Batch means -menetelmä

- Batch means -menetelmää käytetään hyvin yleisesti
- Simulointi suoritetaan tässä yhtenä pitkänä ajona
  - olkoon simuloinnin pituus M
    - tässä ajatellaan, että järjestelmää tarkastellaan asiakaspohjaisesti, jolloin M voi tarkoittaa kiinnostavien havaintojen lukumäärää (yhtä hyvin M voisi tarkoittaa aikaa)
  - olkoon kiinnostuksen kohteena oleva suure X (esim. odotusaika jonossa) ja tehtävänä on arvioida sen odotusarvo $\mu = \mathrm{E}[X]$
- Simuloinnin alusta karsitaan pois alkulämmittelyosuutena K havaintoa
- Hyötyajo (pituus $M - K$) jaetaan N:ään erään (batches) eli jokaisessa erässä havaintoja on

$$n = \frac{M - K}{N}$$

Batch means -menetelmä (jatkoa)

- Erässä i saadaan X:lle otoskeskiarvo (X_{ij} on j:s havainto kyseisessä erässä)

$$\bar{X}_i = \frac{1}{n} \sum_{j=1}^{n} X_{ij}$$

- Lopullinen estimaattori odotusarvolle μ on

$$\hat{\mu}_N = \frac{1}{N} \sum_{i=1}^{N} \bar{X}_i = \frac{1}{nN} \sum_{i=1}^{N} \sum_{j=1}^{n} X_{ij}$$

- Tämä on yksinkertaisesti otoskeskiarvo koko ajosta (alkutransientin karsinnan jälkeen)
  - eriin jakamisella ei ole mitään merkitystä itse estimaattorin arvon kannalta
  - sen ainoana tarkoituksena on saada käsitys estimaattorin luottamusvälistä
- Olettaen, että erät ovat kyllin pitkiä, otoskeskiarvot $\bar{X}_i$ ovat likimain riippumattomia
- Niiden otosvarianssi tarjoaa silloin estimaatin yksittäisen $\bar{X}_i$:n varianssille

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (\bar{X}_i - \hat{\mu}_N)^2$$

Batch means -menetelmä (jatkoa)

- Estimaattorin luottamusväli (luottamustasolla $1 - \beta$) on

$$\hat{\mu}_N \pm z_{1-\beta/2} \frac{S}{\sqrt{N}}$$

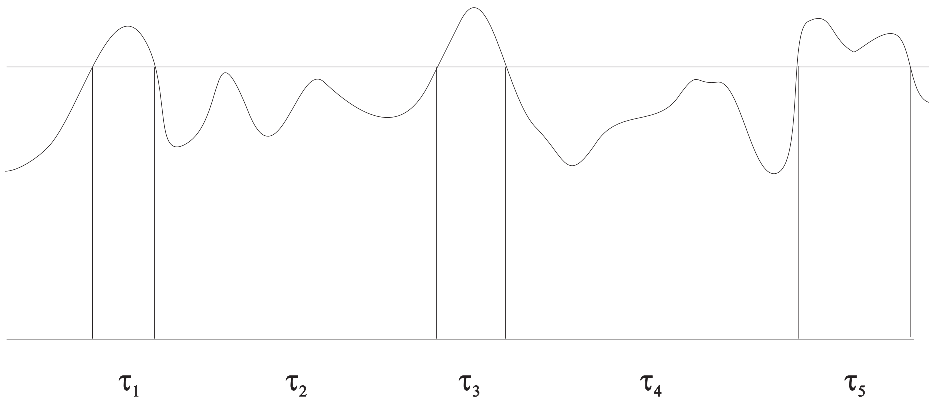
- Menetelmän etuna on, että lämmittelyjaksoja on vain yksi
- Eriä pitäisi olla vähintään luokkaa 20-30 kappaletta, jotta otosvarianssi voidaan järkevästi arvioida
- Erien pitää olla kestoltaan riittävän suuria (paljon pitempiä kuin alkutransientin kesto), jotta $\bar{X}_i$:t olisivat likimain riippumattomia
- Jos riippuvuutta on, korrelaatio on yleensä positiivinen
- Tällöin $\hat{\mu}$:n todellinen luottamusväli on suurempi kuin edellä annettu riippumattomien erien oletukseen perustuva arvio
  - riippuvuus ei mitenkään heikennä itse estimaattorin arvoa
  - se voi ainoastaan johtaa siihen, että arvio estimaattorin tarkkuudesta on liian optimistinen

Regeneratiivinen menetelmä (uusiutumismenetelmä)

- Soveltuu ns. regeneratiivisiin järjestelmiin
- Tällaisella järjestelmällä on olemassa ainakin yksi regeneraatiotila
  - järjestelmän kehitys tästä eteenpäin ei riipu lainkaan siitä, miten tilaan on tultu
  - Markovisten järjestelmien jokainen tila on regeneratiivinen
  - G/G/1-jonossa tila, jossa systeemi on tyhjä, on regeneratiivinen
- Jos regeneratiivisia tiloja on useita, valitaan niistä yksi menetelmän pohjaksi
  - jatkossa regeneraatiotilalla tarkoitetaan kyseistä valittua tilaa
- Järjestelmä aina silloin tällöin palaa regeneraatiotilaan eli “regeneroi itsensä”
  - tästä alkaa “uusi elämä” joka ei riipu entisestä

Regeneratiivinen menetelmä (jatkoa)

- Hetkeä, jolloin järjestelmä palaa regeneraatiotilaan, kutsutaan regeneraatiopisteeksi
- Periodia kahden regeneraatiopisteen välillä kutsutaan regeneraatiojaksoksi
- Regeneraatiojaksojen kulut ovat täysin toisistaan riippumattomia
  - tämä on koko menetelmän varsinainen "idea"



Regeneratiivinen menetelmä: piste-estimaatori

- Olkoon tutkittavan suureen kumulatiivinen arvo regeneraatiojakson aikana X, esim.
  - kokonaisaika, jonka systeemi on ollut estotilassa jakson aikana
  - puskurista ylivuotaneiden solujen lukumäärä jakson aikana
- Olkoon τ regeneraatiojakson “kesto”
  - tämä voi tarkoittaa jakson todellista kestoa ajassa
  - voi olla myös esim. saapumisten lukumäärä regeneraatiojakson aikana
- Tutkittavan suureen odotusarvolle ℓ (esim. aikaeston odotusarvolle) pätee

$$\ell = \frac{\mathrm{E}[X]}{\mathrm{E}[\tau]}$$

- n:n regeneraatiojakson jälkeen saadaan (vahvasti konsistentti) estimaattori

$$\bar{\ell}_n = \frac{\bar{X}}{\bar{\tau}}$$

missä $\bar{X}$ ja $\bar{\tau}$ ovat otoskeskiarvot $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ ja $\bar{\tau} = \frac{1}{n}\sum_{i=1}^{n} \tau_i$

Regeneraatiomenetelmän estimaattorin luottamusväli

- Tarkastellaan muuttujaa $Z_i = X_i - \ell\tau_i$
  - Z_i:t ovat riippumattomia identtisesti jakautuneita sm:ia (odotusarvo 0)
  - samoin X_i:t ja τ_i:t
- Merkitään

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \bar{\tau} = \frac{1}{n}\sum_{i=1}^{n} \tau_i, \qquad \bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i = \bar{X} - \ell\bar{\tau}$$

- Keskeisen raja-arvolauseen mukaan

$$\frac{n^{1/2}\bar{Z}}{\sigma} = \frac{n^{1/2}(\bar{X} - \ell\bar{\tau})}{\sigma} \to \mathrm{N}(0,1), \quad \text{kun } n \to \infty$$

missä σ^2 on Z:n varianssi

$$\sigma^2 = \mathrm{V}[Z] = \mathrm{V}[X] - 2\ell\mathrm{Cov}[X,\tau] + \ell^2\mathrm{V}[\tau]$$

Luottamusväli regeneraatiomentelmässä (jatkoa)

- Jakamalla $\bar{\tau}$:lla saadaan

$$\frac{n^{1/2}(\bar{\ell}_n - \ell)}{\sigma/\bar{\tau}} \to \mathrm{N}(0,1), \quad \text{kun } n \to \infty$$

- n:n regeneraatiojakson mittaukseen perustuvan piste-estimaattorin $\bar{\ell}_n$ luottamusväliksi saadaan (luottamustasolla $1-\beta$)

$$\bar{\ell}_n \pm \frac{z_{1-\beta/2} S}{\sqrt{n}\bar{\tau}}$$

missä S^2 on otoksesta laskettu harhaton σ^2:n estimaattori

$$S^2 = S_{11} - 2\,\bar{\ell}_n\, S_{12} + \bar{\ell}_n^2\, S_{22}$$

ja S_{11}, S_{22} ja S_{12} ovat X:n ja τ:n otosvarianssit ja kovarianssi

$$S_{11} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2, \quad S_{22} = \frac{1}{n-1}\sum_{i=1}^{n}(\tau_i - \bar{\tau})^2, \quad S_{12} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(\tau_i - \bar{\tau})$$

Regeneratiivinen menetelmä: tarkastelua

- Menetelmän etuja
  - erillistä transientin poistoa ei tarvita
  - etukäteen ei tarvitse kiinnittää parametreja kuten batchien lukumäärä
  - asymptoottisesti tarkka
  - helppo ymmärtää ja toteuttaa
- Menetelmällä on kuitenkin seuraavia haittoja
  - käytännössä voi olla vaikea identifioida regeneraatiopisteitä
  - jos sellainen löytyykin
    - regeneraatiojakso voi olla hyvin pitkä (siihen ei voi vaikuttaa)
    - mutkikkaassa järjestelmässä tilan tunnistaminen voi olla laskennallisesti kallista
  - äärellisellä arvolla n estimaattori $\bar{\ell}_n$ on harhainen
    - itse asiassa alkutransienttiongelma on sittenkin olemassa vaikkakin piilossa