

kandi työotsikkko

Topias Karjalainen

11. maaliskuuta 2020

# Sisältö

<b>1</b>	<b>Johdanto</b>	<b>2</b>
<b>2</b>	<b>Teoriaa</b>	<b>3</b>
2.1	Perusmääritelmiä . . . . .	3
2.2	Markovin ketjut . . . . .	4
2.2.1	Äärellinen tilajoukko . . . . .	4
2.2.2	Ääretön jatkuva tilajoukko . . . . .	6
<b>3</b>	<b>Metropolis–Hastings algoritmi</b>	<b>8</b>

# Luku 1

## Johdanto

Tilastotieteissä *frekventistinen* koulukunta oli pitkään vallitseva koulukunta. 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.

Tässä kandidutkielmassa aion selventää MCMC-menetelmien teoriaa, sekä esittää yleisimmät kaksi algoritmia: *Gibbs*- ja *Metropolis–Hastings* algoritmit.

# Luku 2

## Teoriaa

### 2.1 Perusmääritelmiä

Määritellään ensiksi todennäköisyys.

**Määritelmä 2.1.**  *$\sigma$ -algebra.* Olkoot  $\Omega$  mielivaltainen epätyhjä joukko. Sigma-algebra perusjoukolla  $\Omega$  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  $\mathbf{P}$  liittää kuhunkin tapahtumaan  $A$  *todennäköisyyden*, joka on luku suljetulla välillä  $[0,1]$  ja sille pätee:

1.  $\mathbf{P}(\Omega) = 1$
2. Jos  $A$  on 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

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

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

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

## 2.2 Markovin ketjut

### 2.2.1 Äärellinen tilajoukko

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

**Merkintä 2.6.** Merkitään stokastista prosessia merkinnällä  $\{X_n\}_{n \in T}$

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

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

ja *siirtymätodennäköisyyksille* on voimassa

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

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  $n \geq 1$  polun  $(i_0, \dots, i_n)$  todennäköisyys on*

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

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

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

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

$$\mathbf{P}(A_0, A_1, A_2, A_3) = \mathbf{P}(A_0) \mathbf{P}(A_1 | A_0) \mathbf{P}(A_2 | A_1, A_0) \mathbf{P}(A_3 | A_2, A_1, A_0)$$

ja  $n$

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

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

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

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

**Merkintä 2.14.**

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

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

**Määritelmä 2.16.** *Siirtymämatriisi* on matriisi

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

**Lause 2.18.** *Kaikilla ajanhetkillä on voimassa*

$$(2.19) \quad \mathbf{P}^{(m)} = \mathbf{P}^m$$

*Todistus.* Todistus on melko pitkä, joten ohitetaan se.  $\square$

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

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

Yhtälö 2.21 voidaan kirjoittaa myös muotoon

$$(2.22) \quad \pi^T \mathbf{P} = \pi^T$$

**Lause 2.23.** *Äärellisellä Markovin ketjulla on aina jokin tasapainojakauma  $\pi$ .*

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

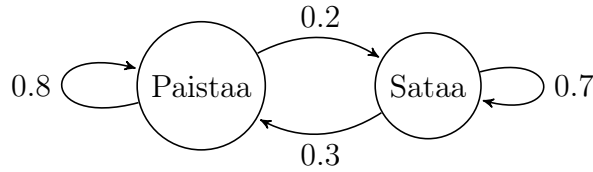
$$(2.25) \quad \lambda_i p_{ij} = \lambda_j p_{ji}, \forall i, j \in S$$

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

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

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

Tämä voidaan visualisoida seuraavanlaisesti:



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

## 2.2.2 Ääretön jatkuva tilajoukko

Kun Markovin ketjun tilajoukko  $S$  ei olekaan 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 kandidatin. Esitetään kuitenkin tarvittavat perustulokset.

**Määritelmä 2.28.** Kun  $S$  on rajoittamaton, siirtymämatriisi on parasta ajatella kuvauksena  $\mathbf{P} : S \times S \rightarrow [0, 1]$ , joka kuvaa tilaparin  $x, y \in S$  todennäköisyydeksi  $p(x, y)$

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

$$(2.30) \quad \int_S p(x, y) dy = 1 \quad \text{ja} \quad \int_S p^{(n)}(x, y) dy = 1, \forall n \geq 1$$

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

$$(2.32) \quad \pi(y) = \int_S \pi(x) p(x, y) dx$$

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

$$(2.34) \quad \pi(x)p(x, y) = \pi(y)p(y, x), \forall x, y \in S$$

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

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

$$(2.36) \quad \int_S \pi(x)p(x, y)dx = \int_S \pi(y)p(y, x)dx = \pi(y) \int_S p(y, x)dx = \pi(y)$$

□



# Luku 3

## Metropolis–Hastings algoritmi

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

**Määritelmä 3.1.** Metropolis–Hastings algoritmi on seuraavanlainen

1. Valitaan aloitus tila  $\theta_0$  ja asetetaan  $t = 0$
2. Generoidaan kandidaatti tila  $\theta'$  satunnaisesti jakaumasta  $J_t(\theta'|\theta_{t-1})$
3. Lasketaan tiheyksien tai todennäköisyyksien suhde

$$r = \frac{p(\theta'|y)/J_t(\theta'|\theta^{t-1})}{p(\theta^{t-1}|y)/J_t(\theta^{t-1}|\theta')}$$

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 3.2.** *Määritelmän 3.1 algoritmi tuottaa Markovin ketjun jolla on uniikki tasapainojakauma, ja jonka tasapainojakauma on haluttu jakauma  $p(\theta|y)$ , 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.  $\square$

# Kirjallisuutta