

Metropolise algoritm kui muutumise vajaduse kriteerium

MARGUS PIHLAK
Tallinna Tehnikaülikool

Elus tuleb sageli ette olukordi, millal tuleb otsustada kas jätkata vana või uut moodi ehk valida *status quo* ja muutuse vahel. Muutus aga on mõistlik vaid juhul, kui see on otstarbekas ehk juhul, kui vanad suhted enam ei toimi. Kuidas mõõta seda otstarbekust? Ühe võimaluse selleks pakub *Metropolise algoritm*. Metropolis-Hastingsi algoritmi rakendati esmakordselt 1953. aastal arvutamaks keemiliste ühendite konfiguratsiooni muutusi. Need arvutused on publitseeritud N. Metropolise teadusartiklis (Metropolis 1953).

Olgu meil k mingi aine molekul konfiguratsiooniga $\theta = (\theta_1, \theta_2, \dots, \theta_k)^\top$. Alternatiivseks positsiooniks olgu $\phi = (\phi_1, \phi_2, \dots, \phi_k)^\top$. Näiteks glükoosi molekuli puhul oleksid need konfiguratsioonid nimedega "tugitool" ja "vann". Kuidas arvutada üleminekut konfiguratsioonist θ konfiguratsiooni ϕ ? Artiklis (Metropolis, 1953) soovitati järgmist meetodit modelleerimaks keeruliste keemiliste süsteemide muutusi:

1. Olgu alghetkel keemiliste ühendite konfiguratsioon $\theta^0 = (\theta_1^0, \theta_2^0, \dots, \theta_k^0)^\top$.
2. Allugu üleminek eelmisest konfiguratsioonist $\theta^{j-1} = (\theta_1^{j-1}, \theta_2^{j-1}, \dots, \theta_k^{j-1})^\top$ uude konfiguratsiooni $\phi = (\phi_1, \phi_2, \dots, \phi_k)^\top$ ühtlasele jaotusele.
3. Leida konfiguratsiooni muutusest tekkinud energia muut ΔE . Aktsepteerida eelmises punktis kirjeldatud konfiguratsiooni muut tõenäosusega

$$\min\{1, \exp(-c\Delta E)\},$$

kus $c = \frac{1}{bT}$, milles b tähistab Boltzmanni konstanti ning T temperatuuri absoluutses (ehk Kelvini) skaalas. Kui muutus toimus, siis $\theta^j = \phi$, vastasel juhul $\theta^j = \theta^{j-1}$.

4. Suurusele j omistatakse suurus $j + 1$ ning minnakse tagasi punkti 2 kuni on saavutatud koondumine.

Metropolise algoritm põhineb tõepära suhtel. See suhe avaldub järgmiselt.

Olgu meil kaks valikut: H_0 ja H_1 . Valik H_0 tähendagu nõ. *status quo* otsust ja valik H_1 muutust. Nende tõenäosusteks olgu $P(H_0)$ ja $P(H_1)$, kusjuures $P(H_0) + P(H_1) = 1$. Viiakse läbi eksperiment, mille käigus toimus sündmus A . Saadi tõepära suhe

$$LR = \frac{P(A | H_1)}{P(A | H_0)}.$$

Teeme tutvust tõepära suhte rakendusega ühe lihtsa näite baasil.

Näide 1. Olgu kassis 8 detaili. Nende detailide kohta esitatakse 2 hüpoteesi:

1. $H_0 =$ "Detailide seas on 2 praakdetaili";
2. $H_1 =$ "Detailide seas on 3 praakdetaili".

Kastist võeti juhuslikult 3 detaili. Juhusliku katse tulemuseks saadi sündmus

$A =$ "Testitud detailidest osutus 1 praagiks ja 2 korras olevaks".

Kuivõrd mõjutab katse tulemus D sündmuste H_0 ja H_1 tõenäosusi? Leiame tõepära suhte

$$LR = \frac{P(A | H_1)}{P(A | H_0)} = \frac{\frac{C_3^1 C_5^2}{C_8^3}}{\frac{C_2^1 C_6^2}{C_8^3}} = \frac{2 \cdot 15}{3 \cdot 10} = 1.$$

Seega ei kalluta katse tulemus A eelistust ei H_0 ega H_1 suunas.

Käsitleme järgnevalt sündmust A mõõtmistulemusena $\mathbf{x} = (x_1, x_2, \dots, x_n)$, mis on juhusliku suuruse X mingi realisatsioon. Olgu selle juhusliku suuruse tihedusfunktsioon $f(x)$. Siis tõepära suhe

$$\text{LR} = \frac{L(\mathbf{x} \mid H_1)}{L(\mathbf{x} \mid H_0)},$$

kus tõepärafunktsioonid on

$$L(\mathbf{x} \mid H_1) = \prod_{i=1}^n f(x_i \mid H_1)$$

ning

$$L(\mathbf{x} \mid H_0) = \prod_{i=1}^n f(x_i \mid H_0).$$

Tõepära suhe leiab rakendust statistilise hüpoteeside kontrollimisel. See rakendus põhineb Neyman-Pearsoni lemmal. Kuna tegemist on olulise tulemusega statistilises analüüsis, siis sõnastame selle lemma teoreemina.

Teoreem. (Neymann-Pearsoni lemma) Olgu meil hüpoteeside paar

$$\begin{cases} H_0 : \theta = \theta_0, \\ H_1 : \theta = \theta_1. \end{cases}$$

Siis tõepära suhte test, mis kummutab hüpoteesi H_0 tingimusel

$$\text{LR} = \frac{L(\mathbf{x} \mid \theta_1)}{L(\mathbf{x} \mid \theta_0)} \geq l,$$

kus

$$P(\text{LR} \geq l \mid H_0) = \beta,$$

on võimsaim statistiline test olulisuse nivool β .

Vaatame lähemalt, kuidas leida teoreemis toodud konstanti l . Osutub, et selle konstandi näol on tegemist vale otuse eest vastutuse määraga. Defineerime järgmised sündmused:

- Sündmus A – tõene on hüpotees H_0 ;
- Sündmus \bar{A} – tõene on hüpotees H_1 ;

- Sündmus B – katsetulemus x sattus nullhüpoteesi piirkonda \mathcal{H}_0 ;
- Sündmus \overline{B} – katsetulemus x sattus sisuka hüpoteesi piirkonda \mathcal{H}_1 .

Olgu juhuslikul suurusel X kaks teadaolevat tinglikku tihedusfunktsiooni:

- $f_0(x)$ - tingimusel, et tõene on hüpotees H_0 ;
- $f_1(x)$ - tingimusel, et tõene on hüpotees H_1 .

Sündmuste korrutistele AB ja \overline{AB} vastab õige otsustus, korrutisele $A\overline{B}$ vastab I liiki viga ning korrutis $\overline{A}B$ on seotud II liiki veaga. Defineerime juhusliku suuruse C , mis iseloomustab valest otsusest tingitud kahju mingis ühikus. Olgu

$$C = \begin{cases} 0, & \text{kui toimus } AB \text{ või } \overline{AB}, \\ c_1, & \text{kui toimus } A\overline{B}, \\ c_2, & \text{kui toimus } \overline{A}B. \end{cases}$$

Juhusliku suuruse C jaotus olgu järgmine: $P(C = 0) = p_0$, $P(C = c_1) = p_1$ ning $P(C = c_2) = p_2$. Tähistagem keskmist kahjumit kui r . Seega

$$r = E(C) = 0p_0 + c_1p_1 + c_2p_2 = c_1p_1 + c_2p_2.$$

Eesmärk on leida optimaalne otsustuse kriteerium, mille puhul keskmine kahjum r oleks minimaalne. Avaldame I ja II liiki vea tegemise tõenäosused (milledeks olgu vastavalt γ_1 ja γ_2) tinglike jaotustiheiduste $f_0(x)$ ning $f_1(x)$ kaudu:

$$\gamma_1 = P(\overline{B} | A) = \int_{\mathcal{H}_1} f_0(x) dx \text{ ja } \gamma_2 = P(B | \overline{A}) = \int_{\mathcal{H}_0} f_1(x) dx.$$

Järgnevalt avaldame tõenäosused p_1 ja p_2 tõenäosuste γ_1 ja γ_2 ning apriorse tõenäosuse $\pi_0 = P(H_0)$ kaudu. Saame

$$p_1 = P(\overline{B}A) = P(A)P(\overline{B} | A) = \pi_0\gamma_1$$

ning

$$p_2 = P(\bar{a}B) = P(\bar{A})P(B | \bar{A}) = (1 - \pi_0)\gamma_2.$$

Kui asendame saadud tõenäosused p_1 ning p_2 keskmise kao valemisse, siis saame

$$r = \pi_0\gamma_1c_1 + (1 - \pi_0)\gamma_2c_2 = \pi_0c_1 \int_{\mathcal{H}_1} f_0(x)dx + (1 - \pi_0)c_2 \int_{\mathcal{H}_0} f_1(x)dx.$$

Seega tuleb meil leida piirkonna \mathcal{H} selline tükeldus piirkondadeks \mathcal{H}_0 ja \mathcal{H}_1 , et keskmine kahjum r oleks minimaalne.

Arvestades jaotustiheduse aditiivsuse omadust

$$\int_{\mathcal{H}} f_0(x)dx = \int_{\mathcal{H}_0} f_0(x)dx + \int_{\mathcal{H}_1} f_0(x)dx = 1,$$

saame

$$\begin{aligned} r &= \pi_0c_1 \left(1 - \int_{\mathcal{H}_0} f_0(x)dx \right) + (1 - \pi_0)c_2 \int_{\mathcal{H}_0} f_1(x)dx = \\ &= \pi_0c_1 + \int_{\mathcal{H}_0} \{ (1 - \pi_0)c_2f_1(x) - \pi_0c_1f_0(x) \}. \end{aligned}$$

Saadud avaldise väärtus on minimaalne sellises nullhüpoteesi \mathcal{H}_0 piirkonnas, mis minimiseerib integraali võrduse parema poole. Selleks aga tuleb piirkonda \mathcal{H}_0 lugeda need ja ainult need x väärtused, mille puhul määratud integraal on negatiivne. Selle negatiivsus on garanteeritud, kui

$$(1 - \pi_0)c_2f_1(x) - \pi_0c_1f_0(x) < 0$$

ehk

$$\frac{f_1(x)}{f_0(x)} < \frac{\pi_0c_1}{(1 - \pi_0)c_2}. \quad (1)$$

Konstant l suhtes (1) on seega järgmine:

$$l = \frac{\pi_0c_1}{(1 - \pi_0)c_2}.$$

See konstant l iseloomustab valest otsusest tingitud maksimaalset lubatud kahju. Kui tõepära suhe on väiksem kui suurus l , siis oleme sunnitud jääma nullhüpoteesi H_0 (*status quo*) juurde. Vastasel juhul aga loeme tõestatuks sisuka hüpoteesi H_1 ehk muutuse.

Metropolise algoritm

Metropolise algoritmi võib kirja panna järgmiselt:

1. Olgu ajahetkel $\theta = 1$ algväärtuseks θ^0 .
2. Genereerime uue väärtuse ϕ tõenäosustiheduse $q(\theta_{j-1})$ põhjal.
3. Leiame vastuvõtmise tõenäosuse

$$\alpha(\theta^{j-1}, \phi) = \min \left\{ 1, \frac{q(\phi)}{q(\theta_{j-1})} \right\}.$$

4. Genereerime juhusliku suuruse U , mis allub ühtlasele jaotusele lõigul $[0; 1]$. Kui selle juhusliku suuruse väärtus $u < \alpha(\theta^{j-1}, \phi)$, siis võtame vastu muutuse ehk $\theta^j = \phi$. Vastasel korral jääb muutus vastu võtmata ehk $\theta^j = \theta^{j-1}$.
5. Omistame $j := j + 1$ ning siirdume algoritmi teise punkti juurde kuni on saavutatud koondumine.

Demonstreerime Metropolise algoritmi ühenduses Bayesi teoreemiga, mille väide tuleneb klassikalisest Bayesi valemist. Selle teoreemi kohaselt on posteriaalne jaotustihedus (ehk järeljaotus) $\pi(\theta)$ proportsionaalne tõepära funktsiooni ning eeljaotuse p tihedusfunktsiooni korrutisega ehk

$$\pi(\theta) \propto L(\theta)p(\theta).$$

Vaatame teist näidet.

Näide 2. Mingit sündmust A katsetatakse sõltumatult n korda. Olgu juhuslik suurus X õnnestunud katsete hulk ning tõenäosus $P(A) = \theta$. Siis

$$P(X = x) = \theta^x (1 - \theta)^{n-x}.$$

Parameetri θ eeljaotuseks olgu beetajaotus $Beta(\alpha, \beta)$. Rakendame Metropolise algoritmi tõenäosusele θ . Allugu muut $\Delta\theta$ normaaljaotusele keskväärtusega 0 ja standardhälbega σ . Olgu esialgne parameeter θ_c ning kavandatav parameeter $\theta_p = \theta_c + \Delta\theta$. Siis saame kirja panna järgmise algoritmi:

- Genereerime juhuslike arvude generaatoriga $\Delta\theta \sim \mathcal{N}(0, \sigma)$. Võtame kavandatavaks parameetriks $\theta_p = \theta_c + \Delta\theta$.

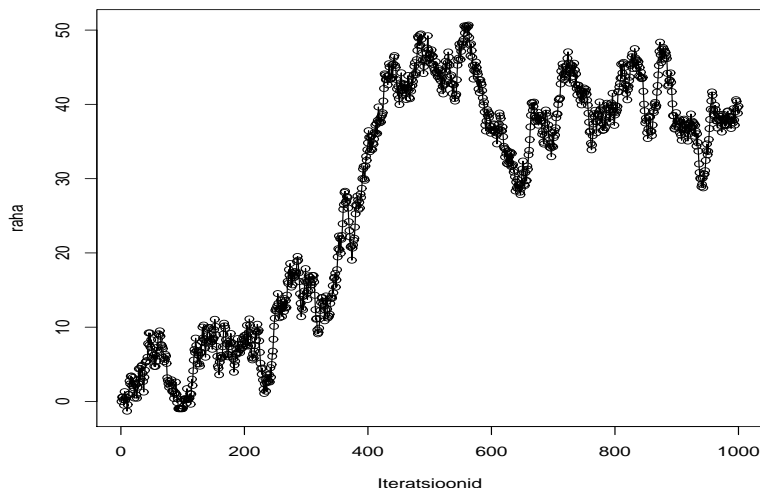
- Moodustame kavandatava parameetri vastuvõtmise tõenäosuse

$$\begin{aligned} \alpha(\theta_c, \theta_p) &= \min \left\{ 1, \frac{P(\theta_p)}{P(\theta_c)} \right\} = \\ &= \min \left\{ 1, \frac{P(X = x | \theta_p)p(\theta_p)}{P(X = x | \theta_c)p(\theta_c)} \right\} = \\ &= \min \left\{ 1, \frac{\theta_p^x (1 - \theta_p)^{n-x} \text{Beta}(\theta_p | \alpha, \beta)}{\theta_c^x (1 - \theta_c)^{n-x} \text{Beta}(\theta_c | \alpha, \beta)} \right\} = \\ &= \min \left\{ 1, \frac{\theta_p^x (1 - \theta_p)^{n-x} \theta_p^{\alpha-1} (1 - \theta_p)^{\beta-1}}{\theta_c^x (1 - \theta_c)^{n-x} \theta_c^{\alpha-1} (1 - \theta_c)^{\beta-1}} \right\}. \end{aligned}$$

- Genereerime juhuslike arvude generaatoriga lõigul $[0; 1]$ ühtlase jaotusega juhusliku suuruse U . Aktsepteerime parameetri väärtust θ_p , kui $u < \alpha(\theta_c, \theta_p)$. Vastasel korral lükkame kavandatud väärtuse tagasi ning jätkame väärtusega θ_c .

Näites 2 toodud algoritmile saab leida mitmeid rakendusi:

- Populatsioonigeneetikas saab sellega uurida mingi geeni alleeli sageduse dünaamikat.
- Sündmuste voo intensiivsuse ajalise muutuse modelleerimisel.
- Samuti saab selle näite algoritmiga modelleerida metsastuse osakaalu muutust aastate lõikes.



Joonis 1. Raha muutus ajas parameetri $a = 2$ korral.

Näide 3. Näidet 3 võib nimetada resursside tekitamise ja kulutamise dünaamikaks. Uurime üht rahakogumise mudelit. Olgu meil alghetkel mingi rahasumma v_0 . Vaatame seda alghetke kui nullseisu ehk $v_0 = 0$. Allugu raha muut Δv ühtlasele jaotusele lõigus $[-a; a]$. Olgu ülemineku tuumaks (ingl. *transitional kernel*) standardsele normaaljaotusele vastav tõenäosus. Nende eelduste korral saame rahahulga muutuse vasuvõtmise kriteeriumiks

$$\alpha(v_i, v_{i+1}) = \min \left\{ 1, \frac{\Phi(v_i + \Delta u) + 0.5}{\Phi(v_i) + 0.5} \right\},$$

$$i = 0, 1, \dots, n.$$

Genereerime taas juhusliku suuruse U , mis allub ühtlasele jaotusele lõigus $[0; 1]$. Kui $u < \alpha(v_i, v_{i+1})$, siis $v_{i+1} = v_i + \Delta v$. Vastasel juhul $v_{i+1} = v_i$.

Raha muutumise üht võimalikku dünaamikat võib näha joonisel 1. Näite 3 puhul on tegemist juhuga, mille puhul rahahulk pigem kasvab, kuid eksisteerib risk oma kasum maha mängida. Mida suurem on rahahulk, seda suuremaks muutub ka raha kulutamine.

Kirjandus

Gurski, J. *Tõenäosusteooria ja matemaatilise statistika elemente*, Tallinn, "Valgus", 1986.

Metropolis, N. et al. Equations of step calculation by fast computing machine. *Journal of Chemical Physics*, **21** (1953), 1087–1091.