

MCMC-menetelmien ongelmakohtia ja ratkaisuja

Alexi Saari, 17.10.2001

Lähteet:

- Mackay: Introduction to Monte Carlo Methods
- Neal: Suppressing Random Walks in Markov Chain Monte Carlo Using Ordered Overrelaxation

Sisältö

- MCMC-menetelmien ongelmakohtia: hitaus, konvergoinnin mittaus ja resurssien jako.
- Tehostettuja MCMC-menetelmiä: hybridi Monte Carlo, simuloitu karkaisu (Simulated annealing), ylirelaksaatio (Overrelaxation) ja järjestetty ylirelaksaatio.

Mistä johtuu MCMC-menetelmien hitaus?

- Usein jakaumassa olevia korrelaatiota ei saada poistettua. Tästä seuraa, että otannassa on käytettävä hyvin pieniä askelia:
$$T \simeq (\sigma^{max} / \sigma^{min})^2.$$
- Menetelmät ovat alttiita satunnaiskululle (random walk).
- Menetelmien konvergoimiseen kuluva aika on vaikea arvioida. Alaraja voidaan onnistua määrittämään, mutta tarkan arvion tekeminen on hankalaa.
- Laskennan mahdollista konvergointia on vaikea mitata. Esim. algoritmi voi juuttua "saarekkeeseen" pitkäksi aikaa, jolloin saadaan väärä vaikutelma konvergoinnista.

Suurten mallien käsittely

- Metropolis -menetelmää voidaan muuttaa siten, että yhden uuden tilan etsintään käytettävän funktion $Q(x'; x)$ sijaan käytetään useita funktiota $Q^{(b)}(x'; x)$.
- Tällöin (ehdollinen) päivitys kohdistetaan yhteen tai useampaan muuttujaan.

- Gibbsin -menetelmän käyttö voi tehostua, jos päivitetään kerralla yhden muuttujan sijasta suurempaa muuttujien osajoukkoa:

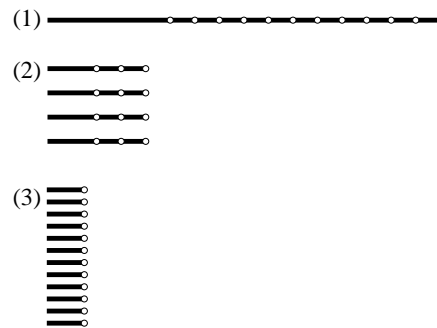
$$x_1^{(t+1)} \dots x_a^{(t+1)} \sim P(x_1 \dots x_a \mid x_{a+1}^t \dots x_K^t)$$

$$x_{a+1}^{(t+1)} \dots x_b^{(t+1)} \sim P(x_{a+1} \dots x_b \mid x_{b+1}^t \dots)$$

Näytteiden määrä ja resurssien jako

- Tarvittavien näytteiden määrä ei ole riippuvainen ongelman dimensiosta, vaan ainoastaan mitattavan suureen varianssista $\sigma^2 = \int P(x)(\phi(x) - \Phi)^2 dx$.
- Tällöin jo pienellä määrällä näytteitä saadaan hyvä arvaus tutkittavalle ongelmalle. Esim. kahdellatoista näytteellä saadaan suureen estimaatille arvo, jonka tarkkuus on $\sigma/\sqrt{12}$.
- Edustavatkaan näytteet eivät tosin paljasta kaikkea jakaumasta ja ongelman luonteesta...

- On olemassa erilaisia strategioita laskea halutut näytteet samalla resurssien määrällä:
 1. Lasketaan yksi pitkä Markovin ketju, josta näytteet otetaan.
 2. Lasketaan monta kohtalaisen pitkää ketjua, joista näytteet otetaan.
 3. Lasketaan jokaista näytettä varten oma Markovin ketju.

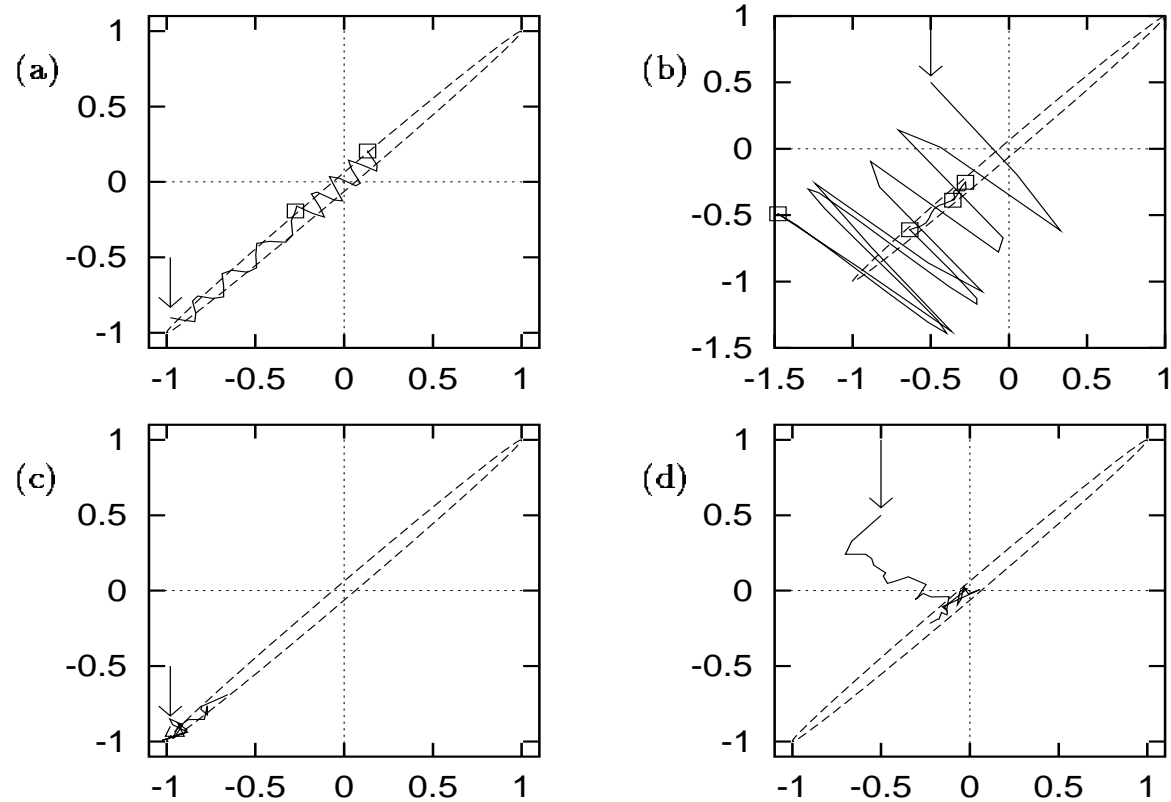


- Pidempi Markovin ketju konvergoi todennäköisemmin oikeaan jakaumaan ja laskennan parametrejä tarvitsee säätää vain kerran, mutta lyhyet Markovin ketjut ovat paremmin rippumattomia toisistaan. Kultaista keskitietä suositaan.

Hybridi Monte Carlo

- Soveltuu jakaumille jotka saadaan kirjoitettua muotoon $P(x) = \frac{e^{-E(x)}}{Z}$. $E(x)$ ja sen gradientti on lisäksi pystyttävä määrittämään.
- Määritellään Hamiltonin funktio $H(x, p) = E(x) + K(p)$, jossa $K(p)$ kuvaa kineettistä energiaa esim. $K(p) = p^T p/2$.
- Saadaan uusi jakauma $P_H(x, p) = \frac{1}{Z_H} \exp[-H(x, p)] = \frac{1}{Z_H} \exp[-E(x)] \exp[-K(p)]$, josta alkuperäinen jakauma saadaan yksinkertaisesti jättämällä potentiaalitermi pois.
- Potentiaalimuuttujan p käytöstä seuraa, että lopputilan etäisyys alkutilasta kasvaa lineaarisesti laskenta-ajan kasvaessa ja laskenta voi nopeutua jopa kertoimella ~ 100 tai ~ 1000 .

- Tilamuuttujaa ja potentiaalimuuttujaa päivitetään kahdessa vaiheessa:
 1. $p \sim \exp[-K(p)]/Z_K$
 2. $\Delta x = p$
 $\Delta p = -\frac{\partial E(x)}{\partial x}$
- 1. vaihe: otetaan uusi potentiaalimuuttuja normaalijakaumasta.
- 2. vaihe: päivitetään tila -ja potentiaalimuuttujaa siten, että Hamiltonin funktio pysyy (likimain) vakiona.
- Tilan x todennäköisyyden ollessa pieni sen muutos on suuri ja toisinpäin. Lisäksi tilaa muutetaan gradientin suuntaan.
- Mackay: 2. vaiheessa tehdään N-kappaletta (hihavakio) "leap frog"-askelia, ts. paloittainen lineaarinen approksimaatio.



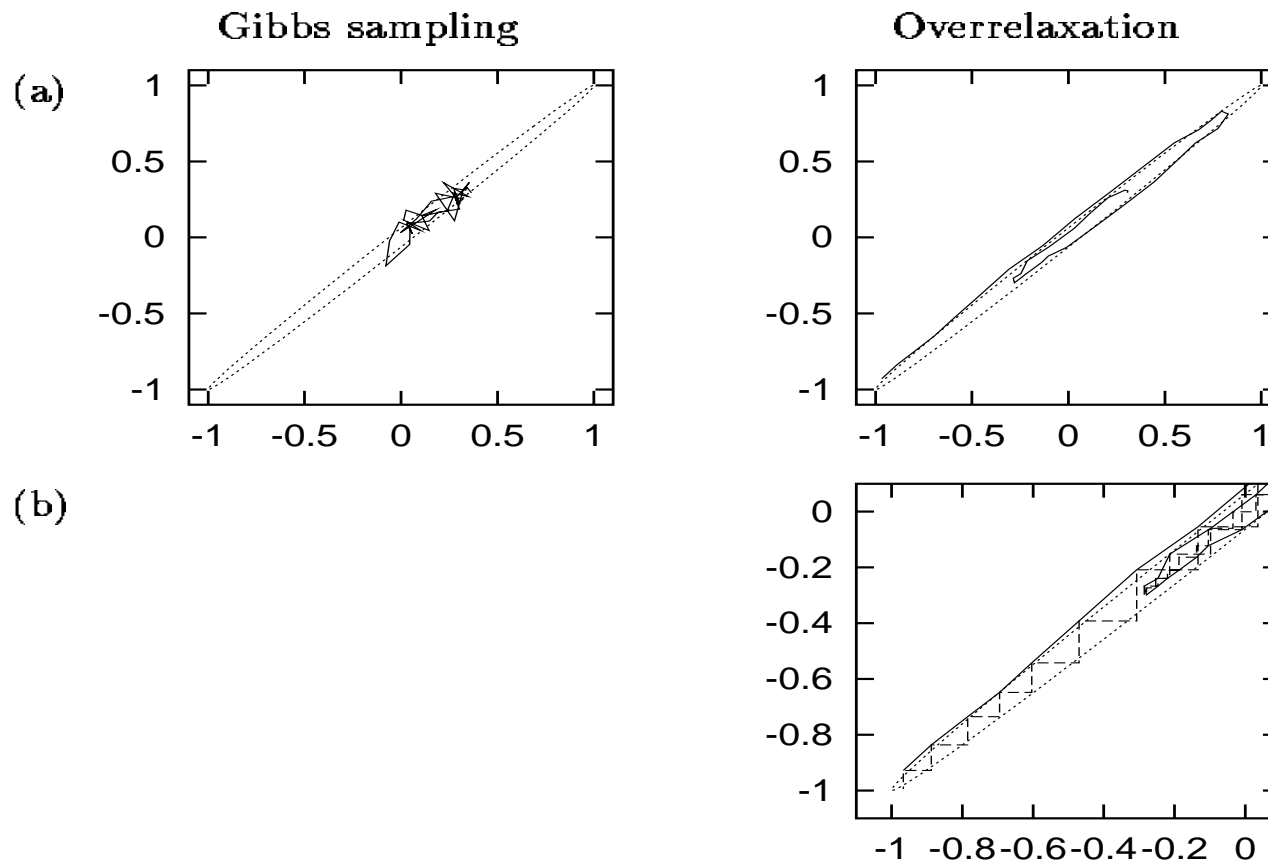
- Kuvat a) ja b) esittävät hybridin Monte Carlo -menetelmän kulkua kun lähtöarvaus on todennäköinen/epätodennäköinen. Kuvat c) ja d) esittävät tavallisen Metropolis -menetelmän kulkua samoilla lähtöarvauksilla.

Simuloitu karkaisu

- Soveltuu jakaumille, jotka saadaan kirjoitettua muotoon $P(x) = \frac{e^{-E(x)}}{Z}$. $E(x)$ on pystyttävä määrittämään.
- Menetelmässä otannan kohteena olevan funktion "lämpötilaa" T lasketaan vähitellen korkeasta lämpötilasta matalampaan, jolloin voidaan välttää juuttuminen epäedustavaan saarekkeeseen.
- Otanta on hieman biasoitunut, mutta sen korjaamiseksi on olemassa menetelmiä.
- Yksinkertaisin muoto: $P_T(x) = \frac{1}{Z(T)} e^{-\frac{E(x)}{T}}$, $t \rightarrow_+ 1$
- Jos $E(x)$ voidaan hajottaa siististi käyttäytyvään ja hankalaan osaan $E(x) = E_0(x) + E_1(x)$, voidaan käyttää muunnosta:
 $P'_T(x) = \frac{1}{Z'(T)} e^{-E_0(x) - E_1(x)/T}$, $t \rightarrow_+ 1$

Ylirelaksaatio

- Ylirelaksaatio on hybridiä Monte Carlo -menetelmää vastaava menetelmä Gibbsin otannan tehostamiseen, jolla pyritään vähentämään satunnaiskulkua.
- Soveltuu menetelmille, joiden ehdolliset todennäköisyysjakaumat ovat gaussisia.
- Adlerin menetelmässä ehdollista gaussista jakaumaa biasoidaan korreloimaan negatiivisesti vanhan arvon kanssa (α on yleensä negatiivinen, jos $\alpha > 0$ puhutaan alirelaksaatiosta).
 - $x_i^{t+1} = \mu + \alpha(x_i^{(t)} - \mu) + (1 - \alpha^2)^{1/2}\sigma\nu$
 - $\nu \sim N(0, 1)$ ja $\alpha \in [-1, 1]$.
- Menetelmä ei ole tarkasti tasapainoinen (detailed balance), mutta se konvergoi $P(x)$:ään joten sitä voidaan käyttää otannassa.

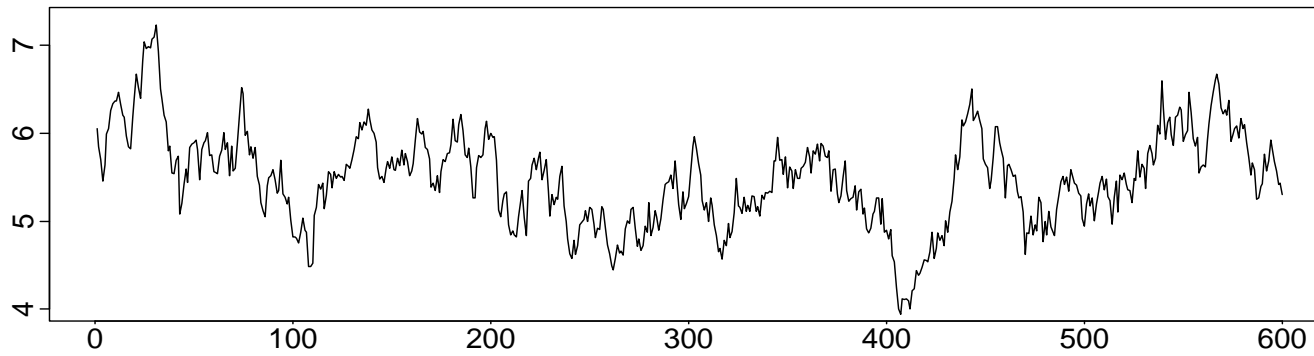


- Kuvissa a) esitetään Gibbsin otannan ja ylirelaksaation toimintaa kun kaksi muuttujaa ovat vahvasti korreloituneita. Kuva b) on suurennus ylirelaksaation kaksiosaisesta toiminnasta.

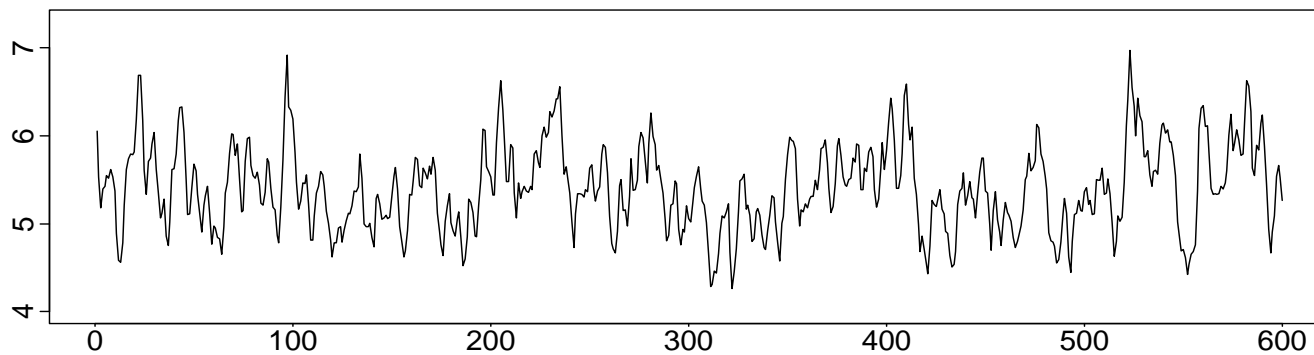
Järjestetty ylirelaksaatio (Neal 1998)

- Ylirelaksaatiomenetelmä joka soveltuu myös tapauksiin, joissa ehdolliset todennäköisyydet eivät ole gaussisia.
- Menetelmän toiminta:
 1. Generoi K riippumatonta arvoa x_i :lle ehdollisesta todennäköisyysjakaumasta $P(x_i | \{x_j\}_{j \neq i})$.
 2. Järjestä lukujoukko vanhan arvon kera järjestykseen:
$$x_i^{(0)} \leq x_i^{(1)} \leq \dots \leq x_i^r = x_i \leq \dots \leq x_i^K,$$
jossa r on vanhan arvon indeksi ko. järjestyksessä.
 3. Aseta uusi arvo: $x'_i = x_i^{(K-r)}$

- Jokaisella kierroksella siis generoidaan K riippumatonta arvoa, mutta useissa tapauksissa menetelmän tuoma nopeutus korvaa laskennan määrän kasvun.
- Jos käytettävissä on kumulatiivinen ehdollinen todennäköisyysfunktio $F(x)$ ja sen käänteisfunktio $F^{-1}(x)$ voidaan uusi arvo x'_i saada laskematta K :ta riippumatonta arvoa.
- Lisäksi joissain tapauksissa K :n riippumattoman arvon generointi ehdollisesta todennäköisyysfunktioista vaatii vähemmän kuin K -kertaisen resurssimäärän.



Plot of τ during Gibbs sampling run



Plot of τ during ordered overrelaxation run with $K = 5$

- Ylemmässä kuvassa Gibbsin menetelmän antama arvo τ :lle ja alemmassa τ :n arvo järjestettyä ylirelaksaatiota käyttäen. Oikea arvo τ :lle on 5.

Yhteenveto

- Perus MCMC -menetelmät ovat monissa tilanteissa liian hitaita, mutta niitä voidaan nopeuttaa.
- Sopivan menetelmän valinta riippuu tutkittavasta jakaumasta.
- Resurssien jako on myös yksi huomioitava seikka MCMC -menetelmiä käytettäessä.