Johdattelua todennäköisyysjakaumien stokastiseen simulointiin

17 syyskuuta armon vuonna 2001
Kari Vasko (vasko@csc.fi)

CSC – Tieteellinen laskenta Oy &
From **D**ata to **K**nowledge (FDK) –yksikkö,
Tietojenkäsittelytieteen laitos, HY

Ongelma

- Olkoon X satunnaismuuttuja, joka noudattaa laskennallisesti hankalaa todennäköisyysjakaumaa $P : \Omega \to \mathbb{R}$ ja olkoon $\phi : \Omega \to \mathbb{R}^q$ funktio, joka kuvaa X:n piirteen $\phi(X)$.
- **Missio:** Evaluoi $\phi(X)$:n odotusarvo

$$E(\phi(X)) = \int_\Omega P(x)\phi(x)dx.$$

- Laskennallisesti hankala tarkoittaa, että ei tunneta
  - analyyttistä menetelmää **tai**
  - muuta laskennallisesti tehokasta menetelmää

  jolla $E(\phi(X))$ saataisiin evaluoitua.
- Kurssin viitekehyksessä P on Bayes-verkon posteriorijakauma.

Ratkaisu 1: Yksinkertaista mallia

- Yksinkertaista mallin oletuksia, jotta tf:n piirteet olisivat helpommin evaluoitavissa. Esimerkiksi
  - Höllennä riippuvuusoletuksia
  - Käytä konjugaattiprioreita (analyyttisen ratkaisun mahdollisuus)
  - Kokeile analyyttisiä approksimatiivisia menetelmiä

Ratkaisu 2: Generoi satunnaisotoksia

- Generoi satunnaisotoksia $X_1, \ldots, X_n$, jotka ovat jakautuneet P:n mukaisesti ja Monte Carlo -integroi. Ts.

$$E(\phi(X)) = \int_\Omega P(x)\phi(x)dx \approx \frac{1}{n}\sum_{i=1}^{n} \phi(X_i) = \hat{\phi}(n),$$

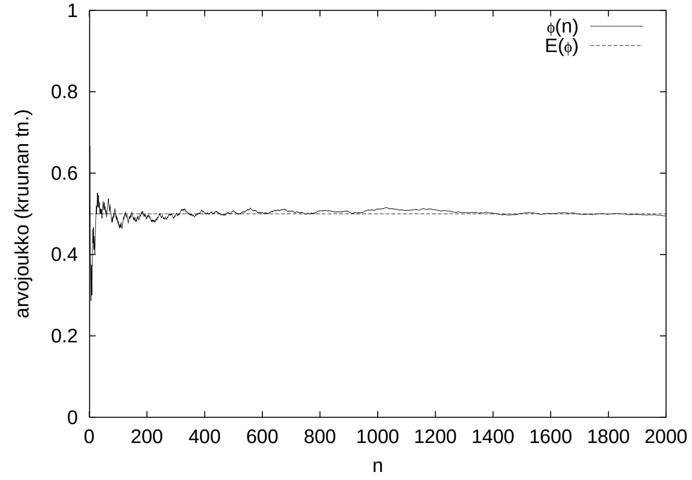
  kun n on suuri (suurten lukujen laki) ja $\phi : \Omega \to \mathbb{R}^q$ on P:n piirre, jonka massakeskipisteestä ollaan kiinnostuneita.
- Monte Carlo –integrointi suppenee stokastisesti (converges in probability) kohti $E(\phi)$:tä ts. kaikilla $\epsilon > 0$ on voimassa

$$\Pr(\|E(\phi) - \hat{\phi}(n)\| > \epsilon) \to 0, \text{ kun } n \to \infty.$$

- X_i:t ovat satunnaismuuttujia $\Rightarrow \hat{\phi}(n)$ on satunnaismuuttuja $(E(\hat{\phi}(n)) = E(\phi))$.

- Esimerkki: $\phi_1(x) = id(x) = x \Rightarrow \hat{\phi_1}(n) \to E(X)$ ja $\phi_2(x) = (x - E(X))^2 \Rightarrow \hat{\phi_2}(n) \to E(V(X)) \equiv V(X)$, jossa $V(X)$ on X:n varianssi, kun $n \to \infty$.
- Huom! Satunnaisotannan sivutuotteena saadaan muitakin data-analyysin kannalta hyödyllisiä P:tä luonnehtivia otuksia kuin piirteen $\phi(X)$ odotusarvo; Esimerkiksi
  - P:n kokonaistodennäköisyysjakauman ja edelleen kaikkien reunajakaumien diskreetit estimaatit
  - Reunajakaumien (tai vaikka kokonaistodennäköisyysjakauman) suurimman tiheyden joukot (high density regions)
  - moodi, mediaani, jne...

Leluesimerkki: kruunan tn:n estimointi



Otanta ei ole helppoa ei...

- Tyypillisiä ongelmia:
  - Tn-massa on keskittynyt pienelle alueelle, kun dimensio on suuri (tyypillinen joukko, dimensiokirous)
  - Konvergenssin arvioiminen
  - Täsmäotanta ei onnistu yleisessä tapauksessa
- Jos tunnettaisiin $P^* \propto P$ diskreetti approksimointi

$$P(x_i) \approx \frac{P^*(x_i)}{\sum_j P^*(x_j)}$$

raa'alla voimalla maksaisi eksponentiaalisen ajankäytön ($\mathcal{O}(2^N)$ jo binäärivektorien tapauksessa) suhteessa aineiston dimensioon, jossa x_i:t ovat arvojoukkoon tasaisesti levitetyn hilan pisteitä

- Vaikka tunnettaisiinkin normalisointitermi $Z_P \in \mathbb{R}$ ts. $P(x) = P^*(x)/Z_P$ kaikilla $x \in \Omega$, olisi otanta edelleen haasteellista suuridimensioisissa tapauksissa (poislukien muutamat erikoistapaukset).
- Ollee selvää, että $\mathcal{O}(c^N)$ $(c > 1)$ on turhan hintavaa, joten jatkossa tarkastelemma satunnaisalgoritmeja, joilla generoidaan satunnaisia pisteitä Ω:sta sopivalla tavalla MC-integroinnin suorittamiseksi (s.e. toivotaan $n \ll c^N$ riittävän).
- Sopiva tapa tulee olemaan käyttää ehdotusjakaumaa Q ja jotain sopivaa muljautusta näille tai stokastista lisäehtoa Q:ksi vinoutuneen satunnaisotannan korjaamiseksi P:n satunnaisotannaksi.

Riippumattoman otannan konvergointi

- Jos otokset X_i ja X_j olisivat riippumattomia kaikille $i \neq j$, $\hat{\phi}(n)$:n varianssi vähenisi verrannollisesti otoksien lukumäärään $V(X)/n$ **dimensiosta riippumatta**.
- Käytännössä osoittautuu, että täsmäotannan (exact sampling) toteuttaminen ei ole helppoa.

Tasajakautunut otanta (Uniform Sampling)

- Ensimmäinen mieleen tuleva satunnaisalgoritmi on tasajakautunut satunnaisotanta. Menetelmässä generoidaan satunnaisesti pisteitä $X_1, \ldots, X_n$ tasaisesta jakaumasta ja evaluoidaan P^* ko. pisteissä ja lasketaan

$$E(\phi(X)) \approx \sum_i \phi(X_i) \frac{P^*(X_i)}{\sum_j P^*(X_j)}$$

- Huom! Tasajakautuneessa satunnaisotannassa joudutaan olettamaan ('pientä' poikkeusjoukkoa lukuunottamatta), että

$$\sup_{x,y\in\Omega} d(x,y) < \infty \text{ (tai } \#\Omega < \infty \text{, kun } \Omega \text{ diskr/nomin).}$$

- Tyypillinen joukko, laihat hännät.

Painotusotanta (Importance Sampling)

- Painotusotanta on tasajakautuneen satunnaisotannan yleistys. Tasajakautuneen ehdotusjakauman sijaan generoidaan otokset $X_1, \ldots, X_n$ jonkin tn-jakauman Q mukaisesti, joka ei välttämättä ole tasainen jakauma.
- Koska näin $X_i \sim Q$ kaikilla $1 \leq i \leq n$, otoksista saadaan esille vain Q:n piirre $E_Q(\phi(Y))$, jossa $Y \sim Q$, josta ei olla kiinnostuneita.
- Suuri idea on käyttää korjauspainoja

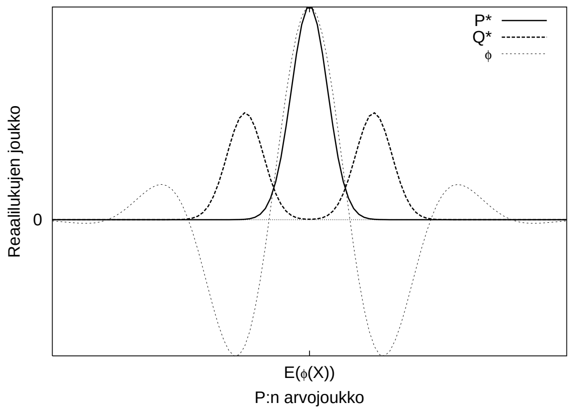
$$w(X_i) := \frac{P^*(X_i)}{Q^*(X_i)},$$

jotka vähentävät (kasvattavat) otoksen X_i merkitystä, jos $P(X_i)$:n tiheys on pienempi (suurempi) kuin $Q(X_i)$:n.

- Painoja $w(X_i)$ $(1 \leq i \leq n)$ käyttäen lasketaan painotettu keskiarvo

$$\hat{\phi}(n) = \frac{\sum_{i=1}^{n} \phi(X_i) w(X_i)}{\sum_{j=1}^{n} w(X_j)}.$$

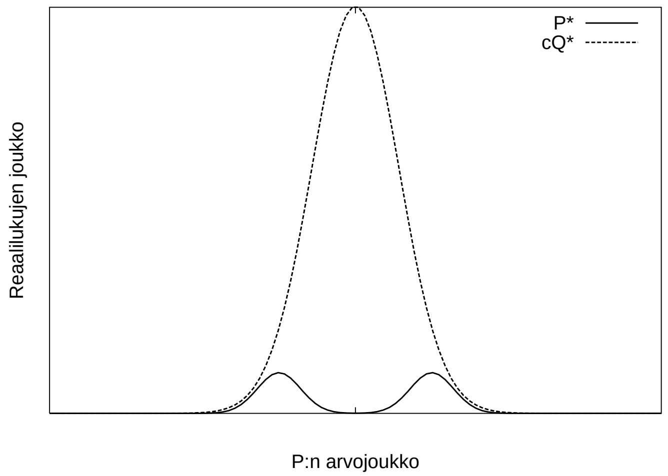
(nähdään, että tasajakautuneessa satunnaisotannassa painot olivat vakiot ja ne supistuivat pois)
- Painotusotanta ei toimi hyvin, jos Q ei noudata karkeasti P:n piirteitä. Esimerkiksi, jos $Q(x)$ on pieni alueessa $A \subset \Omega$, kun $|\phi(x)P(x)|$ on suuri alueessa A (ts. A on 'merkityksellinen' odotusarvo-operaattorin arvon kannalta), on epätodennäköistä, että satunnaisotanta vierailee A:ssa valitulla ehdotusjakaumalla Q.
- Korkeadimensioisissa avaruuksissa lisäksi omat ongelmansa:
  - tyypillinen joukko pienellä alueella
  - tyypillisenkin joukon painoilla suuri varianssi



- Esimerkki, jossa painotusotanta ei toimi järkevässä ajassa hyvin.

Hylkäysotanta (Rejection Sampling)

- Hylkäysotannassa generoidaan satunnaisotoksia ehdotusjakauman Q mukaisesti. Otos X_i hyväksytään osaksi MC-integrointia, jos eräs stokastinen ehto toteutuu. Hylkäysehdolla estetään $\hat{\phi}(n)$:n vinoutuminen Q:n mukaisesti.
- Olkoon c sellainen vakio, että $c \cdot Q^*(x) > P^*(x)$ kaikilla $x \in \Omega$. Huom! Kaikille Q ei ole olemassa vakiota ko. vakiota.
- **Hyväksymisehto:** Otos X_i hyväksytään, jos $P^*(X_i) > U_i$, jossa $U_i \sim \mathrm{Unif}(0, c \cdot Q^*(X_i))$.
- MC-integrointi tehdään hyväksyttyjen otosten joukossa.
- Käytännössä c kasvaa eksponentiaalisesti N:n funktiona $\Rightarrow$ Hyväksymisen tn. on pieni alueilla, joista tulee usein ehdotuksia (eli Q:n ST-alueet).



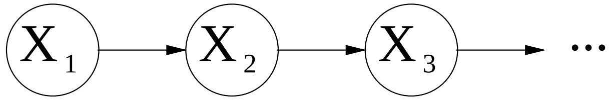
- Esimerkki, jossa hylkäysotanta ei toimi tehokkaasti.

Markovin ketju

- Tähän asti ollaan tarkasteltu täsmäotantaa. Nyt tarkastellaan muutamaa menetelmää autokorreloivien otoksia generoimiseksi.
- Kulminaatiopisteeksi muodostuu autokorrelaation ja Ω:n haravoimisen välissä tasapainottelu (satunnaiskulku).
- Jos ehto

$$P(X_i | X_{i-1}) = P(X_i | X_1, X_2, \ldots, X_{i-1}) \text{ kaikilla } i$$

pätee jonolle $S = X_1, X_2, \ldots$ satunnaismuuttujia, sanotaan, että S on “Markovin ketju”.



Markovin ketju, ehdotus- ja tasapainojakauma

- MCMC-menetelmissä generoidaan Markovin ketju, jolla on *tasapainojakaumana* P.
- Autokorrelaation johdosta konvergenssin arvioiminen matemaattisesti haastavaa. Peukalosääntöjä ja erilaisia mutu-pohjaisia hihafaktoreita löytyy pilvin pimein.
- Tasapainojakauman olemassaolo ja yksikäsitteisyys ei seuraa automaattisesti. On tarkastettava, että Markovin ketjulla on olemassa yksikäsitteinen tasapainojakauma (ja että se on P).

Tasapainojakauman 1-käsitteisyys ja olemassaolo

- Markovin ketjulla on yksikäsitteinen tasapainojakauma, jos se toteuttaa seuraavat ehdot:
  1. Se on ei-väistyvä (irreducible): todennäköisyys saapua tilasta x tilaan y on nollasta poikkeava kaikilla x ja y.
  2. Se on jaksoton (aperiodic): Todennäköisyys liikkua tilasta x itseensä on nollasta poikkeava kaikilla x.
  3. Odotusarvo saapua tilaan x tilasta y on äärellinen kaikilla tiloilla x ja y.

  (Ehdot 1 ja 2 ovat riittävät tasapainojakauman olemassaolon ja yksikäsitteisyyden kannalta, jos $\#\Omega < \infty$. S:n sanotaan olevan ergodinen, jos se toteuttaa ehdot 2 ja 3).

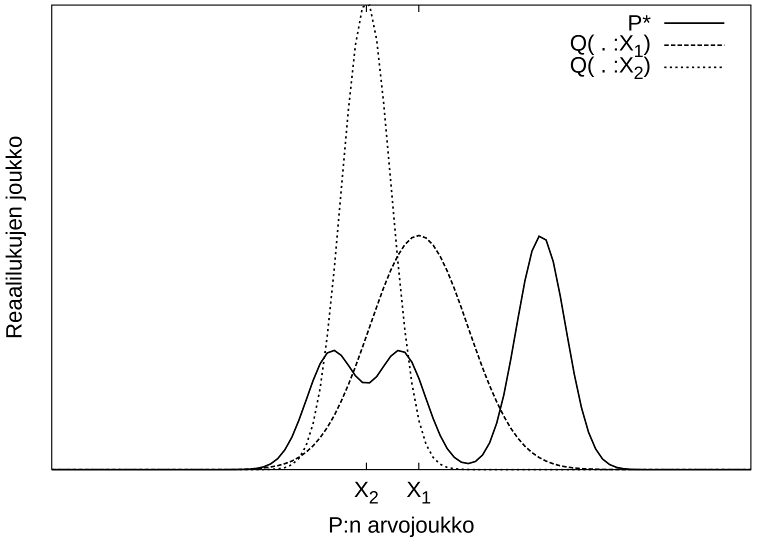
Markov Chain Monte Carlo –otanta

- Generoidaan autokorreloivia satunnaisotoksia käyttäen sopivaa ehdotusjakaumaa $Q(\cdot : \cdot)$ ja sopivaa stokastista lisäehtoa ($Q(y; x)$ kertoo todennäköisyystiheyden ehdottaa tilaa y, kun ollaan tilassa x).
- MCMC-menetelmistä on olemassa suuri repertuaari erilaisia versioita.
  - Perusmenetelmien variaatioita
  - Adaptoituvia menetelmiä
- Tässä keskitytään muutamaan tunnetuimpaan: Metropolis-Hastings –algoritmiin ja sen erääseen erikoistapaukseen Gibbsin otantaan.

Metropolis-Hastingsin otanta

1. Vedä hihasta tai jostain jakaumasta alkuarvo $x_0 \in \Omega$.
2. Juoskoon $i = 1, \ldots, N$
   (a) $x' :=$ ehdotus jakaumasta $Q(\cdot : x_{i-1})$
   (b) osamäärä:$= \frac{P^*(x')Q(x_{i-1}:x')}{P^*(x_{i-1})Q(x':x_{i-1})}$
   (c) $U :=$ satunnaisotos jakaumasta Unif(0,1).
   (d) Jos $U <$min(1,osamäärä)
   - $x_i := x'$ (hyväksy ehdotus)

   muutoin
   - $x_i := x_{i-1}$ (hylkää ehdotus, hyväksy vanha tila uudeksi)
3. Tulosta $x_1, \ldots, x_n$



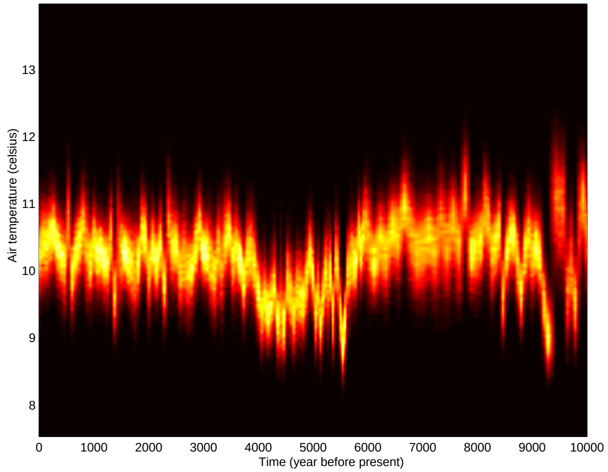
Gibbsin otanta

- Gibbsin otanta on MH-otannan erikoistapaus (kun $N > 1$), jossa dimensiolle k käytetään ehdotusjakaumana mielenkiinnon kohteena olevan tf:n ehdollista jakaumaa $P(x_t(k)|x_{t-1}(1), \dots, x_{t-1}(k-1), x_{t-1}(k+1), \dots, x_{t-1}(N))$, jossa $x_i(k)$ on i:nen otoksen k:s dimensio.
- Dimensioille olemassa luonnollinen järjestys Bayes-verkon tapauksessa. Ts. solmulle i vedetään arvoja jakaumasta $P(\cdot|pa(i))$.
- Fig. 9, pp 191.
- Etu: lokaaleista jakaumista osataan yleensä generoida satunnaisotoksia helpohkosti.

Diskussiota MH:sta & Gibbsistä

- MH kannattaa toteuttaa Bayes-verkoille s.e. evaluoidaan ehdotukset ketjussa juuresta lapsiin.
- Tehokas satunnaiskulku ei ole itsestäänselvää. Kuva.
- MacKay:llä heittoja ja peukaloarvioita efektiivisesti riippumattomien satunnaisotoksien generoimiseksi.

MH –otantaa: Tosielämän esimerkki



Käynnissä olevaa tutkimusta

- Täsmäotanta
- Adaptoituvat menetelmät
- 'Palautuva kulku'-MCMC (reversible jump MCMC)
  - Integroidaan sopivan mallijoukon (kurssin kontekstissa Bayes-verkkojoukon) yli.

Yhteenvetoa

- Satunnaismenetelmien yhteinen nimittäjä on, että generoidaan otoksia $X_1, \ldots, X_n$ jostain ehdotusjakaumasta Q sopivan stokastisen lisäehdon höystämänä (pl. painotusotanta). Yleisluonteinen toive on, että Q arvoja tulisi noukittua tyypillisen joukosta nopeasti ja efektiivisesti.
- Riittävän yleistä täsmäotantamenetelmää ei ole, joten maailma kärsii autokorrelaatiosta.
- Muidenkin tunnuslukujen (tai välien) kuin $E(\phi(X))$:n estimointi satunnaisotannalla kyytipoikana (pl. PO).
- MCMC-veivaukset (ja satunnaisotantamenetelmät yleensä) ovat raskasta taustaoletuksista tinkimätöntä data-analyysiä, joka ei sovellu aikakriittiseen analyysiin.

Lisää tietoa

- MCMC:stä

  `http://www.cs.toronto.edu/~radford/review.abstract.html`
- Täsmäotannasta

  `http://dimacs.rutgers.edu/~dbwilson/exact/`