3,...)$ satunnaismuuttujia on diskreettiaikainen stokastinen prosessi.

Merkintä 2.6. Merkitään stokastista prosessia merkinnällä $\{X_n\}$

Määritelmä 2.7. Stokastinen prosessi $\{X_n\}$ on *Markovin ketju*, jos kaikilla alkuhetkillä m, n ja tiloilla $i, j \in S$ on voimassa

(2.8)
$$\mathbf{P}(X_{n+1} = j | X_0 = i_0, X_1 = i_1, ..., X_{n-1} = i_{n-1}, X_n = i) = \mathbf{P}(X_{n+1} = j | X_n = i)$$

ja siirtymätodennäköisyyksille on voimassa

(2.9)
$$p_{ij} = \mathbf{P}(X_{n+1} = j | X_n = i) = \mathbf{P}(X_{m+1} = j | X_m = i)$$

Yhtälöä 2.8 kutsutaan Markovin-ehdoksi ja yhtälöä 2.9 taas kutsutaan stationarisuusehdoksi, mikä tarkoittaa, että siirtymätodennäköisyys tilojen i ja j välillä ei riipu ajasta m ja n, vaan pelkästään tiloista i ja j.

Määritelmä 2.10. Satunnaismuuttujan X_0 jakaumaa kutsutaan alkujakaumaksi.

Lause 2.11. Ajanhetkellä $n \ge 1$ polun $(i_0, ... i_n)$ todennäköisyys on

(2.12)
$$P(X_0 = i_0, ..., X_n = i_n) = p_{i_0} p_{i_0, i_1} p_{i_1, i_2} ... p_{i_{n-1}, i_n}$$

Todistus. Käyttäen ehdollisen todennäköisyyden kaavaa, saadaan 2:lle tapahtumalle

$$P(A_0, A_1) = P(A_0)P(A_1|A_0)$$

Jos tapahtumia on kolme, saadaan

$$\mathbf{P}(A_0, A_1, A_2) = \mathbf{P}(A_0)\mathbf{P}(A_1|A_0)\mathbf{P}(A_2|A_1, A_0)$$

neljä

$$P(A_0, A_1, A_2, A_3) = P(A_0)P(A_1|A_0)P(A_2|A_1, A_0)P(A_3|A_2, A_1, A_0)$$

ja n

(2.13)
$$\mathbf{P}(A_0, ..., A_n) = \mathbf{P}(A_0)\mathbf{P}(A_1|A_0)...\mathbf{P}(A_n|A_{n-1}, ...A_0)$$

Tämä on yleinen ehdollinen todennäköisyys. Merkataan $A_n := (X_i = i_n)$. Koska käsittelemme Markovin ketjua, niin yhtälö 2.8 pätee, jolloin yhtälöstä 2.13 saadaan

$$\mathbf{P}(X_0 = i_0, ..., X_n = i_n) = \mathbf{P}(X_0 = i_0)\mathbf{P}(X_1 = i_1|X_0 = i_0)...\mathbf{P}(X_n = i_n|X_{n-1} = i_{n-1})$$

jossa $\forall n=0,1,2,...,n: \mathbf{P}(X_n=i_n|X_{n-1}=i_{n-1})$ on siirtymätödennäköisyys p_{i_{n-1},i_n} jolloin tulos seuraa substituoimalla termit.

Merkintä 2.14.

(2.15)
$$p_{ij}^{(m)} := \mathbf{P}(X_m = j | X_0 = i), \ i, j \in S, m \in T$$

on siirtymätodennäköisyys tilasta i tilaan j, kun aikaa kuluu m yksikköä.

Määritelmä 2.16. Siirtymämatriisi on matriisi

(2.17)
$$\mathbf{P}^{(m)} := (p_{ij}^{(m)})_{i,j} = \begin{pmatrix} p_{00}^{(m)} & p_{01}^{(m)} & \dots & p_{0n}^{(m)} \\ p_{10}^{(m)} & p_{11}^{(m)} & \dots & p_{1n}^{(m)} \\ \vdots & & \ddots & \vdots \\ p_{n0}^{(m)} & p_{n1}^{(m)} & \dots & p_{nn}^{(m)} \end{pmatrix}$$

Lause 2.18. Kaikilla ajanhetkillä on voimassa

$$(2.19) P^{(m)} = P^m$$

Todistus. Todistus on melko pitkä, joten ohitetaan se.

Määritelmä 2.20. Todennäköisyysjakauma $\pi = (\pi)_{i \in S}$ on Markovin ketjun $\{X_n\}$ tasapainojakauma, jos

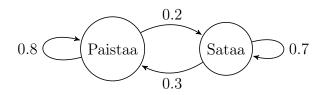
(2.21)
$$\sum_{i \in S} \pi_i p_{ij} = \pi_j, \forall j \in S$$

Yhtälö 2.21 voidaan kirjoittaa myös muotoon

Lause 2.23. Äärellisellä Markovin ketjulla on aina jokin tasapainojakauma π .

Määritelmä 2.24. Markovin ketju on $k\ddot{a}\ddot{a}ntyv\ddot{a}$, jos löytyy sellainen TN-jakauma $\lambda=(\lambda_i)_{i\in S},$ että

$$(2.25) \lambda_i p_{ij} = \lambda_j p_{ji}, \forall i, j \in S$$



Kuva 2.1: Esimerkki 2.27

Lause 2.26. Jos Markovin ketju on kääntyvä, niin $\lambda = \pi$ on sen tasapainojakauma.

Esimerkki 2.27. Pohditaan lyhyttä esimerkkiä, jossa tilajoukko on $S = {\text{"sataa", "paistaa"}}$. Määritellään siirtymätodennäköisyydet siirtymämatriisilla

$$\mathbf{P}^{(1)} = \begin{pmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{pmatrix}$$

Tämä voidaan visualisoida kuvan 2.27 mukaisesti. Ketju on äärellinen, joten sillä on tasapainojakauma. Yhtälö 2.22 implikoi, että jakauma π on siirtymämatriisin **P** vasen ominaisvektori (π^T **P** = $\lambda \pi^T$, jossa λ = 1). Tämä voidaan ratkaista numeerisesti, ja ratkaisu on π^T = (0.4, 0.6). Helposti nyt nähdään, että 2.22 pätee.

2.2.2 Ääretön jatkuva tilajoukko

Kun Markovin ketjun tilajoukko S ei olekkaan rajattu (esimerkiksi, jos halutaan simuloida normaalijakaumasta, joka voi saada minkä vain arvon väliltä $(-\infty, \infty)$), niin teoria muuttuu hieman. Suurin osa tuloksista pätee pienin muutoksin, mutta niiden todistaminen on hankalaa ja ylittää kanditason. Esitetään kuitenkin tarvittavat perustulokset.

Määritelmä 2.28. Kun S on rajoittamaton, siirtymämatriisi on parasta ajatella kuvauksena $T: S \times S \to [0,1]$, joka kuvaa tilaparin $x,y \in S$ todennäköisyydeksi T(x,y)

Määritelmä 2.29. Iistymätodennäköisyys $p_{ij}^{(n)}$ voidaan kirjoittaa siirtymätiheytenä $T^{(n)}(x,y)$, jolle pätee

(2.30)
$$\int_{S} T(x,y)dy = 1$$
 ja $\int_{S} T^{(n)}(x,y)dy = 1, \forall n \ge 1$

Määritelmä 2.31. Jakauma π on Markovin ketjun $\{X_n\}$ kun S on jatkuva, tasapainojakauma jos

(2.32)
$$\pi(y) = \int_{S} \pi(x)T(x,y)dx$$

Määritelmä 2.33. Markovin ketju jatkuvassa S:ssä on kääntyvä, jos on olemassa

$$\pi(x)T(x,y) = \pi(y)T(y,x), \forall x, y \in S$$

Lause 2.35. Jos Markovin ketju $\{X_n\}$ on kääntyvä ja tilajoukko S on jatkuva, niin π on sen tasapainojakauma.

Todistus.Yhtälön 2.30 mukaan $\int_S T(y,x) dx = 1,$ joten

(2.36)
$$\int_{S} \pi(x)T(x,y)dx = \int_{S} \pi(y)T(y,x)dx = \pi(y)\int_{S} T(y,x)dx = \pi(y)$$

Markov Chain Monte Carlo

3.1 Gibbsin otanta-algoritmi

Gibbsin otanta-algoritmi on tapa simuloida Bayesiläistä moniulotteista posteriorijakaumaa (eli ulottuvuuksia vähintään 2), kun suora otanta on hankalaa. Algoritmi on nimetty amerikkalaisen fyysikon, Josiah Willard Gibbs'n (1839-1903) mukaan, mutta sen todellinen kehittäjä on veljekset Donald Geman (1943-) ja Stuart Geman (1949-) vuonna 1984 artikkelissa Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images

Määritelmä 3.1. Olkoot θ parametrivektori, joka jaetaan d:hen osaan tai osavektoriin, eli $\theta = (\theta_1, \theta_2, ..., \theta_d)$. Gibbsin otanta-algoritmi määritellään seuraavanlaisesti:

- 1. Valitaan θ :n osavektoreille järjestys.
- 2. Arvotaan uusi tila jokaiselle osavektorille θ_j ehdollistamalla se jokaiselle muulle parametrille, eli vedetään arvot $\theta = (\theta_1, \theta_2, ..., \theta_d)$ jakaumista

$$(3.2) p(\theta_i | \theta_{-i,n-1}, y)$$

missä $\theta_{-j,n-1}$ on kaikki muut θ :n komponentit paitsi js komponentti, näiden tämänhetkisillä arvoilla eli

$$\theta_{-j,n-1} = (\theta_{1,n},...,\theta_{j-1,n},\theta_{j+1,n-1},...,\theta_{d,n-1})$$

3.2 Metropolis-Hastings algoritmi

Metropolis–Hastings algoritmi on kehittelijöidenssä Nicholas Metropolisksen (1915-1999) ja Wilfred Keith Hastings:n (1930-2016) mukaan nimetty MCMC-menetelmä, jolla voidaan simuloida Bayesiläisessä analyysissa käytettäviä posteriori jakaumia myös silloin kun tiheys on mahdotonta määrittää analyyttisesti.

Algoritmin pohjan kehitti Stanislav Ulam ja Metropolis työskennellessään Los Alamosissa ja myöhemmin Metropolis kehitteli nykyään Metropolis-algoritmina tunnettua

algoritmiä ja esittelivät sen artikkelissa Equation of state calculations by fast computing machines [4]. Tämä versio algoritmista vaati, että pian esiteltävä ehdotusjakauma on symmetrinen. Myöhemmin Hastings laajenti algoritmin koskemaan myös epäsymmetrisiä ehdotusjakaumia artikkelissa Monte Carlo Sampling Methods Using Markov Chains and Their Applications

Merkintä 3.3. TN-jakauma $J_n(\cdot|\cdot)$ on niin sanottu *ehdotusjakauma* (*proposal distribution*, *jumping distribution*), josta *MH-algoritmissa* arvotaan ehdotus tila.

Määritelmä 3.4. Metropolis-Hastings algoritmi on seuraavanlainen

- 1. Valitaan aloitus tila θ_0 ja asetetaan n=0
- 2. Generoidaan kandidaatti tila θ' satunnaisesti jakaumasta $J_n(\theta'|\theta_{n-1})$
- 3. Lasketaan tiheyksien tai todennäköisyyksien suhde

$$r = \frac{p(\theta'|y)/J_n(\theta'|\theta_{n-1})}{p(\theta_{n-1}|y)/J_n(\theta_{n-1}|\theta')}$$

4. Asetetaan

$$\theta_t = \begin{cases} \theta', \text{todennäköisyydellä} & \min(r, 1) \\ \theta_{t-1}, \text{muuten} \end{cases}$$

Jossa $J_t(\theta'|\theta^{t-1})$ on ns. ehdotusjakauma (eng. proposal distribution).

Lause 3.5. Määritelmän 3.4 algoritmi tuottaa Markovin ketjun jolla on uniikki tasapainojakauma, ja jonka tasapainojakauma on posteriorijakauma $p(\theta|y)$, jossa y on data.

Todistus. Ohitamme todistuksen, että kyseessä Markovin ketju jolla yksi tasapainojakauma, mutta todistamme toisen osan, eli että tasapainojakauma on haluttu $p(\theta|y)$ eli posteriori jakauma. Todistus nojautuu Markovin ketjun kääntyvyysominaisuuteen (2.24 ja 2.33), eli

(3.6)
$$T(\theta_n|\theta_{n-1})p(\theta_{n-1}|y) = T(\theta_{n-1}|\theta_n)p(\theta_n|y)$$

joka on siis riittävä ehto tasapainojakauman olemassaololle. Mietitään kahta tapausta: (1) $\theta_n \neq \theta_{n-1}$ ja (2) $\theta_n = \theta_{n-1}$. Tapauksen (2) siirtymä voi tapahtua kahdella tavalla. Joko kohdassa 4. ehdotus θ' hylätään, tai se hyväksytään, mutta osutaan sattumanvaraisesti takaisin samaan kohtaan. Kuitenkin selvästi nähdään, että ehto 3.6 pätee tilanteessa (2).

Tilanteessa (1)Siirtymätodennäköisyys pisteestä θ_{n-1} pisteeseen θ_n on

(3.7)
$$T(\theta_n | \theta_{n-1}) = J_n(\theta_n | \theta_{n-1}) \min \left(\frac{p(\theta_n | y) J_n(\theta_{n-1} | \theta_n)}{p(\theta_{n-1} | y) J_n(\theta_n | \theta_{n-1})}, 1 \right)$$

Jota voidaan muokata helposti

(3.8)
$$T(\theta_{n}|\theta_{n-1}) = J_{n}(\theta_{n}|\theta_{n-1}) \min\left(\frac{p(\theta_{n}|y)J_{n}(\theta_{n-1}|\theta_{n})}{p(\theta_{n-1}|y)J_{n}(\theta_{n}|\theta_{n-1})}, 1\right) \\ = \frac{1}{p(\theta_{n-1}|y)} \min\left(p(\theta_{n}|y)J_{n}(\theta_{n-1}|\theta_{n}), p(\theta_{n-1}|y)J_{n}(\theta_{n}|\theta_{n-1})\right)$$

Nähdään kuitenkin, että yhtälön 3.8 alempi yhtäläisyys on symmetrinen eli

$$(3.9) T(\theta_{n-1}|\theta_n) = \frac{1}{p(\theta_n|y)} \min\left(p(\theta_{n-1}|y)J_n(\theta_n|\theta_{n-1}), p(\theta_n|y)J_n(\theta_{n-1}|\theta_n)\right)$$

joten kerrotaan 3.8 termillä $p(\theta_{n-1}|y)$ ja hyödynnetään 3.9 ominaisuutta

$$T(\theta_{n}|\theta_{n-1})p(\theta_{n-1}|y) = \frac{1}{p(\theta_{n-1}|y)} \min \left(p(\theta_{n}|y) J_{n}(\theta_{n-1}|\theta_{n}), p(\theta_{n-1}|y) J_{n}(\theta_{n}|\theta_{n-1}) \right) p(\theta_{n-1}|y)$$

$$= \frac{1}{p(\theta_{n}|y)} \min \left(p(\theta_{n-1}|y) J_{n}(\theta_{n}|\theta_{n-1}), p(\theta_{n}|y) J_{n}(\theta_{n-1}|\theta_{n}) \right) p(\theta_{n}|y)$$

$$= T(\theta_{n-1}|\theta_{n}) p(\theta_{n}|y)$$

Eli myös tapauksessa (1) yhtälö 3.6 pätee.

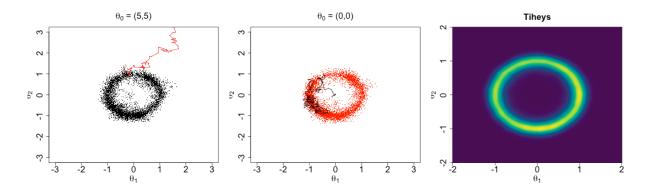
Esimerkki 3.10. Ajatellaan kuvitteellista tapausta, jossa meillä jatkuva kaksiulotteinen todennäköisyysjakauma, jonka tiheysfunktio on

(3.11)
$$p(\theta) \propto \exp(-5|\theta_1^2 + \theta_2^2 - 1|)$$

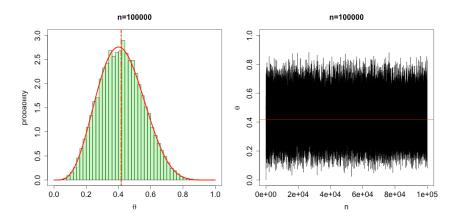
joka muodostaa regasmaisen 2-ulotteisen jakauman. Valitaan ehdotusjakaumaksi $J_n(\theta_n|\theta_{n-1})$ 2d-multinormaalijakauma

$$(3.12) J_n(\theta_n|\theta_{n-1}) \sim N(\theta_{n-1}, \sigma^2 I_2)$$

jossa I_2 on 2x2 yksikkömatriisi ja olkoot $\sigma^2 = 0.01$. Nyt Metropolis Hastings algoritmin avulla voidaan simuloida jakaumaa $p(\theta)$ algoritmilla 3.4. Simuloidaan kaksi Markovin ketjua asettemalla aloitustiloiksi (0,0) ja (5,5), kummastakin 10 000 tilaa. Simuloimme myös 200 000 pistettä aloitusarvolla (0,0), joista luodaan tiheysestimaatti. Tulokset löytyy kuvasta 3.1. Kahdessa ekassa kuvassa viiva on ensimmäisen 250 pisteen polku. Selvästi nähdään, että aloituspisteellä ei ole väliä. Markovin ketjun tasapainojakauma on sama huolimatta aloituspisteestä.



Kuva 3.1: Vasemmalla: (5,5). Keskellä: (0,0) Oikealla: tiheysestimaatti (huomaa eri skaala).



Kuva 3.2: Esimerkin 3.13 tulokset

Esimerkki 3.13. Otetaan toisena esimerkkinä klassinen diskreetti tapaus, jossa oletetaan, että havainnot ovat jauakautuneet Bernoulli-jakauman mukaan, $y_i \sim Bernoulli(\theta)$, ja että priori on tasajakauma. Tällöin tiedetään, että analyyttinen posteriori on

$$p(\theta|y_i) = Beta(\sum_{i=1}^{n} y_i + 1, n - \sum_{i=1}^{n} y_i + 1)$$

Valitaan hieman eksoottinen ehdotusjakauma esimerkin vuoksi:

(3.14)
$$J_n(\theta_n|\theta_{n-1}) \sim \begin{cases} Unif(\theta_{n-1}, 1) & \text{kun } \theta_{n-1} < 0.5 \\ Unif(0, \theta_{n-1}) & \text{kun } \theta_{n-1} \ge 0.5 \end{cases}$$

Toisin kuin esimerkissä 3.10, nyt ehdotusjakauma ei olekkaan symmetrinen.

Oletetaan että, meillä on havainnot (1, 1, 1, 0, 0, 1, 0, 0, 0, 0) ja tarkastellaan sekä analyyttistä että MH-algoritmin tuottamaa jakaumaa ja niiden eroja.

Kuvasta 3.2 nähdään esimerkin tulokset. Huomataan, että vaikka ehdotusjakauma on melko kummallinen, niin kuitenkin tarpeaksi monella iteraatiolla saavutetaan tasapainojakauma. Huomaa, että vasemmassa kuvaajassa on vihreällä simulaatio keskiarvo, ja punaisella analyyttinen keskiarvo, mutta nämä arvot ovat niin lähellä toisiaan, että viivat ovat päällekkäin.

3.2.1 Ehdotusjakauman valinnasta

Kummassakin kappaleen 3.2 esimerkissä valitsimme ehdotusjakauman melko satunnaisesti. Varsinkin esimerkissä 3.13 se on erittäin epätavallinen, mistä syystä hyvän approksimaation saavuttaminen vie todella monta iteraatiota. Yleensä jos haluamme oikeasti tehokkaasti ja ekonomisesti simuloida jakaumia esitetyllä algoritmilla, haluamme valita ehdotusjakauman jollakin järkevällä, systemaattisella tavalla, joka minimoisi tarvittavien iteraatioiden määrän.

Yleisesti ottaen hyvällä ehdotusjakaumalla on muutama ominaisuus[1]

- 1. Kaikilla θ :n arvoilla on helppo arpoa arvo $J(\theta'|\theta)$
- 2. Suhde r on helppo laskea
- 3. Siirtymät ovat tarpeaksi pitkiä. Muuten Markovin Ketju etenee liian hitaasti ja hyvän estimaatin saaminen kestää liian pitkään.
- 4. Siirtymiä ei hylätä liian usein. Muuten Markovin Ketju ei etene vaan seisoo paikallaan.

Lisäksi simulointia voidaan nopeuttaa mm. käyttämällä adaptiivista ehdotusjakaumaa, eli toisinsanoen ehdotusjakaumaa muunnellaan riippuen ketjun liikkeistä.

3.3 Yleisiä käytäntöjä MCMC-menetelmissä

Usein on tapana simuloida useampi kuin yksi ketju kustakin tapauksesta. Tällöin asetetaan näiden ketjujen aloitustilat eriäviksi, jotta saadaan vähennettyä aloitustilan vaikutusta lopputukokseen.

Toinen yleinen käytäntö on jättää ketjun alkupäästä jonkin verran tiloja huomiotta, sillä ketjun alkupäässä ei se ei välttämättä ole vielä saavuttanut tasapainojakaumaa. Tätä kutsutaan *Burn-in periodiksi*.

Joskus taas on hyvä tapa pudottaa joka n:s tila ketjusta. Tilojen tiputtaminen ei vaikuta tasapainojakaumaan, kunhan ketju vain on saavuttanut sen. Tilojen tiputtamista on hyötyä jos mallissa on paljon parametreja, jolloin tietokoneessa voi tulla ongelmia tilan kanssa.

3.4 Konvergenssi ja Diagnostiikka

Kappaleissa 3.1 ja 3.2 esiteltyjen algoritmien kohdalla voi herätä kysymys, että mikä on riittävä määrä iteraatioita, jotta Markovin ketju on saavuttanut tasapainojakaumansa ja otanta on riittävän hyvä aproksimaatio posteriorijaukaumasta. Esimerkiksi kun katsotaan kuvan 3.1 vasemman puolen kuvan punaista polkua, niin voimme sanoa, että se ei ole vielä saavuttanut tasapainojakaumaa sillä tunnemme melko hyvin halutun jakauman, mutta yleensä emme välttämättä osaa sanoa tätä suoraan.

Toinen yleinen ongelma, joka kaipaa pohdintaa on se, että ketjujen sisällä on korrelaatiota, mikä vaikeuttaa päättelyä simuloinnin tuloksista. Otannat eivät siis ole välttämättä täysin riippumattomia.

3.4.1 Gelmanin \hat{R}

Yksi yleisimmistä estimaattoreista, joita käytetään MCMC-metodeissa Markovin ketjujen konvergenssin arvioimiseen on $Andrew\ Gelmanin\ \hat{R}$ [1]. Se mittaa ketjujen sisäistä sekoittumista (eng. mixing) ja erillisten ketjujen välistä sekoittumista.

Määritelmä 3.15. Gelmanin \hat{R} [2][3] lasketaan jakamalla ensin jokin määrä simuloituja ketjua erillisillä aloituspisteillä keskeltä kahtia. Olkoon nyt m ketjujen määrä jaon jälkeen ja n ketjujen pituus. Olkoot $\psi_{ij}(i=1,...,n;j=1,...,m)$ tila i ketjussa j. Nyt merkataan

(3.16a)
$$B = \frac{n}{m-1} \sum_{j=1}^{m} (\overline{\psi}_{.j} - \overline{\psi}_{.})^2, \text{ jossa}$$

$$\overline{\psi}_{.j} = \frac{1}{n} \sum_{i=1}^{n} \psi_{ij}$$

(3.16c)
$$\overline{\psi}_{\cdot \cdot} = \frac{1}{m} \sum_{j=1}^{m} \overline{\psi}_{\cdot j}$$

Ja merkataan

(3.17)
$$W = \frac{1}{m} \sum_{j=1}^{m} s_j^2, \quad jossa \quad s_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\psi_{ij} - \overline{\psi}_{.j})^2$$

Jossa siis B on ketjujen välinen varianssi (between sequence) ja W on ketjujen sisäinen varianssi (within sequence). Näiden painotettuna keskiarvona saadaan estimaattori ψ :n marginaaliselle posteriori varianssille

(3.18)
$$\hat{\sigma}^{+}(\psi|y) = \frac{n-1}{n}W + \frac{1}{n}B$$

Nyt \hat{R} voidaan määritellä

$$\hat{R} = \sqrt{\frac{\hat{\sigma}^+(\psi|y)}{W}}$$

Gelmanin \hat{R} mittaa sitä miten paljon kullakin hetkellä ψ :n jakauma voisi supistua, jos simulaatioiden annettaisiin jatkua loputtomasti. Eli siis kun $\hat{R} \approx 1$, niin voidaan sanoa, että Markovin ketju on todennäköisesti saavuttanut tasapainojakaumansa, sillä tällöin eri pisteitä aloitettujen ketjujen välinen varianssi on vakaantunut.

3.4.2 ESS

Toinen hyödyllinen diagnostiikka MCMC-menetelmissä on niin sanottu effective sample size (ESS), joka mittaa sitä kuinka monta effektiivisesti riippumatonta otosta ketjussa on. Tämä tarkoittaa siis sitä, että se mittaa kuinka paljon ketjun autokorrelaatio vaikuttaa keskivirheeseen verrattuna täysin riippumattomiin otoksiin.

Määritelmä 3.20. ESS lasketaan siten, että lasketaan $\hat{\sigma}^+$ kaavan 3.18 mukaan, ja sitten lasketaan

Loppusanat

Kirjallisuus

- [1] Andrew Gelman. Bayesian Data Analysis. 3. painos. ISBN: 978-1-4398-4095-5.
- [2] Donald Rubin Gelman Andrew. "Inference from Iterative Simulation using Multiple Sequences". Statistical Science 7 (1992), s. 457–511.
- [3] Stephen Brooks Gelman Andrew. "General Methods for Monitoring Convergence of Iterative Simulations". *Journal of Computational and Graphical Statistics* (joulukuu 1998). URL: http://www.stat.columbia.edu/~gelman/research/published/brooksgelman2.pdf.
- [4] Nicholas Metropolis. "Equation of state calculations by fast computing machines" (1953).