Teoria

- Johdanto simulointiin
- Simuloinnin kulku – prosessin realisaatioiden tuottaminen
- Satunnaismuuttujan arvonta annetusta jakaumasta
- Tulosten keruu ja analyysi
- Varianssinreduktiotekniikoista

Varianssinreduktiotekniikat

- Simulointi on yleensä paljon aikaa / laskentatyötä vaativa tekniikka
- Jos näytteenottoa voidaan jollakin tavoin muuttaa siten, että näytteen varianssi pienenee,
  - saadaan tarkempi estimaatti tietyllä toistojen määrällä
  - tietyn tarkkuuden saavuttamiseen tarvittava toistojen määrä pienenee (kääntäen verrannollisesti varianssin reduktioon eli hajonnan reduktion neliöön)
- Tällaisia menetelmiä kutsutaan varianssinreduktiotekniikoiksi
  - ovat arvokkaita edellyttäen, ettei tekniikan soveltaminen itsessään tuo merkittävää lisää laskentatyöhön
- Varianssinreduktiotekniikoissa
  - käytetään älykkyyttä
  - hyödynnetään tunnettuja tuloksia / tietoa systeemistä

sellaisten havaintojen tuottamiseksi, joissa vaihtelua on vähemmän

Varianssinreduktiotekniikoita

- Yhteisten satunnaislukusekvenssien käyttö
- Antiteettiset muuttujat
- Kontrollimuuttujat
- Ehdollistaminen – tunnettujen keskiarvojen käyttö
- Importance sampling (liittyy lähinnä staattiseen simulointiin; ei käsitellä tässä)
- RESTART

VAROITUS

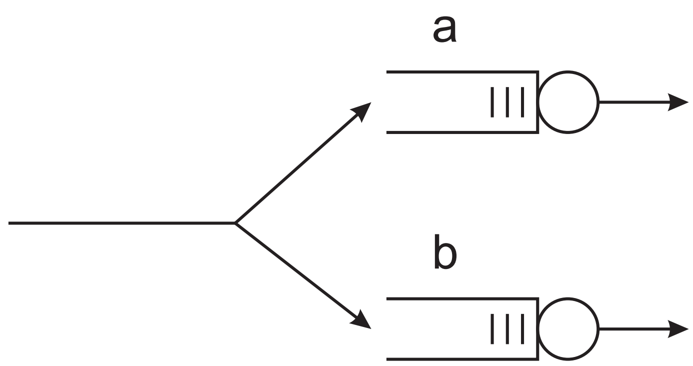
- Varianssinreduktiotekniikoiden virheellinen käyttö saattaa johtaa toivottuun nähden vastakkaiseen tulokseen eli varianssin kasvuun

Yhteisten satunnaislukusekvenssien käyttö

- Tulee kysymykseen vertailtaessa kahden samankaltaisen systeemin suorituskykyä
- Tyypillisessä tapauksessa tutkitaan systeemin käyttäytymistä kahden eri ohjaus- tai toimintapolitiikan A ja B välillä, esim.
  - jonojärjestelmän käyttäytyminen kahdella eri skedulointipolitiikalla:
  - esim. mikä on ero keskimääräisessä odotusajassa kuvan esittämässä kahden jonon järjestelmässä, kun saapuvat asiakkaat ohjataan jonoihin a ja b

    A) vuorotellen

    B) satunnaisesti, 50% todennäköisyydellä joko jonoon a tai b



Yhteisten satunnaislukusekvenssien käyttö (jatkoa)

- Kiinnostavaa on jonkin suorituskykysuureen X odotusarvojen ero näiden kahden politiikan välillä
$$\Delta\mu = \mu_A - \mu_B = \mathrm{E}[X^A] - \mathrm{E}[X^B]$$
missä X^A ja X^B tarkoittavat satunnaismuuttujaa X kyseisten politiikkojen vallitessa ja
$$\mu_A = \mathrm{E}[X^A] \qquad \mu_B = \mathrm{E}[X^B]$$

- Erotuksen $\Delta\mu$ estimoimiseksi on edullista tutkia järjestelmän käyttäytymistä politiikoilla A ja B käyttäen täsmälleen samoja saapumisprosessien realisaatioita
  - jos politiikkaa A simuloitaessa esiintyy saapumisryöppy, niin sama ryöppy esiintyy myös politiikkaa B simuloitaessa
  - eroa ei näinollen pääse muodostumaan simuloinneissa satunnaisesti poikkeavista saapumisprosesseista, vaan ainoastaan politiikkojen eroista

Yhteisten satunnaislukusekvenssien käyttö (jatkoa)

- Simuloinneissa, jotka toistetaan n kertaa saadaan otokset

$$\begin{cases} X_1^A, X_2^A, X_3^A, \ldots, X_n^A \\ X_1^B, X_2^B, X_3^B, \ldots, X_n^B \end{cases}$$

- Politiikalla A otoskeskiarvo ja otosvarianssi ovat

$$\overline{X}_A = \frac{1}{n}\sum_{i=1}^{n} X_i^A \qquad S_A^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i^A - \overline{X}_A)^2$$

ja vastaavat kaavat pätevät politiikalle B

- Erotukselle $\Delta\mu$ saadaan estimaattori

$$\Delta\hat{\mu} = \overline{X}_A - \overline{X}_B$$

jonka varianssi on

$$\mathrm{V}[\Delta\hat{\mu}] = \mathrm{V}[\overline{X}_A] + \mathrm{V}[\overline{X}_B] - 2\mathrm{Cov}[\overline{X}_A, \overline{X}_B]$$

Yhteisten satunnaislukusekvenssien käyttö (jatkoa)

- Otoksen $X_1^A, X_2^A, X_3^A, \ldots, X_n^A$ näytteet ovat toisistaan riippumattomia, samoin otoksen $X_1^B, X_2^B, X_3^B, \ldots, X_n^B$ näytteet
- Samoja realisaatioita käytettäessä kuitenkin näytteet X_i^A ja X_i^B, $i = 1 \ldots n$, ovat korreloituneita
  - korrelaatio on positiivinen, koska samat saapumisryöpyt esiintyvät kummassakin simuloinnissa
  - vastaavasti otoskeskiarvot $\overline{X}_A$ ja $\overline{X}_B$ ovat positiivisesti korreloituneita
  - estimaattorin varianssi on pienempi kuin käytettäessä riippumatomia saapumisprosesseja politiikoilla A ja B
- Estimaattorin varianssin estimaattori on

$$\frac{1}{n}(S_A^2 + S_B^2 - 2S_{AB}^2)$$

missä S_{AB} on otoskovarianssi

$$S_{AB}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i^A - \overline{X}_A)(X_i^B - \overline{X}_B)$$

Yhteisten satunnaislukusekvenssien käyttö (jatkoa)

- Triviaalissa erikoistapauksessa, jossa politiikat A ja B ovat samoja, yhteisten satunnaislukujen käyttö johtaa erotuksen tarkkaan estimaattoriin $\Delta\hat{\mu} = 0$
  - eri satunnaisluvuilla estimaattori on yleensä $\neq 0$

Vertailtaessa kahden järjestelmän suorituskykyä simuloinnit kannattaa tehdä niin identtisissä olosuhteissa kuin mahdollista

- Yhteisten satunnaislukujen käytöllä on pieni vaikutus simulointiohjelman toteutukseen
  - esim. vertailtaessa FIFO- ja LIFO-politiikkoja, palveluaikojen arvonta on suoritettava saapumishetkellä, kun riippumattomien simulointien tapauksessa on vapaus arpoa ajat vasta palvelun alkamishetkellä

Antiteettiset muuttujat

- Tehtävänä on estimoida suureen X odotusarvo $\mu = \mathrm{E}[X]$
- Perusidea: tehdään kaksi eri ajoa siten, että toisen ajon "pienet" havaittavat arvot muuttuvat "suuriksi" arvoiksi toisessa ajossa ja päinvastoin; käytetään näiden keskiarvoa estimaattorina
- Olkoon X eräässä kokeessa havaittu tutkitun satunnaismuuttujan arvo
- Olkoon Y erään toisen kokeen tulos, joka noudattaa samaa jakaumaa
- Tällöin $Z = \frac{1}{2}(X+Y)$ on satunnaismuuttuja, jolla on sama odotusarvo kuin X:llä

$$\mu = \mathrm{E}[Z]$$

- Z:n varianssi on

$$\begin{aligned} \mathrm{V}[Z] &= \tfrac{1}{4}(\mathrm{V}[X] + \mathrm{V}[Y] + 2\mathrm{Cov}[X,Y]) \\ &= \tfrac{1}{2}(\mathrm{V}[X] + \mathrm{Cov}[X,Y]) \end{aligned}$$

- Jos X ja Y ovat negatiivisesti korreloituneita, $\mathrm{Cov}[X,Y] < 0$, niin Z:n varianssi on pienempi kuin X:n ja Y:n ollessa riippumattomia ($\frac{1}{2}\mathrm{V}[X]$)

Antiteettiset muuttujat (jatkoa)

- Olkoon X:n jakaumafunktio $F(x)$ ja X:n arvot generoitu käänteisfunktiomentelmällä

$$X = F^{-1}(U)$$

missä $U \sim \mathrm{U}(0,1)$ (tasainen jakauma)

- Tällöin Y, joka generoidaan kaavalla

$$Y = F^{-1}(1-U)$$

noudattaa samaa jakaumaa kuin X

  - kun $U \sim \mathrm{U}(0,1)$ niin myös $1-U \sim \mathrm{U}(0,1)$

- On selvää, että X ja Y ovat negatiivisesti korreloituneita
  - jos U saa pienen (≈ 0) arvon, niin $1-U$ saa suuren arvon (≈ 1)
  - X ja Y saavat arvot jakauman vastakkaisista reunoista

- X ja Y muodostavat *antiteettisen satunnaismuutujaparin*

Antiteettiset muuttujat (jatkoa)

Esimerkki

- Olkoon $X \sim \mathrm{Exp}(1)$ eli

$$X = -\log U \qquad Y = -\log(1-U)$$

- Voidaan osoittaa, että tässä tapauksessa $\mathrm{Cov}[X,Y] = 1 - \pi^2/6 = -0.645$
- Varianssi pienenee 64 %:lla riippumattomien muuttujien tapaukseen nähden (eikä maksa mitään!)
- Jos X:n odotusarvo halutaan estimoida simuloimalla
  - missä ei tietenkään tämän esimerkin tapauksessa ole järkeä
  - kannattaa jokaisen arvotun satunnaisluvun U_i jälkeen käyttää toisena satunnaislukuna arvoa $1 - U_i$

Antiteettiset muuttujat (jatkoa)

Esimerkki (jatkoa)

i	X_i	Y_i	Z_i
1	2.3340	0.1019	1.2180
2	0.0569	2.8945	1.4757
3	1.1121	0.3988	0.7554
4	0.1235	2.1527	1.1381
5	2.2724	0.1088	1.1906
6	2.0813	0.1333	1.1073
7	0.4510	1.0132	0.7321
8	2.2548	0.1108	1.1828
9	0.1761	1.8231	0.9996
10	0.8221	0.5789	0.7005

Otoskeskiarvot

$\bar{X}$	$\bar{Y}$	$\bar{Z}$
1.1684	0.9316	1.0500

Otosvarianssit

S_X^2	S_Y^2	S_Z^2
0.9489	1.0256	0.0634

Antiteettiset muuttujat (jatkoa)

- Toistokokeessa voidaan generoida arvoja $X_1, X_2, \ldots$ riippumattomilla tasajakautuneilla satunnaisluvuilla $U_1, U_2, \ldots$
  - $X_1, X_2, \ldots$ ovat riippumattomia
- Vastaavasti generoidaan sekvenssi arvoja $Y_1, Y_2, \ldots$ riippumattomilla tasajakautuneilla satunnaisluvuilla $1 - U_1, 1 - U_2, \ldots$
  - $Y_1, Y_2, \ldots$ ovat riippumattomia
- Tällöin myös arvot $Z_i = (X_i + Y_i)/2, \; i = 1, 2, \ldots$ ovat riippumattomia
- Kussakin parissa X_i ja Y_i ovat negatiivisesti korreloituneita
- Z_i:n varianssi pienempi kuin riippumattomilla muuttujilla X_i ja Y_i
- Vastaavasti estimaattorin

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} Z_i$$

varianssi on pienempi kuin riippumattomien muuttujien tapauksessa.

Antiteettiset muuttujat (jatkoa)

- Systeemin simuloinnissa sekvenssejä $U_1, U_2, \ldots$ ja $1-U_1, 1-U_2, \ldots$ on käytettävä eri ajoissa "samaan tarkoitukseen"
  - esim. palveluaikojen generointiin
  - synkronointi vaatii huolellisuutta ohjelman toteutuksessa
- Varianssin reduktion suuruutta on edeltäkäsin vaikea arvioida
  - käytännössä asiaa täytyy tutkia testiajojen avulla
- Antiteettisten muuttujien tekniikan käytöstä aiheutuva laskennallinen kustannus on kuitenkin lähes olematon, joten menetelmää kannattaa kokeilla

Kontrollimuuttujat

- Kontrollimuutujien tekniikassa idea on tavallaan vastakkainen antiteettisiin muuttujiin nähden: menetelmässä pyritään hyödyntämään (vahvaa) positiivista korrelaatiota tutkittavan suureen X ja jonkin apusuureen Y, ns. kontrollimuuttujan välillä
- Olennaista on, että *kontrollimuutujan* Y *tarkka odotusarvo* $\nu = \mathrm{E}[Y]$ *oletetaan tunnetuksi*
- Tehtävänä on jälleen arvioida odotusarvo $\mu = \mathrm{E}[X]$
- Simuloinnista mitataan sekä X:n että Y:n arvot
- Näiden avulla muodostetaan varsinainen havaittava suure

$$Z = X - (Y - \nu)$$

jonka odotusarvo on

$$\mathrm{E}[Z] = \mathrm{E}[X] - (\mathrm{E}[Y] - \nu) = \mu$$

- Havaitsemalla toistetuissa simuloinneissa Z:n arvoja saadaan näiden keskiarvona harhaton estimaatti μ:lle

Kontrollimuuttujat (jatkoa)

- Z:n varianssi on

$$\mathrm{V}[Z] = \mathrm{V}[X - Y] = \mathrm{V}[X] + \mathrm{V}[Y] - 2\mathrm{Cov}[X, Y]$$

- Varianssi on pienempi kuin X:llä edellyttäen, että korrelaatio on kyllin vahva

$$\mathrm{Cov}[X, Y] > \mathrm{V}[Y]/2$$

Kontrollimuuttujat: esimerkki

- Tarkastellaan M/M/1-jonoa, joka hetkellä $t = 0$ on tyhjä
- Olkoon X sadan ensimmäisen saapuvan asiakkaan odotusaikojen keskiarvo
- Otetaan kontrollimuutujaksi Y 99 ensimmäisen asiakkaan keskimääräinen palveluaika
  - 100:nnen asiakkaan palveluajalla ei ole vaikutusta X:ään, joten se kannattaa jättää pois
  - Y:n odotusarvo $\nu = \mathrm{E}[Y]$ eli keskimääräinen palveluaika on tunnettu, kun palveluaikajakauma on annettu
- Jos palveluajat ovat (sattuvat olemaan) pitkiä, niin myös odotusaikojen voi otaksua muodostuvan pitkiksi
  - voidaan siis olettaa, että X ja Y ovat positiivisesti korreloituneita
- Y on otaksuttavasti hyvä kontrollimuuttuja
  - toimivuutta pitää testata käytännön kokein
  - voi huonossa tapauksessa johtaa jopa varianssin kasvuun

Kontrollimuuttujat: parannettu versio

- Parempaan tulokseen päästään sisällyttämällä Z:n määritelmään vapaa kerroin a

$$Z = X - a(Y - \nu)$$

- Jälleen pätee $\mathrm{E}[Z] = \mu$
- Parametri a määrätään siten, että Z:n varianssi minimoituu
- Optimaalinen a:n arvo a^* on

$$a^* = \frac{\mathrm{Cov}[X, Y]}{\mathrm{V}[Y]}$$

- Minimoitu varianssi on

$$\mathrm{V}[Z] = \mathrm{V}[X] - \frac{\mathrm{Cov}[X, Y]^2}{\mathrm{V}[Y]} = (1 - \rho_{X,Y}^2)\mathrm{V}[X]$$

missä

$$\rho_{X,Y} = \frac{\mathrm{Cov}[X, Y]}{(\mathrm{V}[X]\mathrm{V}[Y])^{1/2}}$$

on X:n ja Y:n korrelaatiokerroin

Kontrollimuuttujat: parannettu versio

- Parannetussa versiossa havaittavan muuttujan Z varianssi on aina pienempi (tai korkeintaan yhtäsuuri) kuin X:n varianssi
- Korrelaation ei myöskään tarvitse olla positiivinen
  - negatiivisen korrelaation tapauksessa korjaustermin merkki muuttuu, ts. $a < 0$
- Optimaalisen a:n laskemista varten tarvitaan sekä $\mathrm{V}[Y]$ että $\mathrm{Cov}[X, Y]$
  - näiden arvoja ei yleensä tunneta edeltäkäsin
  - arvot on estimoitava valmistelevissa simulointikokeissa mitatun otosvarianssin ja otoskovarianssin avulla

$$S_Y^2 = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{n-1} \qquad S_{X,Y}^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$$

$$a^* \approx \frac{S_{X,Y}^2}{S_Y^2}$$

Kontrollimuuttujat (esimerkin jatkoa)

i	X_i	Y_i
1	3.84	0.92
2	3.18	0.95
3	2.26	0.88
4	2.76	0.89
5	4.33	0.93
6	1.35	0.81
7	1.82	0.84
8	3.01	0.92
9	1.68	0.85
10	3.60	0.88

- X = 100 ensimmäisen asiakkaan keskim. odotusaika
- Y = 99 ensimmäisen asiakkaan keskim. palveluaika

$$\nu = \mathrm{E}[Y] = 0.9$$

- Keskimääräinen saapumisten väliaika 1 (kuorma 0.9)
- Kymmenen toistoa: $\bar{X} = 3.78, \quad \bar{Y} = 0.89$
- Näistä lasketut otosvarianssit ja kovarianssit:

$$S_X^2 = 13.33$$
$$S_Y^2 = 0.002$$
$$S_{X,Y}^2 = 0.07$$

$\Rightarrow \rho_{X,Y} = 0.43$

$\Rightarrow a^* = 30$

$\Rightarrow \bar{Z} = \bar{X} - a^*(\bar{Y} - \nu) = 4.13$

- Tämä on täsmälleen oikea arvo (kun taas $\bar{X} = 3.78$)

Kontrollimuuttujat (jatkoa)

- Kontrollimuuttujien käyttö saattaa vaatia enemmän työtä (etenkin parannetun version käyttö) kuin edellä käsitellyt reduktiotekniikat
  - menetelmä voi kuitenkin olla varsin tehokas
- Kontrollimuuttuja voidaan valita myös jostakin rinnakkaisesta systeemistä
- Tehtävänä voi esim. olla tutkia keskimääräsitä viipymäaikaa M/M/1-jonossa jonkin kehittyneen skedulointimenetelmän alaisuudessa
  - rinnakkaiseksi järjestelmäksi voidaan ottaa tavallinen M/M/1-FIFO -jono
  - kontrollimuuttujaksi otetaan viipymäaika tämän rinnakkaisen järjestelmän jonossa
  - sen odotusarvo tunnetaan, $1/(\mu - \lambda)$
  - tutkittavaa järjestelmää ja rinnakkaisjärjestelmää ajetaan samoilla syötteillä (samat saapumisvirrat: saapumisväliajat ja palveluajat)

Ehdollistaminen

- Idea: korvataan satunnaismuutuja ehdollisella odotusarvollaan
- Tuloksen kerääminen simulaatioajosta merkitsee yleensä tietyn satunnaismuuttujan X keskiarvojen laskemista
  - viipymäaika jonossa; aika, jonka jono on tyhjä; aika, jona kaikki palvelijat ovat käytössä . . .
- Joissakin tapauksissa kyseisen satunnaismuuttujan ehdollinen odotusarvo (ehdollistettuna jonkin toisen, systeemistä mitattavissa olevan muuttujan Y vallitsevaan arvoon) tunnetaan eksaktisti
- Tällöin on edullista kerätä statistiikkaa ei muuttujan X arvosta itsestään vaan kyseisestä ehdollisesta odotusarvosta $\mathrm{E}[X \mid Y]$
  - havaittava muuttuja on nyt $Z = \mathrm{E}[X \mid Y]$
  - tämä on satunnaismuuttuja, jonka arvot vaihtelevat, kun Y vaihtelee
  - kiinteällä Y:n arvolla se on tunnettu luku (oletetaan tunnetuksi)

Ehdollistaminen (jatkoa)

- Pätee
$$\mathrm{E}[Z] = \mathrm{E}[\mathrm{E}\,[X \mid Y]] = \mathrm{E}[X]$$
- Toisaalta tunnetaan varianssin ehdollistamiskaava
$$\mathrm{V}[X] = \mathrm{E}[\mathrm{V}[X \mid Y]] + \mathrm{V}[\mathrm{E}\,[X \mid Y]]$$
josta
$$\mathrm{V}[Z] = \mathrm{V}[\mathrm{E}\,[X \mid Y]] = \mathrm{V}[X] - \mathrm{E}[\mathrm{V}[X \mid Y]]$$
- Varianssi pienenee
  - ehdollistaminen eliminoi X:n vaihtelun kaikkien niiden tapausten joukon sisällä, joissa ehdollistavalla muuttujalla Y on tietty kiinteä arvo

Ehdollistaminen: esimerkki 1

- Estojärjestelmään saapuu puheluita Poissonisesti intensiteetillä λ
- Halutaan arvioida, kuinka paljon puheluita keskimäärin estyy aikayksikössä
- Suoraviivaisessa simuloinnissa seurataan järjestelmän kehitystä ajanjakson T yli ja rekisteröidään estyneiden puheluiden lukumäärä L
  - estimaatti kysytylle suureelle on L/T
- Koska puheluita estyy vain, kun systeemi on estotilassa, pätee

$$L = \sum_i L_i$$

  missä L_i on i:nnessä estojaksossa estyneiden puheluiden lukumäärä

- Olkoon kyseisen estojakson kesto simuloinnissa B_i
- Koska saapumiset tapahtuvat Poissonisesti, tiedetään

$$\mathrm{E}\left[L_i \,|\, B_i\right] = \lambda B_i$$

Ehdollistaminen: esimerkki 1 jatkuu

- Sen sijaan, että kerättäisiin estyneiden puheluiden lukumäärää, kannattakin kerätä estotilan kokonaiskestoa $\sum_i B_i$
- Estyneiden puheluiden lukumäärän ehdollinen odostusarvo on nyt

$$\bar{L} = \lambda \sum_i B_i$$

  ja uusi estimaattori kysytylle suureelle on $\bar{L}/T$
- Simuloinnin kulku ei muutu mitenkään
  - nyt vain rekisteröidään L_i:n asemesta suureen $\mathrm{E}\left[L_i \,|\, B_i\right] = \lambda B_i$ arvoja
  - estojakson aikana saapuvien puheluiden virran arvonnasta aiheutuva varianssi eliminoituu

Ehdollistaminen: esimerkki 2

- Tehtävänä on arvioida keskimääräinen viipymäaika */M/1-FIFO -jonossa
- Suoraviivaisessa simuloinnissa seurataan järjestelmän kehitystä, kunnes n asiakasta on palveltu
  - mitataan kunkin asiakkaan viipymäaika $T_i,\ i = 1, 2, \ldots, n$
  - estimaattori on $\frac{1}{n}\sum_i T_i$
- Jos i:nnen asiakkaan saapuessa jonoon jonossa on edellä m_i asiakasta, niin tähän lukumäärään ehdollistettuna pätee

$$\mathrm{E}\left[T_i \,|\, m_i\right] = (m_i + 1)/\mu$$

missä $1/\mu$ on yhden asiakkaan keskimääräinen palveluaika jonossa (huom! muistittomuus)

Ehdollistaminen: esimerkki 2 jatkuu

- T_i:n asemesta on edullista ottaa havaittavaksi suureeksi $\mathrm{E}\left[T_i \,|\, m_i\right]$ eli olennaisesti m_i
- Estimaattoriksi saadaan
$$\frac{1}{n}\sum_{i=1}^{n}(m_i+1)/\mu$$
- Taaskaan simuloinnin kulku ei muutu
  - jonoon saapuva asiakas näkee saman jononpituuden kuin ennenkin
  - nyt vain viipymäajan rekisteröinnistä jää pois se vaihtelu minkä jonossa olevien asiakkaiden palveluaikojen arpominen aiheuttaa
  - kaikki palveluajat on korvattu odotusarvollaan $1/\mu$

Ehdollistaminen (jatkuu)

Statistiikan keruussa on aina kun vain mahdollista edullista korvata satunnaismuuttuja (kulloisenkiin dynaamiseen tilanteeseen ehdollistetulla) odotusarvollaan

- Pieni varaus edelliseen: varianssin reduktio ei ole aina taattu
  - simuloinnin aikana saatavat näytteet satunnaismuuttujasta eivät ole riippumattomia vaan yleensä positiivisesti korreloituneita, mikä kasvattaa keskiarvon varianssia riippumattomiin näytteisiin verrattuna
  - ehdollistaminen voi tehdä korrelaatiot vielä voimakkaammaksi
  - periaatteessa on mahdollista, että tämä kumoaa yksittäisen pisteen varianssin reduktion (käytännössä lienee harvinaista)

RESTART (Repetitive Simulation Trials After Reaching Thresholds)

- RESTART on harvinaisten tapausten simulointiin tarkoitettu kiihdytysmenetelmä
- Menetelmä on perusidealtaan yksinkertainen ja helppo soveltaa mihin tahansa tapaukseen (ei vaadi monimutkaista analyysiä)
- Tässäkin on kysymyksessä eräänlainen ehdollistamismenetelmä
  - mutta nyt ehdollinenkin odotusarvo estimoidaan simuloimalla
- Oletetaan, että ollaan kiinnostuneita harvinaisesta tapahtumasta A
  - esim. A voi tarkoittaa, että muuttuja X_t ylittää jonkin (korkean) tason L

$$A = \{X_t > L\}$$

- Olkoon $C \supset A$ toinen, vähemmän harvinainen tapahtuma, jolle pätee

$$1 \gg \mathrm{P}\{C\} \gg \mathrm{P}\{A\}$$

  - esim. C voi tarkoittaa, että muuttuja X_t ylittää jonkin alemman kynnystason T

$$C = \{X_t > T\}$$

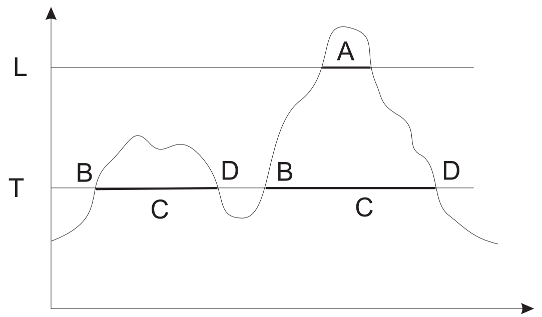
RESTART (jatkuu)

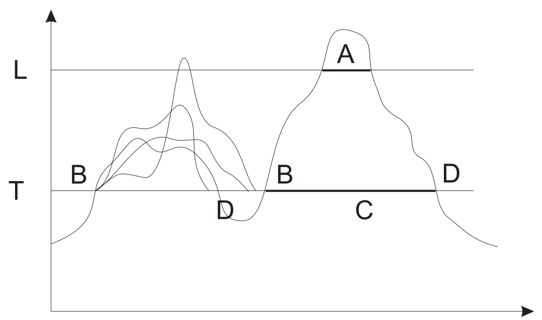
- Ehdollisen todennäköisyyden määritelmän mukaan pätee

$$\mathrm{P}\{A\} = \mathrm{P}\{C\}\mathrm{P}\{A \,|\, C\}$$

- Tavallisella simuloinnilla on helppo saada tarkka estimaatti todennäköisyydelle $\mathrm{P}\{C\}$
- Sen sijaan ehdollista todennäköisyyttä $\mathrm{P}\{A \,|\, C\}$ on vaikea estimoida, koska C esiintyy harvoin
- RESTARTissa $\mathrm{P}\{A \,|\, C\}$:n estimaattia tarkennetaan tekemällä useita toistoja niistä simulointiajon (prosessin X_t) osista, joissa C esiintyy
  - koska C itsessään on oletettu suhteellisen harvinaiseksi C-tapahtuman toistot eivät simulointiajan kannalta ole kovin raskaita

RESTART (jatkuu)





Tavallinen simulointi (ylempi kuva) ja RESTART-menetelmä (alempi kuva)

– tapahtuman B sattuessa systeemin tila talletetaan

– kun on päästy pisteeseen D eli tapahtuma C loppuu, edellisen tapahtuman B tila palautetaan ja jakso B–D simuloidaan uudelleen

– tämä proseduuri toistetaan n kertaa

– n:nnellä kerralla simulointia jatketaan tavalliseen tapaan pisteestä D eteenpäin, kunnes tullaan uuteen pisteeseen B, josta eteenpäin tehdään jälleen n toistoa seuraavaan pisteeseen D asti

– harvinaisen tapahtuman A esiinsaamiseksi ne jaksot, joissa se voi ylipäätään esiintyä (jakso C), toistetaan n kertaa

RESTART (jatkuu)

- RESTART-menetelmällä voidaan saavuttaa varsin dramaattisia säästöjä simulointiajassa
- Jos järjestelmä on mutkikas, niin merkittävä kustannus laskennassa muodostuu tilan B tallettamisesta ja palauttamisesta
  - järjestelmän koko tila (kaikkien muuttujien arvot) on talletettava
- Menetelmän tehostamiseksi siitä on kehitetty variantteja, joissa käytetään useampia kynnystasoja ja/tai hystereesiä

Tärkeysotanta (importance sampling, IS) (1)

- Tärkeysotanta on yksi tehokkaimmista keinoista näytteiden varianssin pienentämiseksi
  - se vaatii myös huolellisimman analyysin systeemin ominaisuuksista
- Tarkastellaan diskreettiä sm:aa X, jolla on tila-avaruus $\mathcal{S}$, ja olkoon $X \in \mathcal{S} \sim p(x)$
- Tarkastellaan odotusarvoa

$$\alpha = \mathrm{E}[1(X \in \mathcal{A})] = \mathrm{P}\{X \in \mathcal{A}\}$$

- Estimaattori α:n arvioimiseksi on tällöin

$$\hat{\alpha} = \frac{1}{N} \sum_{n=1}^{N} 1(X_n \in \mathcal{A}),$$

jossa X_n ovat i.i.d.

Tärkeysotanta (2)

- Olkoon $X^* \sim p^*(x)$ toinen sm siten että $p^*(x) > 0, \forall x \in \mathcal{A}$
- Tällöin

$$\begin{aligned}\alpha &= \mathrm{E}[1(X \in \mathcal{A}] \\ &= \sum_{x \in \mathcal{S}} p(x)\, 1(x \in \mathcal{A}) \\ &= \sum_{x \in \mathcal{S}} p^*(x)\, \frac{p(x)}{p^*(x)} 1(x \in \mathcal{A}) \\ &= \mathrm{E}_{p^*}\left[w(X^*) 1(X^* \in \mathcal{A}\right],\end{aligned}$$

missä $w(x) = p(x)/p^*(x)$ on ns. uskottavuusmäärä (engl. likelihood ratio).

- Tästä saadaan uusi estimaattori

$$\hat{\alpha} = \sum_{n=1}^{N} w(X_n^*)\, 1(X_n^* \in \mathcal{A})$$

- IS mahdollistaa siis näytteiden generoimisen jakaumasta $p^*(x)$, ja sen aiheuttama harha korjataan painottamalla näytteet sopivasti.

Optimaalinen IS (1)

- Oletetaan että

$$p^*(x) = \mathrm{P}\{X = x \,|\, X \in \mathcal{A}\} = \frac{p(x)}{\mathrm{P}\{X \in \mathcal{A}\}}$$

- Tällöin $w(x) = p(x)/p^*(x) = \mathrm{P}\{X \in \mathcal{A}\}$ ja

$$\begin{aligned}\hat{\alpha} &= \frac{1}{N}\sum_{n=1}^{N} w(X_n^*)1(X_n^* \in \mathcal{A}) \\ &= \frac{1}{N}(N \cdot \mathrm{P}\{X \in \mathcal{A}\} = \mathrm{P}\{X \in \mathcal{A}\} \qquad \text{(tarkasti!)}\end{aligned}$$

- Ongelma: $w(x)$:n laskemiseksi täytyy tuntea $\mathrm{P}\{X \in \mathcal{A}\}$, joka on juuri se suure, jota pyritään estimoimaan.
- Idea: hyvä IS jakauma $p^*(x)$ pyrkii aproksimoimaan optimaalista jakaumaa mahdollisimman paljon, mutta uskottavuusmäärä $w(x)$ on silti laskettavissa helposti.

Optimaalinen IS (2)

- Tarkastellaan muuttujaa $\mathrm{V}[w(X^*)\,1(X^* \in \mathcal{A})]$. Sille voidaan johtaa

$$\begin{aligned} \mathrm{V}[w(X^*)\,1(X^* \in \mathcal{A})] &= \frac{\alpha^2}{\alpha^*} - \alpha^2 + \alpha^*(\sigma^*)^2, \quad \text{missä} \\ \alpha &= \mathrm{E}[1(X \in \mathcal{A})], \\ \alpha^* &= \mathrm{E}_{p^*}\left[1(X^* \in \mathcal{A})\right], \\ (\sigma^*)^2 &= \mathrm{V}_{p^*}\left[w(X^*) \,|\, X^* \in \mathcal{A}\right]. \end{aligned}$$

- Havaintoja
  - $p^*(x) = p(x) \Rightarrow \mathrm{V}[*] = \alpha - \alpha^2$
  - Kasvattamalla $\alpha^* \Rightarrow \alpha^2/\alpha^*$ pienenee
  - Ideaalisesti, jos $\alpha^* = 1$ ja $\sigma^* = 0 \Rightarrow \mathrm{V}[*] = 0$
  - Käytännössä pyritään kasvattamaan α^*:n arvoa ja pitämään huoli, että σ^* ei kasva liian suureksi.

IS:n sovelluksia

- Staattinen simulointi
  - Edellä tarkasteltiin tilannetta, jossa näytteet olivat i.i.d. Simuloinnissa tämä vastaa staattista Monte Carlo simulointia.
  - Sovelluksia: monidimensioisten integraalien (summien) arviointi, esimerkiksi estotodennäköisyydet häviömalleissa
- Dynaaminen simulointi
  - Joillekin yksinkertaisille systeemeille on mahdollista johtaa ns. asymptoottisesti optimaalisia IS jakaumia (esim. M/M/1 jono)
  - Tyypillisesti tarkasteltavat systeemit ovat niin monimutkaisia, että asymptoottisesti optimaalisia IS jakaumia ei voida johtaa. Tällöin on mahdollista kehittää adaptiivisia menetelmiä, jotka iteratiivisesti pyrkivät etsimään parhaan ratkaisun.