Algoritmi za uvjetno uzorkovanje u skrivenim Markovljevim modelima

Banek, Erik

SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO–MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Erik Banek

ALGORITMI ZA UVJETNO UZORKOVANJE U SKRIVENIM MARKOVLJEVIM MODELIMA

Diplomski rad

Voditelj rada:
prof. dr. sc. Miljenko Huzak

Zagreb, travanj, 2020.

Ovaj diplomski rad obranjen je dana ____________________ pred ispitnim povjerenstvom u sastavu:

1. ____________________________, predsjednik
2. ____________________________, član
3. ____________________________, član

Povjerenstvo je rad ocijenilo ocjenom _______________.

Potpisi članova povjerenstva:

1. ____________________________
2. ____________________________
3. ____________________________

Sadržaj

Uvod

Skrivene Markovljeve modele (skraćeno SMM-ove) dobivamo kada se zapitamo: što ako ne možemo direktno promatrati stanja Markovljevog lanca? Jedan od čestih primjera je Markovljev lanac kojemu je dodan gaussovski šum.

Šumovita stanja koja direktno promatramo u SMM-ovima, naša opažanja, nazivamo **opaženim stanjima**. Vrijednosti koje su opažena stanja poprimila su naši podaci. Stanja unutarnjeg Markovljevog lanca, koja su skrivena šumom, nazivamo **skrivenim stanjima**. U SMM-ovima se često pitamo što možemo saznati o skrivenim stanjima iz vrijednosti onih opaženih.

Skriveni Markovljevi modeli su model za stohastički proces koji opisuje nastanak niza podataka. Primjene su našli u prijenosu govora u tekst [17], ekonometriji [8], računarskoj biologiji [13], medicini [14] i mnogim drugim područjima [2].

Dva algoritma vrlo raširene uporabe, **Kalmanov filter** i **čestični filter**, se mogu primijeniti na SMM-ove. Ti algoritmi su prirodno podložni matematičkoj obradi u kontekstu SMM-ova i to nam daje motivaciju za dublje proučavanje tih modela.

U ovom diplomskom radu dat ćemo uvod u SMM-ove nad izmjerivim prostorom stanja i posebno obratiti pozornost na familiju algoritama u koju spada čestični filter. Za računanje vrijednosti skrivenih stanja ćemo primjeniti metode uzorkovanja (Monte Carlo metode). Čestični filter nad SMM-ovima bit će jedan od specijalnih slučajeva algoritma **sekvencijalnog uzorkovanja s reuzorkovanjem** koji je glavni fokus ovoga rada. Za taj algoritam ćemo dati dokaze asimptotske konzistentnosti i normalnosti.

Glavni izvor na kojemu se ovaj rad bazira jest [3].

Poglavlje 1

Osnovna svojstva skrivenih Markovljevih modela

U ovom poglavlju opisujemo i definiramo skrivene Markovljeve modele koji poprimaju vrijednosti na općenito izmjerivom prostoru. Lanci nemaju uvijek konačan broj stanja, te zato prijelazne vjerojatnosti ne možemo više opisivati samo matricama.

Za poopćenje definicija poput *stacionarnosti, ireducibilnosti, strogog Markovljevog svojstva, svojstva zaboravljanja* u Markovljevim lancima, čitatelja upućujemo na poglavlje 14 u [3] ili detaljnije na [16]. Ta svojstva ovdje nećemo koristiti.

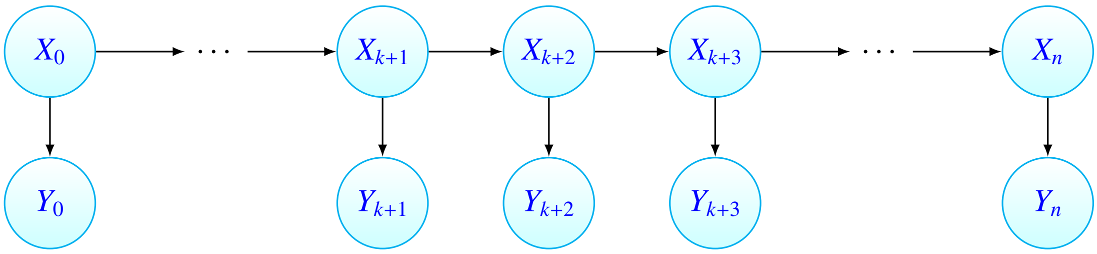
Ali prvo, objasnimo što su SMM-ovi i motivirajmo naš rad s nekim primjerima.

1.1 Primjeri skrivenih Markovljevih modela

Skriven Markovljev model opisuje način konstrukcije bivarijatnog stohastičkog procesa $\{X_n, Y_n\}_{n\geq 0}$ u diskretnom vremenu. Stanja proces $\{X_n\}$ nazivamo skrivenim stanjima te ona zasebno čine proces koji je Markovljev lanac.

Stanja procesa $\{Y_n\}$ nazivamo opaženim stanjima, jer su vrijednosti koje ta stanja poprime zapravo naši podaci. Ako su y_i brojevi koji su naši podaci, u SMM-ovima često uvjetujemo na događaje oblika $\{Y_0 = y_0, Y_1 = y_1, \dots, Y_k = y_k\}$. Uvjetovanjem na takve događaje, cilj je saznati što više o skrivenim stanjima $\{X_n\}$.

SMM-ovi su generativni modeli jer direktno opisuju kako nastaju podaci. Jezikom vjerojatnosnih grafičkih modela [11], gdje modele definiramo strukturom uvjetne zavisnosti, SMM-ove možemo grafički definirati kao na slici 1.1. Dokaz takve strukture je posljedica Markovljevosti procesa $\{X_n\}$ i propozicije 1.3.5.



Slika 1.1: Skriven Markovljev model s n unutarnjih stanja, prikazan grafički pomoću uvjetnih zavisnosti

Način na koji shvaćamo opažena stanja $\{Y_n\}$ može se mijenjati od primjene do primjene. Već smo prije spomenuli slučaj kada su ona jednaka skrivenim stanjima s dodanim šumom, npr. $Y_k = X_k + U_k$, gdje je U_k gaussovska slučajna varijabla za svaki k.

Konačni skriveni Markovljevi modeli

Ako je broj stanja u modelu konačan (i proces je homogen kroz vrijeme), ponašanje takvog SMM-a možemo u potpunosti opisati s dvije matrice. To su matrica prijelaznih vjerojatnosti skrivenog lanca, i matrica vjerojatnosti opažanja s obzirom na skriveno stanje.

Iduća dva primjera su iz područja procesiranja signala. Takve modele kanala možemo koristiti za statističke procjene njihove učinkovitosti.

Primjer 1.1.1 (Gilbert-Elliotov model kanala)**.** Ovaj primjer predstavlja situaciju kada želimo preko nekog šumovitog kanala prenijeti bitovnu poruku. Bitovnu poruku ćemo predstaviti kao niz nezavisnih i jednako distribuiranih Bernoullijevih varijabli s vjerojatnošću $\frac{1}{2}$. Odnosno, za svako unutarnje stanje vrijedi $X_k \sim B(\frac{1}{2})$ te su unutarnja stanja nezavisna i jednako distribuirana (n.j.d.).

Poruku koja je prenesena, onu šumovitu, modelirat će opažena stanja $\{Y_n\}$, tako da za svaki k vrijedi

$$P(Y_k = 1|X_k = 1) = P(Y_k = 0|X_k = 0) = q$$

te

$$P(Y_k = 1|X_k = 0) = P(Y_k = 0|X_k = 1) = 1 - q.$$

Broj q predstavlja vjerojatnost uspjeha prenošenja trenutnog *bita* poruke i to je neki broj blizu 1.

Primjer 1.1.2 (Gilbert-Elliotov model kanala 2)**.** Sada ćemo doraditi prethodni primjer tako da uzimamo u obzir da su greške u kanalu često zaredane. Osim poruke koju šaljemo, unutarnji će lanac $\{X_n\}$ također pamtiti je li u stanju visoke ili niske učestalosti grešaka.

Za skriveno stanje neka vrijedi da je $X_k = (X_{1,k}, X_{2,k})$ gdje je prva komponenta stanje učestalosti grešaka, s kodovima 0 = rijetke greške, 1 = česte greške. Druga komponenta od X_k je, kao u prethodnom primjeru, trenutni bit poruke koja se prenosi. Vrijednost p_0 neka bude vjerojatnost prijelaza iz stanja visoke učestalosti grešaka (1) u nisku (0), a p_1 obrnuto. Matrica prijelaza unutarnjeg lanca tada izgleda ovako

$$\begin{array}{c c}
\begin{array}{cccc} (0,0) & (0,1) & (1,0) & (1,1) \end{array} & \\
\begin{pmatrix}
\frac{1-p_1}{2} & \frac{1-p_1}{2} & \frac{p_1}{2} & \frac{p_1}{2} \\
\frac{1-p_1}{2} & \frac{1-p_1}{2} & \frac{p_1}{2} & \frac{p_1}{2} \\
\frac{p_0}{2} & \frac{p_0}{2} & \frac{1-p_0}{2} & \frac{1-p_0}{2} \\
\frac{p_0}{2} & \frac{p_0}{2} & \frac{1-p_0}{2} & \frac{1-p_0}{2}
\end{pmatrix} & \begin{array}{c} (0,0) \\ (0,1) \\ (1,0) \\ (1,1). \end{array}
\end{array}$$

Za opažena stanja vrijedi:

$$P(Y_k = b | X_k = (b, s)) = q_s$$

tako da je vjerojatnost uspjeha slanja poruke puno veća u stanju 0 nego u stanju 1, to jest vrijedi $q_0 >> q_1$.

Gaussovski linearni modeli

Promotrimo općenitu formulaciju modela oblika

$$\begin{aligned}
X_k &= AX_{k-1} + RU_k, && k \geq 1, \\
Y_k &= BX_k + SV_k, && k \geq 0,
\end{aligned}$$

gdje vrijedi:

- $\{U_n\}_{n\geq 1}$ su n.j.d. slučajni vektori dimenzije d_U tako da vrijedi $U_k \sim N(0, I_{d_U})$. Njih nazivamo procesni šum.
- $\{V_n\}_{n\geq 0}$ su n.j.d. slučajni vektori dimenzije d_V tako da vrijedi $V_k \sim N(0, I_{d_V})$. Njih nazivamo opservacijski šum.
- $X_0 \sim N(M_0, S_0)$.
- Svi $\{V_n\}, \{U_n\}, X_0$ su međusobno nezavisni.
- A, B, R, S su neslučajne matrice odgovarajućih dimenzija.

Time smo dobili opis jednog SMM-a, kojega nazivamo gaussovski linearan model. Ovaj tip modela nam je od posebnog značaja jer je jedan od modela nad realnim prostorom stanja kojemu znamo egzaktno, odnosno analitički izračunati integrale nad uvjetnim prostorima.

Algoritam računanja tih integrala naziva se *Kalmanov filter* [10], a svođenje nelinearnog modela na takav linearni pomoću prve derivacije jest *prošireni Kalmanov filter*.

Primijetimo da su R, S zapravo korijeni kovarijacijskih matrica šumova.

Primjer 1.1.3 (Autoregresijski model)**.** Definirajmo autoregresivni model $\{Z_n\}$ reda p takav da zadovoljava

$$Z_{k+1} = \phi_1 Z_k + \phi_2 Z_{k-1} + \cdots + \phi_p Z_{k-p+1} + U_k.$$

Ako označimo $X_k = (Z_k, \ldots, Z_{k-p+1})^t$ tada ekvivalentno u formi za SMM možemo zapisati

$$\begin{aligned} X_{k+1} &= AX_k + (1, 0, \ldots, 0)^t U_k, \\ Y_{k+1} &= (1, 0, \ldots, 0) X_{k+1} \end{aligned}$$

gdje je matrica A jednaka

$$A = \begin{pmatrix} \phi_1 & \phi_2 & \ldots & \phi_{p-1} & \phi_p \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix}.$$

Ako postoji šum u opažanju vrijednosti koje autoregresivni model poprima, tada ima smisla definirati $Y_{k+1} = (1, 0, \ldots, 0) X_{k+1} + V_k$.

Neprekidni skriveni Markovljevi modeli

Primjer 1.1.4 (Stohastička volatilnost)**.** ARCH model [6] je široko korišten u ekonometriji. Jedna njegova verzija je

$$\begin{aligned} X_k &= \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{k-i}^2, \\ Y_k &= \sqrt{X_k} V_k \end{aligned}$$

što jasno nije SMM jer X_k ovisi o prijašnjim Y_k. S druge strane, kao moguća alternativa ARCH-u je predložen [9] model

$$\begin{aligned} X_k &= \alpha + \delta X_{k-1} + \sigma_U U_k, & U_k &\sim N(0, 1), \\ Y_k &= exp(X_k/2) V_k, & V_k &\sim N(0, 1). \end{aligned}$$

Primijetimo da ovaj model stohastičke volatilnosti ne spada ni u jednu od prijašnje dvije kategorije konačnih i gaussovskih linearnih SMM-ova. Zbog postojanja obilja takvih općenitijih modela često se koriste aproksimativne metode za računanje integrala. Naime, razvijanje analitičkih rješenja za svaku zasebnu vrstu modela često je vremenski vrlo zahtjevno ili jednostavno nemoguće.

Mi ćemo se u ovom radu pozabaviti upravo tim aproksimativnim metodama koje rade na svim navedenim primjerima.

1.2 Markovljevi lanci nad izmjerivim prostorom stanja

Kao inače, podloga svega što radimo je vjerojatnosni prostor $(\Omega, \mathcal{F}, P)$. Poopćenje pojma prijelaznih vjerojatnosti daje nam jezgra prijelaza (*engl. transition kernel*).

Definicija 1.2.1 (Jezgra prijelaza). *Neka su* $(\mathsf{X}, \mathcal{X}), (\mathsf{Y}, \mathcal{Y})$ *izmjerivi prostori. Neka je* T *funkcija* $T : \mathsf{X} \times \mathcal{Y} \to [0, +\infty]$, *takva da vrijedi*

- $T(\cdot, A)$ *je izmjeriva funkcija za svaki* $A \in \mathcal{Y}$,
- $T(x, \cdot)$ *je mjera nad* $\mathcal{Y}$ *za svaki* $x \in \mathsf{X}$.

Tada T *zovemo nenormaliziranom jezgrom prijelaza iz* X *u* Y. *Ako još vrijedi da je* $T(x, \cdot)$ *vjerojatnosna mjera, tada* T *zovemo vjerojatnosnom jezgrom prijelaza, ili kraće, jezgrom prijelaza. Na kraju, ako je* $\mathsf{Y} = \mathsf{X}$, *tada* T *nazivamo Markovljevom jezgrom prijelaza, ili kraće Markovljevom jezgrom.*

Sada koristimo jezgre prijelaza da bismo definirali općeniti Markovljev lanac u diskretnom vremenu. Ovdje koristimo proširenu, mjerovnu definiciju uvjetnog očekivanja i uvjetne vjerojatnosti, o kojima detalje možemo pronaći u [20], poglavlje 15.

Definicija 1.2.2 (Markovljev lanac). *Neka je* $\{X_n\}_{n\geq 0}$ *slučajni proces u diskretnom vremenu koji poprima vrijednosti u izmjerivom prostoru* $(\mathsf{X}, \mathcal{X})$. *Neka je* $\{\mathcal{F}_n\}_{n\geq 0}$ *filtracija i neka je taj slučajan proces adaptiran na nju. Neka je* Q *Markovljeva jezgra na* X. *Ako za svaki* n *i svaki* $A \in \mathcal{X}$ *vrijedi*

$$P(X_{n+1} \in A | \mathcal{F}_n) = Q(X_n, A) \tag{1.1}$$

tada $\{X_n\}$ *zovemo homogeni Markovljev lanac, ili kraće Markovljev lanac. Distribuciju varijable* X_0 *zovemo inicijalnom distribucijom tog lanca. Često ju označavamo s* ν.

Moguće je imati i nehomogen Markovljev lanac gdje prijelazne vjerojatnosti ovise o vremenu. On se definira analogno, ali s jezgrama prijelaza Q_k koje ovise o vremenu. Konstrukciju Markovljevog lanca s danom inicijalnom distribucijom i Markovljevom jezgrom daje nam teorem 2.1.5. u [3].

Napomena 1.2.3 (Skraćeni zapis vektora). *Na nekim mjestima radi štednje prostora koristimo "računarsku" notaciju za vektore: umjesto da funkciju* f *nad* n *varijabli zapisujemo s* $f(x_1, x_2, \ldots, x_n)$ *često ćemo koristiti* $f(x_{1:n})$. *Pojavit će se i integriranje po mjeri na* k*-dimenzionalnom prostoru, te ćemo često umjesto* $\mu(dx_0, dx_1, \ldots, dx_{k-1})$ *zapisivati* $\mu(dx_{0:k-1})$.

Kada koristimo tu notaciju, radi dodatnog skraćivanja pisanja vektora duljine jedan, uvijek će biti $k : k$ *isto što i* k.

Napomena 1.2.4 (Konvencije o σ-algebrama). *Neka je* X *slučajna varijabla s vrijednostima u izmjerivom prostoru* $(\mathsf{X}, \mathcal{X})$. *Ta varijabla praslikama skupova iz* $\mathcal{X}$ *generira* σ*-algebru nad skupom* Ω. *Tu* σ*-algebru označavamo sa* $\sigma(X)$.

Ako je $(\mathsf{Y}, \mathcal{Y})$ *neki drugi izmjerivi prostor, sa* $\mathcal{X} \otimes \mathcal{Y}$ *označavamo klasičnu produktnu* σ*-algebru* $\mathcal{X} \otimes \mathcal{Y} = \sigma(\mathcal{X} \times \mathcal{Y})$ *(potpoglavlje 10.6. [20]). Ako na jednom prostoru gledamo produkt sa samim sobom, to ćemo označavati s* $\mathcal{X}^{\otimes n} = \mathcal{X} \otimes \mathcal{X} \otimes \cdots \otimes \mathcal{X}$*, gdje se* $\mathcal{X}$ *pojavljuje* n *puta.*

Spomenimo sada dvije leme. Prva govori o zapisu funkcija preko kompozicije sa slučajnim varijablama (teorem 8.6. [20]), a druga je zakon potpune vjerojatnosti u uvjetnom očekivanju (korolar 15.1. [20]).

Lema 1.2.5. *Ako je* X *slučajna varijabla s vrijednostima u izmjerivom prostoru* $(\mathsf{X}, \mathcal{X})$*, a* Y *(realna) slučajna varijabla te ako je* $\sigma(Y) \subset \sigma(X)$*, tada postoji izmjeriva funkcija* $g : \mathsf{X} \to \mathbb{R}$ *takva da je* $Y = g(X)$*.*

Dokaz. Dokaz je analogan prvom dijelu dokaza teorema 8.6. [20]. □

Lema 1.2.6. *Neka su* $\mathcal{D}_1$ *i* $\mathcal{D}_2$ σ*-algebre takve da je* $\mathcal{D}_1 \subset \mathcal{D}_2 \subset \mathcal{F}$*. Ako je* Y *integrabilna slučajna varijabla, tada vrijedi*

$$\mathbb{E}[Y|\mathcal{D}_1] = \mathbb{E}[\mathbb{E}[Y|\mathcal{D}_1]|\mathcal{D}_2] = \mathbb{E}[\mathbb{E}[Y|\mathcal{D}_2]|\mathcal{D}_1] \ (g.s.)$$

Poseban slučaj dobivamo kada uvrstimo $\mathcal{D}_1 = \{\emptyset, \Omega\}$*. Tada vrijedi*

$$\mathbb{E}[Y] = \mathbb{E}[Y|\mathcal{D}_2] \ (g.s.)$$

Sada pomoću tih lema dokazujemo izraz o očekivanju funkcije stanja Markovljevog lanca.

Propozicija 1.2.7. *Neka je* $\{X_n\}_{n\geq 0}$ *Markovljev lanac s vrijednostima u prostoru* $(\mathsf{X}, \mathcal{X})$*, jezgrom prijelaza* Q *i inicijalnom distribucijom* ν*. Tada za svaki* $n \in \mathbb{N}_0$ *i svaku ograničenu* $\mathcal{X}^{\otimes(n+1)}$*-izmjerivu funkciju* f *vrijedi*

$$\mathbb{E}[f(X_0, X_1, X_2, \ldots, X_n)] = \underset{\mathsf{X}^{n+1}}{\int \cdots \int} f(x_0, x_1, \ldots, x_n) \nu(dx_0) \prod_{k=1}^{n} Q(x_{k-1}, dx_k). \tag{1.2}$$

Dokaz. Distribucija od X_0 je ν, stoga bazu indukcije za $n = 0$ dobijemo pomoću teorema o zamjeni varijabli (teorem 10.11. [20])

$$\mathbb{E}[f(X_0)] = \int_{\mathsf{X}} f(x_0) \nu(dx_0).$$

Dokažimo korak indukcije. Označimo s g izmjerivu funkciju takvu da vrijedi $g(X_{0:n}) = \mathbb{E}[f(X_{0:n+1})|X_{0:n}]$, koja postoji prema 1.2.5. Koristeći redom 1.2.6 i pretpostavku indukcije primijenjenu na $g(X_{0:n})$ imamo da vrijedi

$$\mathbb{E}[f(X_{0:n+1})] = \mathbb{E}[\mathbb{E}[f(X_{0:n+1})|X_{0:n}]] = \int_{\mathsf{X}^{n+1}} \cdots \int g(x_{0:n})\nu(dx_0) \prod_{k=1}^{n} Q(x_{k-1}, dx_k).$$

Korištenjem svojstva lanca (1.1), definiciju funkcije g i izmjerivost funkcije f dobivamo

$$g(x_{0:n}) = \int_{\mathsf{X}} f(x_{0:n+1})Q(x_n, dx_{n+1})$$

što dovršava dokaz koraka indukcije. □

1.3 Skriveni Markovljevi procesi

Skriven Markovljev proces definirat ćemo kao Markovljev lanac s posebnom strukturom jezgre prijelaza. To nam olakšava izvod svojstava, te će ranije spomenuto svojstvo uvjetne nezavisnosti biti jedna od posljedica koju ćemo dokazati.

Napomena 1.3.1 (Definiranje mjera)**.** *U daljnjem tekstu, mjere ćemo uvijek opisivati preko njihovih integrala izmjerivih ograničenih funkcija nad odgovarajućim prostorima: $\mu(f) = \int f(x)\mu(dx)$. To ima smisla činiti jer integral ograničenih funkcija uvijek postoji na vjerojatnosnom prostoru.*

Tako jednostavnije vidimo koje je očekivanje funkcija, a ništa ne gubimo jer uvijek vrijedi $\mu(\mathbb{1}_A) = \mu(A)$ za izmjerivi skup A.

Također, ako je $(\mathsf{X}, \mathcal{X})$ izmjeriv prostor, prostor ograničenih, $\mathcal{X}$-izmjerivih realnih funkcija označavat ćemo s $\mathcal{F}^b(\mathsf{X})$.

Definicija 1.3.2 (Skriven Markovljev proces)**.** *Neka su $(\mathsf{X}, \mathcal{X})$, $(\mathsf{Y}, \mathcal{Y})$ izmjerivi prostori te neka je T Markovljeva jezgra na $(\mathsf{X} \times \mathsf{Y}, \mathcal{X} \otimes \mathcal{Y})$ dana s*

$$T((x, y), f) := \int_{\mathsf{X}\times\mathsf{Y}} f(x', y)Q(x, dx')G(x', dy)$$

gdje je Q Markovljeva jezgra na X, a G jezgra prijelaza iz X u Y. Tada Markovljev lanac $\{X_n, Y_n\}_{n\geq 0}$ s vrijednostima u $\mathsf{X} \times \mathsf{Y}$ i jezgrom prijelaza T nazivamo skrivenim Markovljevim procesom. Po konvenciji ćemo uzimati da je inicijalna distribucija tog lanca dana s $(\nu\times G)$, gdje je ν distribucija od X_0 te za $f \in \mathcal{F}^b(\mathsf{X} \times \mathsf{Y})$ vrijedi

$$(\nu \times G)(f) = \int_{\mathsf{X}\times\mathsf{Y}} f(x, y)G(x, dy)\nu(dx).$$

Ova definicija pokriva vrlo široku klasu modela. Ipak, kada modeli nisu konačni, zahtijevamo malo jači uvjet kao što slijedi.

Definicija 1.3.3 (Parcijalno dominiran skriven Markovljev proces)**.** *Ako za jezgru prijelaza G iz definicije 1.3.2 postoji $\mathcal{X} \otimes \mathcal{Y}$-izmjeriva funkcija $g : \mathsf{X} \times \mathsf{Y} \to [0, \infty]$ i mjera μ na $(\mathsf{Y}, \mathcal{Y})$, takvi da za svaki $x \in \mathsf{X}$ vrijedi*

$$G(x, f) = \int_{\mathsf{Y}} f(y) g(x, y) \mu(dy)$$

tada taj skriven Markovljev proces nazivamo ***parcijalno dominiran skriven Markovljev proces****. Funkciju g nazivamo prijelaznom funkcijom gustoće.*

Daljnju analizu i algoritme razrađivat ćemo upravo na parcijalno dominiranim SMM-ovima.

SMM-ovi u kojima je prostor Y konačan nisu parcijalno dominirani, ali se za takve konačne modele mogu izvesti algoritmi koji nisu aproksimativni (poglavlje 3 u [3]), stoga ne gubimo puno u vidu općenitosti u kojoj radimo.

Uzimajući parcijalno dominiran skriven model iz 1.3.3 i primjenjujući propoziciju 1.2.7 o integralima nad stanjima Markovljevog lanca direktno dobivamo

$$\begin{aligned}
&\mathbb{E}[f(X_0, Y_0, X_1, Y_1, \dots, X_n, Y_n)] \\
&= \int \cdots \int_{\mathsf{X}^n \times \mathsf{Y}^n} f(x_0, y_0, x_1, y_1, \dots, x_n, y_n) \nu(dx_0) g(x_0, y_0) \mu(dy_0) \prod_{k=1}^{n} Q(x_{k-1}, dx_k) g(x_k, y_k) \mu(dy_k)
\end{aligned} \tag{1.3}$$

Prije dokaza uvjetne nezavisnosti opaženih stanja SMM-a, spomenimo lemu koju koristimo u dokazima jednakosti koje uključuju uvjetno očekivanje.

Lema 1.3.4. *Neka su $(\mathsf{X}, \mathcal{X}), (\mathsf{Y}, \mathcal{Y})$ izmjerivi prostori, te $f : \mathsf{X} \to \mathbb{R}, g : \mathsf{Y} \to \mathbb{R}$ izmjerive funkcije. Neka je redom X slučajna varijabla s vrijednostima u X, a Y slučajna varijabla s vrijednostima u Y. Ako za svaku izmjerivu funkciju $h : \mathsf{Y} \to \mathbb{R}$ vrijedi da $\mathbb{E}[f(X)h(Y)] = \mathbb{E}[g(Y)h(Y)]$, tada je $\mathbb{E}[f(X)|Y] = g(Y)$ $(g.s.)$.*

Dokaz. Dovoljno je dokazati da vrijedi $E[f(X)Z] = E[g(Y)Z]$ za svaku $\sigma(Y)$-izmjerivu slučajnu varijablu Z. Budući da je Z $\sigma(Y)$-izmjeriva varijabla, tada vrijedi $\sigma(Z) \subset \sigma(Y)$ te prema lemi 1.2.5 postoji h' takva da vrijedi $Z = h'(Y)$. Tvrdnja sada slijedi iz pretpostavke primijenjene na funkciju h'. □

Sada smo spremni dokazati nezavisnost opaženih stanja uvjetno na pripadna skrivena stanja.

Propozicija 1.3.5. *Neka je* $\{X_n, Y_n\}_{n\geq 0}$ *skriven Markovljev proces s vrijednostima u* $\mathsf{X}\times\mathsf{Y}$ *i oznakama kao u 1.3.2. Tada za svaki neprazan podskup* $\{k_1,\dots,k_p\}\subseteq\{0,1,\dots,n\}$ *vrijedi*

$$\mathbb{E}[\prod_{k\in\{k_1,\dots,k_p\}} f_k(Y_k)|X_{k_1},\dots,X_{k_p}] = \prod_{k\in\{k_1,\dots,k_p\}}\int_Y f_k(y_k)G(X_k,dy_k)\ (g.s.)$$

Dokaz. Koristimo lemu 1.3.4. Neka je h funkcija iz $\mathcal{F}^b(\mathsf{X}^{p+1})$. Prema (1.3) vrijedi, za $n\geq k_p$

$$\begin{aligned}
&\mathbb{E}[h(X_{k_1},\dots,X_{k_p})\prod_{k\in\{k_1,\dots,k_p\}} f_k(Y_k)]\\
&=\int\cdots\int_{\mathsf{X}^n\times\mathsf{Y}^n}\prod_{k\in\{k_1,\dots,k_p\}} f_k(y_k)h(x_{k_1},\dots,x_{k_p})\nu(dx_0)G(x_0,dy_0)\prod_{l=1}^n Q(x_{l-1},dx_l)G(x_l,dy_l).
\end{aligned}$$

Integriranjem konstantne funkcije pod integralom za svaki $k\notin\{k_1,\dots,k_p\}$ po $G(x_k,dy_k)$, i zatim algebarskim manipulacijama (vađenje konstanti iz integrala) taj je izraz jednak

$$\begin{aligned}
&\int\cdots\int_{\mathsf{X}^n\times\mathsf{Y}^p}\prod_{k\ \in\{k_1,\dots,k_p\}} f_k(y_k)G(x_k,dy_k)h(x_{k_1},\dots,x_{k_p})\nu(dx_0)\prod_{l\ =1}^n Q(x_{l-1},dx_l)\\
&=\int\cdots\int_{\mathsf{X}^n}\left(\prod_{k\in\{k_1,\dots,k_p\}}\int_{\mathsf{Y}} f_k(y_k)G(x_k,dy_k)\right)h(x_{k_1},\dots,x_{k_p})\nu(dx_0)\prod_{l=1}^n Q(x_{l-1},dx_l)\\
&=\mathbb{E}[h(X_{k_1},\dots,X_{k_p})\prod_{k\in\{k_1,\dots,k_p\}}\int f_k(y_k)G(X_k,dy_k)].
\end{aligned}$$

U posljednjoj jednakosti smo koristili Markovljevu strukturu nad skrivenim stanjima u SMM-u, to jest izraz (1.2). □

Izglađivanje

Ponovimo izraz (1.3) za očekivanje nad stanjima SMM-a

$$\begin{aligned}
&\mathbb{E}[f(X_0,Y_0,X_1,Y_1,\dots,X_n,Y_n)]\\
&=\int\cdots\int_{\mathsf{X}^{n+1}\times\mathsf{Y}^{n+1}} f(x_0,y_0,x_1,y_1,\dots,x_n,y_n)\nu(dx_0)g(x_0,y_0)\mu(dy_0)\prod_{k=1}^n Q(x_{k-1},dx_k)g(x_k,y_k)\mu(dy_k).
\end{aligned}$$

U ostatku ovoga rada opažena će nam stanja biti fiksna i uvijek ćemo na neki način uvjetovati na događaje oblika $\{Y_0=y_0,\dots,Y_n=y_n\}$. Zbog toga će nam funkcija L_n, gdje je

$$L_n(y_0,y_1,\dots,y_n)=\int\cdots\int_{\mathsf{X}^{n+1}}\nu(dx_0)g(x_0,y_0)\prod_{k=1}^n Q(x_{k-1},dx_k)g(x_k,y_k). \tag{1.4}$$

biti od velike važnosti. Ako u izrazu (1.3) marginaliziramo X_i-eve dobivamo

$$\mathbb{E}[f(Y_0, Y_1, \ldots, Y_n)] = \int_{\mathsf{Y}^{n+1}} \cdots \int f(y_0, y_1, \ldots, y_n) L_n(y_0, y_1, \ldots, y_n) \mu_n(dy_0, dy_1, \ldots, dy_n). \quad (1.5)$$

gdje smo s μ_n označili produktnu mjeru $\mu_n = \mu \times \ldots \times \mu$ nad $(\mathsf{Y}^{n+1}, \mathcal{Y}^{\otimes(n+1)})$. Gornja jednakost pokazuje da je L_n zapravo (marginalna) funkcija gustoće slučajnog vektora $(Y_{0:n})$ s obzirom na mjeru μ_n.

Sada kada smo pripremili teren, konačno imamo sve materijale za definiranje problema kojime se bavimo u ovome radu. A taj problem je računanje integrala $\mathbb{E}[f(X_{l:k}) | Y_{0:n} = y_{0:n}]$. To računanje integrala skraćeno ćemo nazivati *računanje distribucije stanja*.

Želimo saznati očekivanje proizvoljne funkcije nad skrivenim stanjima, uvjetno na poprimljene vrijednosti opaženih stanja. Funkcija f opisuje ono što želimo saznati o tim skrivenim stanjima.

Ovaj problem ima i bayesovsku interpretaciju [5]. Imamo prijašnje (*prior*) znanje o skrivenim stanjima $\{X_n\}$, a to znanje je zapravo distribucija skrivenih stanja. Također, znamo kako skrivena stanja proizvode opažanja (*likelihood*), to jest poznajemo distribuciju $\{Y_n\}$ uvjetno na $\{X_n\}$. Na kraju, imamo vrijednosti opservacija, vrijednosti koje su $\{Y_n\}$ poprimili. Cilj nam je vidjeti kako to mijenja naše znanje o skrivenim stanjima, i to opisujemo distribucijom od $\{X_n\}$ uvjetno na $\{Y_0 = y_0, \ldots, Y_n = y_n\}$ (*posterior*).

Definirajmo izraz $\mathbb{E}[f(X_{l:k}) | Y_{0:n}]$ kojim se bavimo.

Definicija 1.3.6. *[Izglađivanje] Neka su k, l, n pozitivni indeksi tako da $k \leq l$. Neka je $\{X_n, Y_n\}_{n \geq 0}$ skriveni Markovljev proces kao u 1.3.3. Neka je $\phi_{k:l|n}$ jezgra prijelaza iz Y^{n+1} u X^{l-k+1} takva da vrijedi*

$$\mathbb{E}[f(X_{k:l}) | Y_{0:n}] = \int_{\mathsf{X}^{l-k+1}} \cdots \int f(x_{k:l}) \phi_{k:l|n}(Y_{0:n}, dx_{k:l}) \ (g.s.)$$

Takvu jezgru zovemo jezgra izglađivanja. Ovisno o izboru k, l, n postoji nekoliko slučajeva kojima dodjeljujemo posebna imena:

- *(zajedničko izglađivanje) $\phi_{0:n|n}$ - računanje distribucije svih stanja koja su proizvela naša opažanja,*
- *(marginalno izglađivanje) $\phi_{k|n}$ - računanje distribucije jednog skrivenog stanja,*
- *(filtriracija) $\phi_{n|n}$, pišemo i ϕ_n - računanje distribucije zadnjeg skrivenog stanja koje je proizvelo naša opažanja,*
- *(p-koračna predikcija) $\phi_{n+p|n}$, računanje distribucije budućih stanja.*

Sada dokazujemo formulu za izglađivanje

Propozicija 1.3.7. *Neka je $\{X_k, Y_k\}_{k\geq 0}$ skriven Markovljev proces kao u definiciji 1.3.3 i neka je $n \in \mathbb{N}$. Također neka vrijedi $L_n(y_{0:n}) > 0$ za svaki niz opservacija $y_{0:n}$ iz Y^{n+1}. Tada za svaki $y_{0:n}$ vrijedi*

$$\phi_{0:n|n}(y_{0:n}, f) = L_n^{-1}(y_{0:n}) \int \cdots \int_{\mathsf{X}^{n+1}} f(x_{0:n})\nu(dx_0)g(x_0, y_0) \prod_{l=1}^{n} Q(x_{l-1}, dx_l)g(x_l, y_l). \quad (1.6)$$

Dokaz. Ponovno koristimo lemu 1.3.4 te uzimamo funkciju h iz $\mathcal{F}^b(\mathsf{X}^{n+1})$. Prema (1.3) vrijedi

$$\begin{aligned}
\mathbb{E}[f(X_{0:n})h(Y_{0:n})] &= \int \cdots \int_{\mathsf{X}^{n+1}\times\mathsf{Y}^{n+1}} f(x_{0:n})h(y_{0:n})\nu(dx_0)g(x_0, y_0) \prod_{l=1}^{n} Q(x_{l-1}, dx_l)g(x_l, y_l)\mu_n(dy_{0:n}) \\
&= \int \cdots \int_{\mathsf{Y}^{n+1}} \left(\int \cdots \int_{\mathsf{X}^{n+1}} f(x_{0:n})\nu(dx_0) \prod_{l=1}^{n} Q(x_{l-1}, dx_l)g(x_l, y_l) \right) h(y_{0:n})\mu_n(dy_{0:n}) \\
&= \int \cdots \int_{\mathsf{Y}^{n+1}} \phi_{0:n|n}(y_{0:n}, f)L_{\nu,n}(y_{0:n})h(y_{0:n})\mu_n(dy_{0:n}) \\
&= \mathbb{E}[\phi_{0:n|n}(Y_{0:n}, f)h(Y_{0:n})]
\end{aligned}$$

gdje smo u zadnjoj jednakosti koristili (1.5). □

Za dobivanje slučajeva filtriranja i marginalnog izglađivanja dovoljno je marginalizirati izraz za zajedničko izglađivanje. Za proizvoljnu $f \in \mathcal{F}^b(\mathsf{X})$ definiramo $f'(x_{0:n}) := f(x_k)$. Očito vrijedi da onda i $f' \in \mathcal{F}^b(\mathsf{X}^{n+1})$ stoga imamo

$$\begin{aligned}
&\phi_{k|n}(Y_{0:n}, f) = \mathbb{E}[f(X_k)|Y_{0:n}] = \mathbb{E}[f'(X_{0:n})|Y_{0:n}] = \phi_{0:n|n}(Y_{0:n}, f') \\
&= \int \cdots \int_{\mathsf{X}^{n+1}} f'(x_{0:n})\phi_{0:n|n}(Y_{0:n}, dx_{0:n}) = \int \cdots \int_{\mathsf{X}^{n+1}} f(x_k)\phi_{0:n|n}(Y_{0:n}, dx_{0:n})\ (g.s.)
\end{aligned}$$

Eksplicitni algoritmi za računanje izraza (1.6) su mogući u slučaju modela s konačnim brojem stanja i gaussovskih linearnih modela. Ti algoritmi često koriste linearnu vremensku strukturu opažanja i redom računaju $\phi_{0:k|k}$ prema rastućem k. U tim situacijama je cilj dobiti numerički stabilne rekurzije (potpoglavlje 3.2. u [3]).

U mnogim drugim slučajevima, analitička formula za vjerodostojnost $L_k(\cdot)$ nije moguća, i koriste se aproksimativni algoritmi za račun integrala $\phi_{0:n|n}(y_{0:n}, f)$. Mi ćemo se baviti upravo takvim algoritmima.

Zanemarivanje funkcijske ovisnosti o opaženim stanjima

Izrazi $L_k(y_{0:k})$, $\phi_{0:n|n}(f, y_{0:n})$ ovise o vrijednostima koja su opažena stanja poprimila. Da još skratimo zapis, nećemo više zapisivati funkcijsku ovisnost o $y_{0:n}$, nego ju samo implicitno pretpostaviti, s obzirom na to da su te vrijednosti iste kroz tijek jednog algoritma. Uz to što ih ne navodimo, drukčije ćemo govoriti o objektima kojima baratamo.

Tako će funkcija $L_k(\cdot)$ postati konstanta L_k, a jezgra prijelaza $\phi_{0:n|n}$ postati mjera nad prostorom $(\mathsf{X}, \mathcal{X})$. S druge strane, funkcija gustoće jezgre prijelaza opažanja g eksplicitno ovisi o vrijednosti opaženog stanja y_k za neki indeks k, pa ćemo odsada nadalje zapisivati: $g_k(x) := g(x, y_k)$ za svaki indeks k.

U nekim drugim slučajevima ima smisla provoditi analizu u ovisnosti o $y_{0:n}$, a za to upućujemo čitatelja na potpoglavlje 3.1.4. u [3].

Poglavlje 2

Uzorkovanje po važnosti

U ovom poglavlju opisujemo metodu uzorkovanja po važnosti i njena dva proširenja, samonormalizirajuću verziju te uzorkovanje po važnosti s reuzorkovanjem, i za svaku od tih metoda dokazujemo njihovu asimptotsku normalnost.

U potpoglavlju 2.3 dokazujemo asimptotsku normalnost metode s reuzorkovanjem. Dijelove tog dokaza ćemo iskoristiti kasnije za dokaz asimptotske normalnosti sekvencijalnog uzorkovanja s reuzorkovanjem. Za analizu će nam trebati dvostruki nizovi, stoga njih i prikladne definicije konzistentnosti i asimptotske normalnosti uvodimo u potpoglavlju 2.3.

Za analizu metoda koje navodimo postoje i neasimptotski rezultati greške koji su proširenja Hoeffdingove i Marcinkiewicz-Zygmundove nejednakosti. Čitatelja upućujemo na poglavlje 9. u [3].

2.1 Osnovni algoritam i asimptotika uzorkovanja po važnosti

U ovom potpoglavlju opisujemo dvije metode uzorkovanja po važnosti. Te metode čine osnovu Monte Carlo algoritama za izglađivanje.

Računanje integrala uzorkovanjem po važnosti

Cilj nam je izračunati očekivanje neke integrabilne funkcije $f : \mathsf{X} \to \mathbb{R}$ s obzirom na vjerojatnosnu mjeru μ na izmjerivom prostoru $(\mathsf{X}, \mathcal{X})$. U svim metodama implicitno pretpostavljamo da je evaluacija funkcije f jednostavna, odnosno da taj dio znamo riješiti, a fokus stavljamo na dio uzorkovanja.

Procjenitelj

$$\mu_n^{MC}(f) = \frac{1}{n}\sum_{i=1}^{n} f(\zeta_i),$$

nazivamo Monte Carlo procjeniteljem očekivanja od f, gdje su varijable ζ_i nezavisne jednako distribuirane slučajne varijable (n.j.d.), s distribucijom po μ i vrijednostima u X.

Pretpostavimo da ne možemo uzorkovati direktno iz μ, a želimo procijeniti integral $\mu(f)$. Ako postoji neka druga vjerojatnosna mjera ν na $(\mathsf{X}, \mathcal{X})$, takva da je $\mu \ll \nu$ (μ je apsolutno neprekidna u odnosu na ν), te ako sa $\frac{d\mu}{d\nu}$ označimo verziju Radon-Nikodymove derivacije, tada vrijedi jednakost

$$\mu(f) = \int_{\mathsf{X}} f(x)\mu(dx) = \int_{\mathsf{X}} f(x)\frac{d\mu}{d\nu}(x)\nu(dx).$$

To nam sugerira drugačiji procjenitelj

$$\tilde{\mu}_n^{UV}(f) = \frac{1}{n}\sum_{i=1}^{n} f(\xi_i)\frac{d\mu}{d\nu}(\xi_i) \tag{2.1}$$

gdje su ξ_i n.j.d. po ν. Za praktično korištenje ovog procjenitelja pretpostavljamo da možemo uzorkovati iz ν te da nam je poznata Radon-Nikodymova derivacija. Korištenje procjenitelja $\tilde{\mu}_{UV}^n$ za račun $\mu(f)$ nazivamo *uzorkovanje po važnosti*. Mjeru ν nazivamo *instrumentalna distribucija*.

Direktnom primjenom Kolmogorovljevog jakog zakona velikih brojeva dobivamo

$$\frac{1}{n}\sum_{i=1}^{n} \frac{d\mu}{d\nu}(\xi_i) \xrightarrow[n\to\infty]{(g.s.)} 1.$$

Ako zamijenimo dijeljenje s n u (2.1) dijeljenjem sa sumom $\sum_{j=1}^{n} \frac{d\mu}{d\nu}(\xi_j)$, tada dobivamo *samo-normalizirajuću verziju* uzorkovanja po važnosti

$$\hat{\mu}_n^{UV}(f) = \sum_{i=1}^{n} f(\xi_i)\frac{\frac{d\mu}{d\nu}(\xi_i)}{\sum_{j=1}^{n}\frac{d\mu}{d\nu}(\xi_j)}. \tag{2.2}$$

Faktori koji množe vrijednosti $f(\xi_i)$ nam govore s kojom količinom različiti $f(\xi_i)$ sudjeluju u računu integrala. Stoga ćemo ih nazivati *težinama*. Ovaj procjenitelj zahtjeva poznavanje Radon-Nikodymove derivacije do na konstantu, odnosno za njegov izračun je dovoljno da znamo evaluirati funkciju h tako da vrijedi $h(x) = c\frac{d\mu}{d\nu}(x)$, gdje je $c \in \mathbb{R}$.

Asimptotika osnovnog algoritma

U navedena dva procjenitelja su uzorci, za razliku od ostatka rada, nezavisni i jednako distribuirani. Stoga je dokazivanje direktna primjena zakona velikih brojeva i centralnog graničnog teorema.

Teorem 2.1.1 (Asimptotika procjenitelja uzorkovanja po važnosti)**.** *Neka su μ, ν vjerojatnosne mjere nad prostorom $(\mathsf{X}, \mathcal{X})$ takve da je $\mu \ll \nu$, a $\frac{d\mu}{d\nu}$ pozitivna verzija Radon-Nikodymove derivacije. Neka je $f : \mathsf{X} \to \mathbb{R}$ izmjeriva funkcija takva da vrijedi $f \in \mathcal{L}^1(\mathsf{X}, \mu)$. Neka je $\xi_1, \xi_2, \dots$ niz n.j.d. varijabli distribuiranih po ν.*

Tada vrijedi:

1. *Procjenitelj uzorkovanja po važnosti je jako konzistentan:* $\tilde{\mu}_n^{UV}(f) = \frac{1}{n} \sum_{i=1}^{n} f(\xi_i) \frac{d\mu}{d\nu}(\xi_i) \xrightarrow[n\to\infty]{(g.s.)} \mu(f)$.
2. *Uz dodatni uvjet $\nu(f^2 \frac{d\mu}{d\nu}^2) < \infty$ vrijedi centralni granični teorem*

$$\sqrt{n}(\tilde{\mu}_n^{UV}(f) - \mu(f)) \xrightarrow[n\to\infty]{\mathcal{D}} N(0, \tilde{\sigma}^2(f)),$$

gdje je varijanca $\tilde{\sigma}^2(f)$ jednaka

$$\tilde{\sigma}^2(f) = \int_{\mathsf{X}} (f(x)\frac{d\mu}{d\nu}(x) - \mu(f))^2 \nu(dx).$$

3. *Procjenitelj samo-normalizirajućeg uzorkovanja po važnosti je jako konzistentan:* $\hat{\mu}_n^{UV}(f) = \sum_{i=1}^{n} f(\xi_i) \frac{\frac{d\mu}{d\nu}(\xi_i)}{\sum_{j=1}^{n} \frac{d\mu}{d\nu}(\xi_j)} \xrightarrow[n\to\infty]{(g.s.)} \mu(f)$.
4. *Uz dodatni uvjet $\nu(f^2 \frac{d\mu}{d\nu}^2 + \frac{d\mu}{d\nu}^2) < \infty$ vrijedi centralni granični teorem*

$$\sqrt{n}(\hat{\mu}_n^{UV}(f) - \mu(f)) \xrightarrow[n\to\infty]{\mathcal{D}} N(0, \hat{\sigma}^2(f))$$

gdje je varijanca $\hat{\sigma}^2(f)$ jednaka

$$\hat{\sigma}^2(f) = \int_{\mathsf{X}} \frac{d\mu}{d\nu}(x)^2 (f(x) - \mu(f))^2 \nu(dx).$$

Dokaz.

1. Direktno iz Kolmogorovljevog jakog zakona velikih brojeva (teorem 12.15. u [20]).
2. Direktno iz Lévyevog centralnog graničnog teorema (teorem 14.1. u [20]).
3. Iz Kolmogorovljevog jakog zakona velikih brojeva vrijedi

$$\frac{1}{n} \sum_{i=1}^{n} \frac{d\mu}{d\nu}(\xi_i) \xrightarrow[n\to\infty]{(g.s.)} \nu(\frac{d\mu}{d\nu}) = 1. \tag{2.3}$$

Dijeljenjem rezultata jake konzistentnosti $\frac{1}{n}\sum_{i=1}^{n} f(\xi_i)\frac{d\mu}{d\nu}(\xi_i) \xrightarrow[n\to\infty]{(g.s.)} \mu(f)$ s gornjim izrazom (2.3) dobivamo

$$\hat{\mu}_n^{UV}(f) \xrightarrow[n\to\infty]{(g.s.)} \mu(f).$$

4. Zbog uvjeta $\nu(f^2\frac{d\mu}{d\nu}^2 + \frac{d\mu}{d\nu}^2) < \infty$ možemo primijeniti Lévyev centralni granični teorem na funkciju $f(x)\frac{d\mu}{d\nu}(x) - \mu(f)\frac{d\mu}{d\nu}(x)$ te vrijedi

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f(\xi_i)\frac{d\mu}{d\nu}(\xi_i) - \frac{d\mu}{d\nu}(\xi_i)\mu(f) \xrightarrow[n\to\infty]{\mathcal{D}} N(0, \tilde{\sigma}^2(f)). \tag{2.4}$$

Konvergencija (g.s.) implicira konvergenciju po vjerojatnosti pa iz (2.3) slijedi

$$\frac{1}{n}\sum_{i=1}^{n}\frac{d\mu}{d\nu}(\xi_i) \xrightarrow[n\to\infty]{\mathcal{P}} 1$$

i onda dijeljenjem (2.4) s gornjim izrazom, i korištenjem Slutskyeve leme A.7 dobivamo

$$\sqrt{n}(\hat{\mu}_n^{UV}(f) - \mu(f)) \xrightarrow[n\to\infty]{\mathcal{D}} N(0, \hat{\sigma}^2(f)).$$

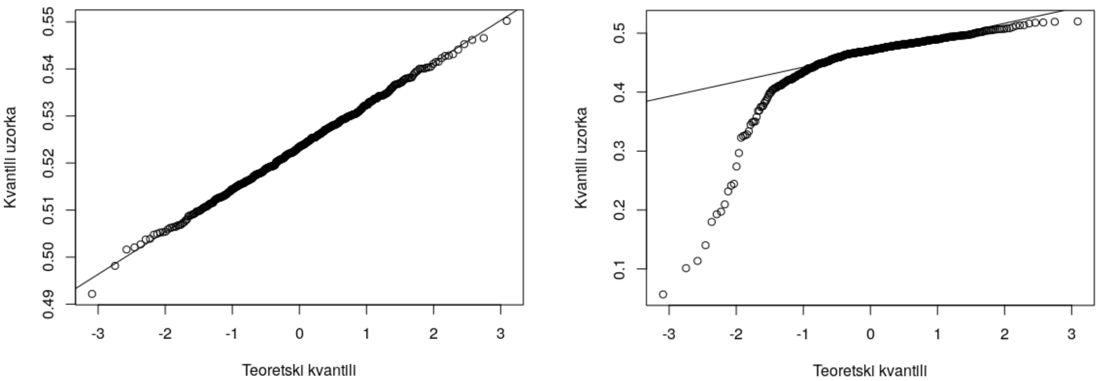
□

Opišimo ukratko korist metode uzorkovanja po važnosti u situaciji *rijetkih događaja*, odnosno kada funkcija koju integriramo poprima velike vrijednosti samo na rijetkim događajima. Uz odabir instrumentalne distribucije ν koja je koncentrirana oko tih rijetkih događaja, uzorkovanjem po važnosti dobivamo procjenitelj koji ima varijancu $\tilde{\sigma}^2(f\frac{d\mu}{d\nu})$. Ta varijanca u nekim slučajevima može biti manja od varijance osnovnog Monte Carlo procjenitelja $Var_\mu(f)$. S druge strane, ako loše odaberemo instrumentalnu distribuciju, uzorkovanje po važnosti može davati loše procjene kao što idući primjer pokazuje.

Primjer 2.1.2 (Uzorkovanje iz Cauchyeve i normalne distribucije)**.** Neka je $\mathcal{C}$ Cauchyeva distribucija nad $\mathbb{R}$ s lokacijom 0 i skalom 1, a $\mathcal{G}$ normalna distribucija $N(0, 1)$. Zato što su obje distribucije apsolutno neprekidne u odnosu na Lebesgueovu mjeru s nosačem jednakim $\mathbb{R}$, one su i međusobno apsolutno neprekidne. Radon-Nikodymova derivacija je razlomak pripadnih funkcija gustoće.

Promotrimo funkciju $f(x) = exp(-|x|)$ čiji integral računamo s obzirom na mjere $\mathcal{C}, \mathcal{G}$, koristeći redom mjere $\mathcal{G}, \mathcal{C}$ kao instrumentalne. Za $x \in \mathbb{R}$ vrijedi

$$\begin{aligned} \frac{d\mathcal{C}}{d\mathcal{G}}(x) &= \sqrt{2\pi}\frac{exp(x^2/2)}{\pi(1+x^2)}, \\ \frac{d\mathcal{G}}{d\mathcal{C}}(x) &= \frac{1}{\sqrt{2\pi}}\frac{\pi(1+x^2)}{exp(x^2/2)}. \end{aligned}$$



(a) procjenitelji $\hat{\mathcal{G}}^{UV}_{1000}(f)$ *(b) procjenitelji* $\hat{C}^{UV}_{1000}(f)$

Slika 2.1: Normalni vjerojatnosni grafovi procjenitelja uzorkovanja po važnosti.

Za upotrebu dijela 4. teorema 2.1.1 potrebno nam je $\nu(f^2 \frac{d\mu}{d\nu}^2 + \frac{d\mu}{d\nu}^2) < \infty$. Ako koristimo $\mathcal{G}$ kao instrumentalnu, to ne vrijedi. Teorem tada ne garantira asimptotsku normalnost i očekujemo loše ponašanje procjenitelja $\hat{C}^{UV}_{1000}(f)$. S druge strane, u slučaju distribucije C kao instrumentalne vrijedi uvjet $\nu(f^2 \frac{d\mu}{d\nu}^2 + \frac{d\mu}{d\nu}^2) < \infty$, te onda očekujemo dobro ponašanje procjenitelja $\hat{\mathcal{G}}^{UV}_{1000}(f)$.

Simulacijski eksperiment s 500 replikacija potvrđuje to ponašanje. Simuliramo procjenitelje $\hat{\mathcal{G}}^{UV}_{1000}(f)$ i $\hat{C}^{UV}_{1000}(f)$ i promatramo njihove normalne vjerojatnosne grafove, slika 2.1.

2.2 Uzorkovanje po važnosti s reuzorkovanjem

U ovom potpoglavlju opisujemo algoritam koji je nastao s ciljem izvlačenja uzoraka iz distribucije iz koje ne znamo to učiniti direktno. Uzorkovanje po važnosti je osnova te metode. Razvijena je u radu [19].

Pretpostavimo da imamo vjerojatnosnu mjeru μ iz koje želimo izvući uzorak duljine n, i vjerojatnosnu mjeru ν tako da vrijedi $\mu \ll \nu$, te nam je funkcija $\frac{d\mu}{d\nu}$ poznata do na množenje konstantom.

Iz distribucije ν izvlačimo uzorak duljine m s pripadnim težinama

$$\omega_i = \frac{\frac{d\mu}{d\nu}(\xi_i)}{\sum_{j=1}^m \frac{d\mu}{d\nu}(\xi_j)}.$$

U samo-normalizirajućem uzorkovanju po važnosti parove (ξ_i, ω_i) koristimo za procjenjivanje integrala neke funkcije po μ.

S druge strane, trik za transformiranje niza ξ_i u uzorak iz μ je da iz niza ξ_i slučajno odaberemo novi niz elemenata pomoću težina ω_i. Novi niz $\tilde{\xi}_1, \tilde{\xi}_2, \ldots, \tilde{\xi}_n$ dobivamo uzorkovanjem s ponavljanjem iz populacije $\xi_1, \xi_2, \ldots, \xi_m$, gdje nam je vjerojatnost odabira i-tog člana jednaka ω_i.

Taj proces nazivamo *reuzorkovanje*, i njime smo dobili niz slučajnih varijabli $\tilde{\xi}_1, \ldots, \tilde{\xi}_n$ koje predstavljaju uzorak iz μ.

Algoritam 2.2.1 (Uzorkovanje po važnosti s reuzorkovanjem)**.**
Inicijalizacija:

Generirajmo nezavisan i jednako distribuiran slučajan uzorak $\xi_1, \xi_2, \ldots, \xi_m$ takav da $\xi_i \sim \nu$. Izračunajmo

$$\omega_i = \frac{\frac{d\mu}{d\nu}(\xi_i)}{\sum_{j=1}^{m} \frac{d\mu}{d\nu}(\xi_j)}. \tag{2.5}$$

Reuzorkovanje:

Generirajmo prirodne brojeve $I_1, I_2, \ldots, I_n$ tako da za svaki par (i, j) gdje je $1 \leq i \leq m$ i $1 \leq j \leq n$ vrijedi

$$P(I_j = i | \xi_1, \xi_2, \ldots, \xi_m) = \omega_i$$

te za svaki j definirajmo

$$\tilde{\xi}_j = \xi_{I_j}.$$

Željeni uzorak jest $(\tilde{\xi}_1, \tilde{\xi}_2, \ldots, \tilde{\xi}_n)$.

Korak reuzorkovanja je zapravo multinomijalno uzorkovanje.

Slučajne varijable $\tilde{\xi}_i$ nisu marginalno distribuirane po μ, ali i dalje možemo koristiti taj uzorak za procjenu integrala nad μ, u smislu koji ćemo pokazati u idućem potpoglavlju. Slučajne varijable $\tilde{\xi}_i$ nisu nezavisne, stoga će analiza procjenitelja koji koriste ovu metodu biti nešto složenija.

Primijetimo da za funkciju f vrijedi

$$\mathbb{E}[f(\tilde{\xi}_i) | \xi_1, \xi_2, \ldots, \xi_m] = \sum_{j=1}^{m} \frac{\frac{d\mu}{d\nu}(\xi_j)}{\sum_{j=1}^{m} \frac{d\mu}{d\nu}(\xi_j)} f(\xi_j) = \hat{\mu}_n^{UV}(f).$$

To pokazuje da je uvjetno očekivanje svakog od $f(\tilde{\xi}_i)$ jednako samo-normalizirajućem procjenitelju uzorkovanja po važnosti. Procjenitelj od $\mu(f)$ korištenjem uzorkovanja po važnosti s reuzorkovanjem jest

$$\mu_n^{UVR}(f) = \frac{1}{n}\sum_{i=1}^{n} f(\tilde{\xi}_i), \tag{2.6}$$

i vrijedi

$$\mathbb{E}[\mu_n^{UVR}(f)|\xi_1, \xi_2, \ldots, \xi_m] = \hat{\mu}_n^{UV}(f).$$

Uz shvaćanje uvjetnog očekivanja kao projekcije na prostor (i dovoljne uvjete da to možemo napraviti) dobivamo

$$\mathbb{E}[(\mu_n^{UVR}(f) - \mu(f))^2] \geq \mathbb{E}[(\hat{\mu}_n^{UV}(f) - \mu(f))^2].$$

Dakle, $\mu_n^{UVR}(f)$ ima veću grešku procjenjivanja od procjenitelja $\hat{\mu}_n^{UV}(f)$. S druge strane, za procjenitelj $\mu_n^{UVR}(f)$ trebamo samo varijable $\tilde{\xi}_1, \tilde{\xi}_2, \ldots, \tilde{\xi}_n$, dok za procjenitelj $\hat{\mu}_n^{UV}(f)$ trebamo varijable $\xi_1, \xi_2, \ldots, \xi_m$ i dodatno njihove pripadne težine $\omega_1, \omega_2, \ldots, \omega_m$.

2.3 Dvostruki nizovi i asimptotska analiza uzorkovanja po važnosti s reuzorkovanjem

Procjenitelj integrala uzorkovanjem po važnosti s reuzorkovanjem $\mu_n^{UVR}(f)$ koristi slučajne varijable $\tilde{\xi}_1, \tilde{\xi}_2, \ldots, \tilde{\xi}_n$ koje nisu međusobno nezavisne. Dosad korišteni zakoni velikih brojeva kao uvjet imaju nezavisnost, stoga ćemo ih morati proširiti, a to radimo pomoću uvjetne nezavisnosti.

Za bavljenje asimptotskom analizom procjenitelja od sada pa nadalje trebat će nam i proširene definicije pojma konzistentnosti i asimptotske normalnosti, a njih uvodimo pomoću dvostrukih nizova.

Definicija 2.3.1 (Dvostruki niz)**.** *Neka je $\{M_n\}_{n\geq 0}$ niz prirodnih brojeva koji ide u beskonačnost u smislu da za svaki pozitivan realan broj r postoji n_r takav da je svaki element niza $\{M_n\}_{n\geq n_r}$ veći od r. Slučajni proces $\{\xi_{n,i}\}_{1\leq i\leq M_n}$, gdje slučajne varijable $\xi_{n,i}$ poprimaju vrijednosti u općenitom izmjerivom prostoru $(\mathsf{X}, \mathcal{X})$, nazivamo dvostrukim nizom. Dvostruki niz s težinama je dvostruki niz $\{(\xi_{n,i}, \omega_{n,i})\}_{1\leq i\leq M_n}$ gdje su $\omega_{n,i}$ nenegativne realne slučajne varijable takve da $\sum_{i=1}^{M_n} \omega_{n,i} > 0$ (g.s.) za svaki n. Ovisno o izgledu dvostrukog niza, implicitno ćemo shvaćati radi li se o dvostrukom nizu s težinama ili bez njih.*

Definicija 2.3.2 (Uvjetna nezavisnost dvostrukog niza)**.** *Neka je $\{\mathcal{F}_n\}_{n\geq 0}$ niz σ-algebri iz $\mathcal{F}$. Kažemo da je dvostruki niz $\{\xi_{n,i}\}_{1\leq iM_n}$ uvjetno nezavisan s obzirom na $\{\mathcal{F}_n\}$ ako vrijedi da je za svaki n skup $\{\xi_{n,1}, \xi_{n,2}, \ldots, \xi_{n,M_n}\}$ uvjetno nezavisan s obzirom na σ-algebru $\mathcal{F}_n$, to jest ako za proizvoljne funkcije $f_1, f_2, \ldots, f_{M_n} \in \mathcal{F}^b(\mathsf{X})$ vrijedi*

$$\mathbb{E}[f_1(\xi_{n,1}) \cdots f_{M_n}(\xi_{n,M_n})|\mathcal{F}_n] = \mathbb{E}[f_1(\xi_{n,1})|\mathcal{F}_n] \cdots \mathbb{E}[f_{M_n}(\xi_{n,M_n})|\mathcal{F}_n] \ (g.s.) \tag{2.7}$$

Dvostruki se nizovi također nazivaju i *trokutastim nizovima* jer ih možemo zamišljati kao

$$\begin{matrix} \xi_{1,1} & \xi_{1,2} \\ \xi_{2,1} & \xi_{2,2} & \xi_{2,3} & \xi_{2,4} \\ \xi_{3,1} & \xi_{3,2} & \xi_{3,3} & \xi_{3,4} & \xi_{3,5} \\ \vdots & \ddots \end{matrix}$$

Izgled ne mora zaista biti trokutast, i M_n ne mora nužno biti rastući niz, ali najčešće to jeste slučaj.

Poopćimo sada pojam konzistentnosti na dvostruke nizove. Prije nego to učinimo, definirajmo na kakvim skupovima funkcija promatramo konzistentnost.

Definicija 2.3.3 (Cjelovit skup funkcija)**.** *Neka je* $(\mathsf{X}, \mathcal{X})$ *izmjeriv prostor, i* C *neki skup realnih izmjerivih funkcija na tom prostoru. Neka za* C *vrijedi*
- *ako su* $\alpha, \beta \in \mathbb{R}$ *te* $f, g \in C$ *tada je i* $\alpha f + \beta g \in C$,
- *ako je* $f \in C$ *te* $g : \mathsf{X} \to \mathbb{R}$ *izmjeriva funkcija takva da je* $|g| \leq |f|$, *tada vrijedi* $g \in C$.

Tada C *nazivamo cjelovitim skupom funkcija.*

Definicija 2.3.4 (Konzistentnost dvostrukog niza)**.** *Neka je* $\{\xi_{n,i}, \omega_{n,i}\}_{1 \leq i \leq M_n}$ *dvostruki niz. Neka je* (ν, C) *uređeni par, gdje je* ν *vjerojatnosna mjera nad* $(\mathsf{X}, \mathcal{X})$, *a* C *cjelovit skup funkcija s vrijednostima u* X *takav da* $C \subseteq \mathcal{L}^1(\mathsf{X}, \nu)$. *Ako za svaku funkciju* $f \in C$ *vrijedi*

$$\sum_{i=1}^{M_n} \frac{\omega_{n,i} f(\xi_{n,i})}{\sum_{j=1}^{M_n} \omega_{n,j}} \xrightarrow[n\to\infty]{\mathcal{P}} \nu(f)$$

tada dvostruki niz $\{\xi_{n,i}, \omega_{n,i}\}_{1 \leq i \leq M_n}$ *nazivamo konzistentnim za par* (ν, C).

Pojam konzistentnosti je koristan, ali ne sadrži nikakvu informaciju o tome koliko povećanje broja uzoraka n povećava točnost procjene. Zato se u ovom radu više posvećujemo dokazivanju svojstva asimptotske normalnosti.

Definicija 2.3.5 (Asimptotska normalnost dvostrukog niza)**.** *Neka je* $\{\xi_{n,i}, \omega_{n,i}\}_{n,i}$ *dvostruki niz. Neka je* $(\nu, A, \sigma, \{\alpha_n\})$ *uređena četvorka takva da je*
- ν *vjerojatnosna mjera na* $(\mathsf{X}, \mathcal{X})$,
- A *cjelovit skup funkcija s vrijednostima u* X, $A \subseteq \mathcal{L}^1(\mathsf{X}, \nu)$,
- σ *nenegativna realna funkcija na* A, *takva da za svaki* $f \in A$ *vrijedi* $\sigma(f) < \infty$,
- $\{\alpha_n\}$ *niz brojeva koji ide u beskonačnost.*

Ako za svaku funkciju $f \in A$ *vrijedi*

$$\alpha_n \sum_{i=1}^{M_n} \frac{\omega_{n,i}}{\sum_{j=1}^{M_n} \omega_{n,j}} (f(\xi_{n,i}) - \nu(f)) \xrightarrow[n\to\infty]{\mathcal{D}} N(0, \sigma^2(f)).$$

tada dvostruki niz $\{\xi_{n,i}, \omega_{n,i}\}$ *nazivamo asimptotski normalnim za* $(\nu, A, \sigma, \{\alpha_n\})$.

Kao primjer asimptotski normalnog dvostrukog niza, u 2.1.1 smo dokazali da za skup $A = \{f \in \mathcal{L}^1(\mathsf{X}, \mu) : \nu(f^2 \frac{d\mu}{d\nu}^2 + \frac{d\mu}{d\nu}^2) < \infty\}$ i funkciju $\hat{\sigma}^2(f) = \int\limits_{\mathsf{X}} \frac{d\mu}{d\nu}(x)^2 (f(x) - \mu(f))^2 \nu(dx)$ vrijedi

$$\sqrt{n}\Big(\sum_{i=1}^{n} \frac{\frac{d\mu}{d\nu}(\xi_i)}{\sum_{j=1}^{n} \frac{d\mu}{d\nu}(\xi_j)} [f(\xi_i) - \mu(f)]\Big) \xrightarrow[n\to\infty]{\mathcal{D}} N(0, \hat{\sigma}^2(f)).$$

što znači da je dvostruki niz $\{\xi_i, \frac{d\mu}{d\nu}(\xi_i)\}_{1\le i\le n}$ asimptotski normalan za $(\mu, A, \hat{\sigma}, \{\sqrt{n}\})$. Cjelovitost skupa A se dokazuje upotrebom Jensenove nejednakosti.

U vidu definicije dvostrukog niza, promotrimo algoritam s reuzorkovanjem. Prvo, imamo dvostruki niz $\{\xi_{m,i}, 1\}_{1\le i\le m}$ gdje su $\xi_{m,i}$ n.j.d iz ν. Zatim tom nizu pridijelimo odgovarajuće težine $\omega_{m,i}$ kao u opisu algoritma (2.5). Pridjeljivanjem težina dobivamo dvostruki niz $\{\xi_{m,i}, \omega_{m,i}\}_{1\le i\le m}$. Na kraju, reuzorkovanjem dobivamo niz $\{\tilde{\xi}_{n,i}, 1\}_{1\le i\le n}$ te želimo proučiti pod kojim uvjetima i za koji skup funkcija je on asimptotski normalan za μ.

Napišimo ponovno algoritam uzorkovanja po važnosti s reuzorkovanjem, pritom pazeći na σ-algebre u odnosu na koje su nizovi uvjetno nezavisni. Zadnja napomena prije samog algoritma jest da ćemo dodatno poopćiti analizu, i počinjati s dvostrukim nizom $\{\xi_{n,i}, \omega_{n,i}\}_{1\le i\le M_n}$, gdje su $\omega_{n,i}$ neke proizvoljne težine.

Algoritam 2.3.6 (Poopćeni algoritam uzorkovanja po važnosti s reuzorkovanjem)**.**
Inicijalizacija:

Neka je $\{\xi_{n,i}, \omega_{n,i}\}_{1\le i\le M_n}$ *dvostruki niz uvjetno nezavisan za* $\mathcal{G}_n$.

Pridjeljivanje težina:

Generiramo drugi dvostruki niz $\{\xi_{n,i}, \omega_{n,i}\frac{d\mu}{d\nu}(\xi_{n,i})\}_{1\le i\le M_n}$ *koji je po konstrukciji uvjetno nezavisan za* $\mathcal{G}_n$.

Reuzorkovanje:

Za svaki n *generirajmo prirodne brojeve* $I_{n,1}, I_{n,2}, \ldots, I_{n,\tilde{M}_n}$ *tako da su uvjetno nezavisni s obzirom na* $\xi_{n,1}, \ldots, \xi_{n,M_n}$ *i tako da za svaki par indeksa* (i, j) *za koji* $1 \le j \le \tilde{M}_n$, $1 \le i \le M_n$ *vrijedi*

$$P(I_{n,j} = i | \xi_{n,1}, \ldots, \xi_{n,M_n}) = \frac{\omega_{n,i}\frac{d\mu}{d\nu}(\xi_{n,i})}{\sum_{l=1}^{M_n} \omega_{n,l}\frac{d\mu}{d\nu}(\xi_{n,l})}$$

te za j *definirajmo*

$$\tilde{\xi}_{n,j} = \xi_{n,I_{n,j}}.$$

Time smo dobili novi dvostruki niz $\{\tilde{\xi}_{n,i}, 1\}_{1 \le i \le \tilde{M}_n}$ *koji je uvjetno nezavisan s obzirom na niz* $\{\mathcal{F}_n\}_{n \ge 0}$ *gdje vrijedi* $\mathcal{F}_n = \sigma(\mathcal{G}_n \cup \sigma(\xi_{n,1}, \ldots, \xi_{n,M_n}))$.

Dokažimo sada dvije leme koje govore o dovoljnim uvjetima za zadržavanje asimptotske normalnosti individualnih koraka gornjeg algoritma. Njih ćemo kasnije spojiti.

Lema 2.3.7 (Asimptotska normalnost koraka dodjeljivanja težina)**.** *Neka su* μ, ν *vjerojatnosne mjere nad* $(\mathsf{X}, \mathcal{X})$ *takve da* $\mu \ll \nu$, *i neka postoji* $\nu - (g.s.)$ *pozitivna verzija Radon-Nikodymove derivacije* $\frac{d\mu}{d\nu}$.

Neka je dvostruki niz $\{\xi_{n,i}, \omega_{n,i}\}_{1 \le i \le M_n}$ *konzistentan za* (ν, C) *te neka je asimptotski normalan za* $(\nu, A, \sigma, \{\alpha_n\})$. *Također, neka za pozitivnu verziju* $\frac{d\mu}{d\nu}$ *vrijedi* $\frac{d\mu}{d\nu} \in C$ *te* $\frac{d\mu}{d\nu} \in A$.

Tada je $\{\xi_{n,i}, \omega_{n,i}\frac{d\mu}{d\nu}(\xi_{n,i})\}_{1 \le i \le M_n}$ *asimptotski normalan za* $(\mu, \hat{A}, \hat{\sigma}, \{\alpha_n\})$, *gdje vrijedi*

$$\hat{A} = \left\{ f \in \mathcal{L}^2(X, \mu) : \frac{d\mu}{d\nu} f \in A \right\},$$
$$\hat{\sigma}(f) = \sigma\left(\frac{d\mu}{d\nu}[f - \mu(f)]\right).$$

Dokaz. Dokaz da je $\hat{A}$ cjelovit slijedi upotrebom Jensenove nejednakosti. Neka je f funkcija u $\hat{A}$. Tada vrijedi $\frac{d\mu}{d\nu} f \in A$. Zbog $\frac{d\mu}{d\nu} \in A$, $\mu(|f|) < \infty$ i cjelovitosti od A imamo da je i $h = \frac{d\mu}{d\nu}(f - \mu(f))$ u A. Zato vrijedi konvergencija

$$\alpha_n \sum_{i=1}^{M_n} \frac{\omega_{n,i} h(\xi_{n,i})}{\sum_{j=1}^{M_n} \omega_{n,j}} = \alpha_n \sum_{i=1}^{M_n} \frac{\omega_{n,i} \frac{d\mu}{d\nu}(\xi_{n,i})}{\sum_{j=1}^{M_n} \omega_{n,j}} (f(\xi_{n,i}) - \mu(f)) \xrightarrow[n\to\infty]{\mathcal{D}} N(0, \sigma^2(h)).$$

Zbog konzistentnosti $\{\xi_{n,i}, \omega_{n,i}\}$ za (ν, C) i $\frac{d\mu}{d\nu} \in C$ možemo gornji izraz podijeliti s lijevom stranom konvergencije $\sum_{i=1}^{M_n} \frac{\omega_{n,i} \frac{d\mu}{d\nu}(\xi_{n,i})}{\sum_{j=1}^{M_n} \omega_{n,j}} \xrightarrow[n\to\infty]{\mathcal{P}} \nu(\frac{d\mu}{d\nu}) = 1$ što sa Slutskyevom lemom A.7 dovršava dokaz propozicije. □

Korak reuzorkovanja će se pojaviti u istom obliku u glavnom algoritmu ovoga rada, sekvencijalnom uzorkovanju s reuzorkovanjem. Stoga nam je sljedeća lema važna i u idućem poglavlju.

Lema 2.3.8 (Asimptotska normalnost koraka reuzorkovanja)**.** *Neka je dvostruki niz* $\{\xi_{n,i}, \omega_{n,i}\}_{1 \le i \le M_n}$ *konzistentan za* (ν, C). *Neka za pozitivnu verziju* $\frac{d\mu}{d\nu}$ *vrijedi* $\frac{d\mu}{d\nu} \in C$ *i* $f^2 \frac{d\mu}{d\nu} \in C$ *te neka su* $\tilde{\xi}_{n,i}$ *definirani kao u opisu algoritma reuzorkovanja. Tada za svaki realan broj* u *vrijedi*

$$\mathbb{E}[exp(\mathrm{i}u\tilde{M}_n^{-\frac{1}{2}} \sum_{i=1}^{\tilde{M}_n} f(\tilde{\xi}_{n,i}) - \mathbb{E}[f(\tilde{\xi}_{n,i})|\mathcal{F}_n])|\mathcal{F}_n] \xrightarrow[n\to\infty]{\mathcal{P}} exp(-u^2 Var_\mu(f)/2).$$

Dokaz. Koristimo teorem A.5 iz dodatka. Onda je dovoljno dokazati svojstva A.5.i., A.5.ii., A.5.iii. tog teorema za $V_{n,i} = \tilde{M}_n^{-\frac{1}{2}} f(\tilde{\xi}_{n,i})$. Po konstrukciji $\tilde{\xi}_{n,i}$ u algoritmu 2.3.6 vrijedi da je dvostruki niz $\{V_{n,i}\}$ uvjetno nezavisan s obzirom na niz σ-algebri $\{\mathcal{F}_n\}$. Raspisom očekivanja multinomijalnog reuzorkovanja slijedi

$$\mathbb{E}[V_{n,i}^2|\mathcal{F}_n] = \frac{1}{\tilde{M}_n}\mathbb{E}[f(\tilde{\xi}_{n,i})^2|\mathcal{F}_n] = \frac{1}{\tilde{M}_n}\sum_{j=1}^{M_n}\frac{\omega_{n,j}\frac{d\mu}{d\nu}(\xi_{n,j})}{\sum_{l=1}^{M_n}\omega_{n,l}\frac{d\mu}{d\nu}(\xi_{n,l})}f(\xi_{n,j})^2 \ (g.s.) \tag{2.8}$$

te imamo

$$\mathbb{E}[V_{n,i}^2|\mathcal{F}_n] - (\mathbb{E}[V_{n,i}|\mathcal{F}_n])^2 = \frac{1}{\tilde{M}_n}(\mathbb{E}[f(\tilde{\xi}_{n,i})^2|\mathcal{F}_n] - (\mathbb{E}[f(\tilde{\xi}_{n,i})|\mathcal{F}_n])^2) =$$

$$= \frac{1}{\tilde{M}_n}(\sum_{j=1}^{M_n}\frac{\omega_{n,j}\frac{d\mu}{d\nu}(\xi_{n,j})}{\sum_{l=1}^{M_n}\omega_{n,l}\frac{d\mu}{d\nu}(\xi_{n,l})}f(\xi_{n,j})^2 - \left(\sum_{j=1}^{M_n}\frac{\omega_{n,j}\frac{d\mu}{d\nu}(\xi_{n,j})}{\sum_{l=1}^{M_n}\omega_{n,l}\frac{d\mu}{d\nu}(\xi_{n,l})}f(\xi_{n,j})\right)^2) \ (g.s.)$$

Primijetimo da zadnji redak ne ovisi o indeksu i kojeg imamo na lijevoj strani. Prva jednakost (2.8) pokazuje svojstvo A.5.i. konačne varijance, jer zdesne strane imamo konačnu sumu funkcija.

Zbog $f^2\frac{d\mu}{d\nu} \in C$ imamo konvergenciju

$$\sum_{i=1}^{M_n}\frac{\omega_{n,i}\frac{d\mu}{d\nu}(\xi_{n,i})}{\sum_{j=1}^{M_n}\omega_{n,j}}f(\xi_{n,i})^2 \xrightarrow[n\to\infty]{\mathcal{P}} \nu(\frac{d\mu}{d\nu}f^2) = \mu(f^2). \tag{2.9}$$

Zato što $\frac{d\mu}{d\nu} \in C$ vrijedi $\sum_{j=1}^{M_n}\frac{\omega_{n,j}\frac{d\mu}{d\nu}(\xi_{n,j})}{\sum_{l=1}^{M_n}\omega_{n,l}} \xrightarrow[n\to\infty]{\mathcal{P}} 1$ što dijeljenjem s (2.9) daje

$$\sum_{i=1}^{M_n}\frac{\omega_{n,i}\frac{d\mu}{d\nu}(\xi_{n,i})}{\sum_{j=1}^{M_n}\omega_{n,j}\frac{d\mu}{d\nu}(\xi_{n,j})}f(\xi_{n,i})^2 \xrightarrow[n\to\infty]{\mathcal{P}} \mu(f^2). \tag{2.10}$$

Zbog nejednakosti $|f|\frac{d\mu}{d\nu} \leq \frac{d\mu}{d\nu} + f^2\frac{d\mu}{d\nu}$, $\frac{d\mu}{d\nu} \in C$, $f^2\frac{d\mu}{d\nu} \in C$ i cjelovitosti od C slijedi da je i $f\frac{d\mu}{d\nu} \in C$. Stoga, analogno s izvodom od (2.10), vrijedi

$$\sum_{i=1}^{M_n}\frac{\omega_{n,i}\frac{d\mu}{d\nu}(\xi_{n,i})}{\sum_{j=1}^{M_n}\omega_{n,j}\frac{d\mu}{d\nu}(\xi_{n,j})}f(\xi_{n,i}) \xrightarrow[n\to\infty]{\mathcal{P}} \mu(f)$$

pa imamo

$$\sum_{i=1}^{\tilde{M}_n}\mathbb{E}[V_{n,i}^2|\mathcal{F}_n] - \mathbb{E}[V_{n,i}|\mathcal{F}_n]^2 =$$

$$\sum_{i=1}^{M_n}\frac{\omega_{n,i}\frac{d\mu}{d\nu}(\xi_{n,i})}{\sum_{j=1}^{M_n}\omega_{n,j}\frac{d\mu}{d\nu}(\xi_{n,j})}f(\xi_{n,i})^2 - \left(\sum_{i=1}^{M_n}\frac{\omega_{n,i}\frac{d\mu}{d\nu}(\xi_{n,i})}{\sum_{j=1}^{M_n}\omega_{n,j}\frac{d\mu}{d\nu}(\xi_{n,j})}f(\xi_{n,i})\right)^2 \xrightarrow[n\to\infty]{\mathcal{P}} \mu(f^2) - \mu(f)^2.$$

što dokazuje svojstvo A.5.ii. konvergencije sume uvjetnih varijanci.

Na kraju, promotrimo $\mathbb{E}[V_{n,i}^2 \mathbb{1}_{\{|V|\geq\epsilon\}}|\mathcal{F}_n]$ za fiksan $\epsilon > 0$. Vrijedi, kao i u izvodu (2.8),

$$\sum_{i=1}^{\tilde{M}_n} \mathbb{E}[V_{n,i}^2 \mathbb{1}_{\{|V|\geq\epsilon\}}|\mathcal{F}_n] = \sum_{i=1}^{M_n} \frac{\omega_{n,i}\frac{d\mu}{d\nu}(\xi_{n,i})}{\sum_{j=1}^{M_n}\omega_{n,j}\frac{d\mu}{d\nu}(\xi_{n,j})} f(\xi_{n,i})^2 \mathbb{1}_{\left\{|f(\xi_{n,i})|\geq\epsilon\sqrt{\tilde{M}_n}\right\}} (g.s.)$$

Neka je K proizvoljan pozitivan realan broj. Za dovoljno velik n vrijedi $\epsilon\sqrt{\tilde{M}_n} \geq K$, pa onda za dovoljno velik n imamo i nejednakost $\mathbb{1}_{\left\{|f(\xi_{n,i})|\geq\epsilon\sqrt{\tilde{M}_n}\right\}} \leq \mathbb{1}_{\{|f(\xi_{n,i})|\geq K\}}$. Zbog cjelovitosti skupa C i $f^2\frac{d\mu}{d\nu} \in C$ vrijedi $f^2\frac{d\mu}{d\nu}\mathbb{1}_{\{|f|\geq K\}} \in C$ i onda

$$\sum_{i=1}^{\tilde{M}_n} \mathbb{E}[V_{n,i}^2 \mathbb{1}_{\{|V|\geq\epsilon\}}|\mathcal{F}_n] \leq \sum_{i=1}^{M_n} \frac{\omega_{n,i}\frac{d\mu}{d\nu}(\xi_{n,i})}{\sum_{j=1}^{M_n}\omega_{n,j}\frac{d\mu}{d\nu}(\xi_{n,j})} f(\xi_{n,i})^2 \mathbb{1}_{\{|f(\xi_{n,i})|\geq K\}} \xrightarrow[n\to\infty]{\mathcal{P}} \mu(f^2\mathbb{1}_{\{|f|\geq K\}}).$$

Zato što je K proizvoljan i jer $\mu(f^2\mathbb{1}_{\{|f|\geq K\}})$ postaje proizvoljno malen kad $K \to \infty$, upotrebom leme A.2 dobivamo i svojstvo A.5.iii., da za svaki $\epsilon > 0$ vrijedi

$$\sum_{i=1}^{\tilde{M}_n} \mathbb{E}[V_{n,i}^2 \mathbb{1}_{\{|V|\geq\epsilon\}}|\mathcal{F}_n] \xrightarrow[n\to\infty]{\mathcal{P}} 0$$

što dovršava dokaz leme. □

Teorem 2.3.9. *Neka vrijede uvjeti leme 2.3.7 i leme 2.3.8. Također neka $\alpha_n^2/\tilde{M}_n$ ima limes u $[0,\infty]$ koji označavamo sa α. Definirajmo $\tilde{A}$ sa*

$$\tilde{A} = \left\{ f \in \mathcal{L}^2(\mathsf{X},\mu) : \frac{d\mu}{d\nu}f \in A, \frac{d\mu}{d\nu}f^2 \in C \right\}.$$

Tada je dvostruki niz $\{\tilde{\xi}_{n,i}, 1\}_{1\leq i\leq\tilde{M}_n}$ asimptotski normalan za

- $(\mu, \tilde{A}, \tilde{\sigma}(f), \{\alpha_n\})$ *u slučaju $\alpha < 1$, gdje vrijedi*

$$\tilde{\sigma}(f) = \hat{\sigma}(f) + \alpha Var_\mu(f).$$

- $(\mu, \tilde{A}, \tilde{\sigma}(f), \{\tilde{M}_n^{1/2}\})$ *u slučaju $\alpha \geq 1$, gdje vrijedi*

$$\tilde{\sigma}(f) = \alpha^{-1}\hat{\sigma}(f) + Var_\mu(f).$$

Dokaz. Zapišimo $\mu_n^{UVR}(f) - \mu(f) = A_n + B_n$ gdje smo uzeli oznake

$$A_n = \frac{1}{\tilde{M}_n}\sum_{i=1}^{\tilde{M}_n}\mathbb{E}[f(\tilde{\xi}_{n,i})|\mathcal{F}_n] - \mu(f) = \sum_{i=1}^{M_n} \frac{\omega_{n,i}\frac{d\mu}{d\nu}(\xi_{n,i})}{\sum_{j=1}^{M_n}\omega_{n,j}\frac{d\mu}{d\nu}(\xi_{n,j})} f(\xi_{n,i}) - \mu(f),$$

$$B_n = \frac{1}{\tilde{M}_n}\sum_{i=1}^{\tilde{M}_n} f(\tilde{\xi}_{n,i}) - \mathbb{E}[f(\tilde{\xi}_{n,i})|\mathcal{F}_n].$$

Prisjetimo se

$$\hat{\sigma}^2(f) = \sigma^2(\frac{d\mu}{d\nu}[f - \mu(f)]),$$
$$Var_\mu(f) = \int_{\mathsf{X}} (f(x) - \mu(f))^2 \mu(dx).$$

Lema 2.3.7 dokazuje da $\alpha_n A_n \xrightarrow[n\to\infty]{\mathcal{D}} N(0, \hat{\sigma}^2(f))$, dok lema 2.3.8 za svaki $\nu \in \mathbb{R}$ daje $\mathbb{E}[exp(iv\tilde{M}_n^{1/2} B_n)|\mathcal{F}_n] \xrightarrow[n\to\infty]{\mathcal{P}} \exp(-\nu^2 Var_\mu(f)/2)$. Zato što je $\alpha_n A_n$ $\mathcal{F}_n$-izmjeriv za svaki n, možemo primijeniti lemu A.6 iz koje dobivamo

$$\begin{pmatrix} \alpha_n A_n \\ \tilde{M}_n^{1/2} B_n \end{pmatrix} \xrightarrow[n\to\infty]{\mathcal{D}} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \hat{\sigma}^2(f) & 0 \\ 0 & Var_\mu(f) \end{pmatrix}\right).$$

Uz shvaćanje karakteristične funkcije $\varphi_{\alpha_n A_n, \tilde{M}_n^{1/2} B_n}(cK_1, cK_2)$ kao funkcije po c, dobivamo da je i svaka linearna kombinacija varijabli $K_1 \alpha_n A_n + K_2 \tilde{M}_n^{1/2} B_n$ asimptotski normalna. Sada zapisom $b_n(A_n + B_n) = b_n \alpha_n^{-1} \alpha_n A_n + b_n \tilde{M}_n^{-1/2} \tilde{M}_n^{1/2} B_n$ dobivamo da je dvostruki niz $\{\tilde{\xi}_{n,i}, 1\}$ asimptotski normalan za $(\mu, \tilde{A}, \tilde{\sigma}, \{b_n\})$, gdje je

$$\tilde{\sigma}^2(f) = (lim_n b_n \alpha_n^{-1})^2 \hat{\sigma}^2(f) + (lim_n b_n \tilde{M}_n^{-1/2})^2 Var_\mu(f).$$

U slučaju $\alpha < 1$ odabiremo $b_n = \alpha_n$, a ako $\alpha \leq 1$ odabiremo $b_n = \tilde{M}_n^{1/2}$ iz čega slijedi tvrdnja teorema. □

Primijenimo sada gornji teorem na algoritam s reuzorkovanjem. Tvrdnja je direktna kombinacija asimptotske normalnosti i konzistentnosti običnog Monte Carlo uzorkovanja s prijašnjim teoremom, stoga ne navodimo dokaz.

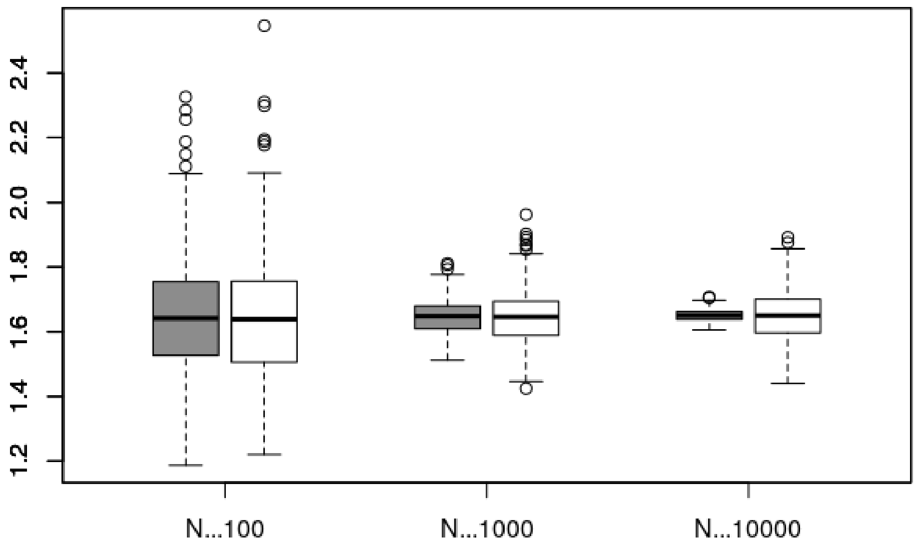
Propozicija 2.3.10 (Uzorkovanje po važnosti s reuzorkovanjem čuva asimptotsku normalnost)**.** *Neka su μ, ν vjerojatnosne mjere na $(\mathsf{X}, \mathcal{X})$ tako da $\mu \ll \nu$. Neka postoji strogo pozitivna verzija Radon-Nikodymove derivacije $\frac{d\mu}{d\nu}$. Neka je $\frac{d\mu}{d\nu} \in \mathcal{L}^2(\mathsf{X}, \mu)$ te $\tilde{A} = \{f \in \mathcal{L}^2(\mathsf{X}, \mu) : f\frac{d\mu}{d\nu} \in\}$. Neka su $\{\xi_{n,i}\}_{1\leq i \leq M_n}$ n.j.d. varijable distribuirane po ν. Neka $M_n/\tilde{M}_n$ ima limes u $[0, \infty]$ koji označavamo sa α Tada je dvostruki niz $\{\tilde{\xi}_{n,i}, 1\}_{1\leq i\leq \tilde{M}_n}$, dan u algoritmu 2.3.6, asimptotski normalan za*

- $(\mu, \tilde{A}, \tilde{\sigma}(f), \{M_n^{1/2}\})$ *u slučaju $\alpha < 1$, gdje za $f \in \tilde{A}$ vrijedi*

$$\tilde{\sigma}(f) = Var_\nu(\frac{d\mu}{d\nu}[f - \mu(f)]) + \alpha Var_\mu(f).$$

- $(\mu, \tilde{A}, \tilde{\sigma}(f), \{\tilde{M}_n^{1/2}\})$ *u slučaju $\alpha \geq 1$, gdje za $f \in \tilde{A}$ vrijedi*

$$\tilde{\sigma}(f) = \alpha^{-1} Var_\nu(\frac{d\mu}{d\nu}[f - \mu(f)]) + Var_\mu(f).$$



Slika 2.2: Tamno su procjenitelji $\hat{\mathcal{G}}_n^{UV}(f)$, a svijetlo su procjenitelji $\mathcal{G}_n^{UVR}(f)$ s fiksnom veličinom krajnjeg uzorka 1000*. Rezultati su dobiveni replikacijom eksperimenta iz primjera 2.3.11* 500 *puta.*

Primjer 2.3.11 (O odnosu veličine instrumentalnog i konačnog uzorka)**.** Nastavljamo primjer 2.1.2 u kojemu uzorkujemo iz Cauchyeve distribucije C kako bismo izračunali integral od $\exp(x)$ pod mjerom $\mathcal{G}$. Prema prethodnoj propoziciji 2.3.10, granična varijanca procjenitelja $\mu_n^{UVR}(f)$ je suma varijance procjenitelja uzorkovanja po važnosti $Var_\nu(\frac{d\mu}{d\nu}[f-\mu(f)])$ i varijance $Var_\mu(f)$ uzrokovane korakom reuzorkovanja. Ovisno o omjeru veličine instrumentalnog uzorka M_n i krajnjeg uzorka $\tilde{M}_n$ ti dijelovi su različitom količinom prisutni u varijanci od $\mu_n^{UVR}(f)$, a ta količina ovisi o parametru $\alpha := lim_n M_n/\tilde{M}_n$.

Fiksirajmo sada veličinu krajnjeg uzorka $\tilde{M} = 1000$ i provedimo eksperiment u kojemu računamo $\mathcal{G}_n^{UVR}(f)$ i $\hat{\mathcal{G}}_n^{UV}(f)$, s instrumentalnom distribucijom C. Na slici 2.2 primjećujemo da kada je instrumentalni uzorak malen u odnosu na krajnji ($\alpha < 1$), tada većinu varijance u $\mathcal{G}_n^{UVR}(f)$ čini varijanca od $\hat{\mathcal{G}}_n^{UV}(f)$. S druge strane, kako se veličina instrumentalnog uzorka povećava, varijanca od $\hat{\mathcal{G}}_n^{UV}(f)$ postaje sve manja, ali varijanca od $\mathcal{G}_n^{UVR}(f)$ postiže neki minimum i na njemu ostaje, a taj minimum je varijanca uzrokovana korakom reuzorkovanja.

Poglavlje 3

Sekvencijalne Monte Carlo metode

U ovom poglavlju primijenit ćemo algoritme inspirirane uzorkovanjem po važnosti za račun izglađivanja $\mathbb{E}[f(X_{l:k})|Y_{0:n}]$. Prvo proširujemo uzorkovanje po važnosti na sekvencijalno uzorkovanje po važnosti, a zatim pokazujemo da je dodatak reuzorkovanja bitna komponenta u radu s duljim nizovima. Za sekvencijalno uzorkovanje po važnosti s reuzorkovanjem dajemo dokaz asimptotske normalnosti u slučaju problema filtriranja.

3.1 Sekvencijalno uzorkovanje po važnosti

Metode koje slijede poznate su pod nazivom *sekvencijalne Monte Carlo metode* i primjenjive su u općenitijim situacijama u kojima uvjetna vjerojatnost $p(x_{0:n}|y_{0:n})$ zadovoljava neku rekurzivnu strukturu. Mi te metode opisujemo sa stajališta skrivenih Markovljevih modela navodeći kakvu rekurzivnu strukturu zadovoljava izglađivanje $\phi_{0:n|n}$. Radimo s parcijalno dominiranim skrivenim Markovljevim modelima iz definicije 1.3.3, to jest onima u kojima jezgra prijelaza iz skrivenog u opaženo stanje ima svojevrsnu funkciju gustoće koju nazivamo prijelazna funkcija gustoće.

U 1.3.7 smo dokazali da za $\phi_{0:k|k}$ vrijedi formula

$$\phi_{0:k|k}(f) = \frac{1}{L_k} \int \cdots \int_{\mathsf{X}^{k+1}} f(x_{0:k}) \nu(dx_0) g_0(x_0) \prod_{i=1}^{k} Q(x_{i-1}, dx_i) g_i(x_i). \tag{3.1}$$

Uzimajući $k = 0$ i uvrštavajući izraz (1.4) za L_0 dobivamo bazu rekurzije

$$\phi_0(f) = \frac{\int_{\mathsf{X}} f(x_0)\nu(dx_0)g_0(x_0)}{\int_{\mathsf{X}} \nu(dx_0)g_0(x_0)},$$

a korištenjem jednakosti (3.1) za $\phi_{0:k|k}$ i $\phi_{0:k+1|k+1}$ dobivamo rekurziju

$$\phi_{0:k+1|k+1}(f) = \frac{L_k}{L_{k+1}} \underset{\mathsf{X}^{k+2}}{\int \cdots \int} f(x_{0:k+1}) \phi_{0:k|k}(dx_{0:k}) Q(x_k, dx_{k+1}) g_{k+1}(x_{k+1}). \tag{3.2}$$

Cilj nam je sada pomoću te rekurzije dobiti instrumentalnu distribuciju za $\phi_{0:n|n}$. Ako označimo nenormaliziranu jezgru prijelaza, prisutnu u (3.2), s

$$T_k^u(x, f) = \int_{\mathsf{X}} f(x') Q(x, dx') g_{k+1}(x') \tag{3.3}$$

izraz rekurzije tada poprima idući oblik

$$\phi_{0:k+1|k+1}(f) = \frac{L_k}{L_{k+1}} \underset{\mathsf{X}^{k+2}}{\int \cdots \int} f(x_{0:k+1}) \phi_{0:k|k}(dx_{0:k}) T_k^u(x_k, dx_{k+1}). \tag{3.4}$$

Normaliziranu (Markovljevu) jezgru $T_k(x, f) = \frac{T_k^u(x,f)}{T_k^u(x,\mathbb{1})}$ ćemo kasnije posebno promatrati.

Osim u specifičnim slučajevima skrivenih Markovljevih modela, onih s konačnim brojem stanja i linearnih gaussovskih, vrijednosti L_i općenito nije moguće analitički izračunati. Zato izraz (3.4) ne možemo iskoristiti za direktno računanje. S druge strane, imamo prijelaz iz distribucije $\phi_{0:k|k}$ u distribuciju $\phi_{0:k+1|k+1}$ u nekom smislu, te bi nam uzorkovanje po važnosti, a posebno samo-normalizirajuća verzija, moglo pomoći. Promotrimo što bi onda bila instrumentalna distribucija.

Neka je R_k Markovljeva prijelazna jezgra takva da je za svaki x mjera $T_k^u(x, \cdot)$ apsolutno neprekidna u odnosu na $R_k(x, \cdot)$. Tada pomoću rekurzije (3.4) i instrumentalne distribucije za $\phi_{0:k|k}$, možemo dobiti i instrumentalnu distribuciju za $\phi_{0:k+1|k+1}$. Ako imamo niz takvih prijelaznih jezgri R_k, one s nekom inicijalnom distribucijom ρ_0 na $(\mathsf{X}, \mathcal{X})$, čine nehomogen Markovljev lanac. Taj lanac nazivamo *instrumentalni lanac*, a prijelazne jezgre R_k nazivamo *instrumentalne jezgre*. Instrumentalni lanac ima mjeru (analogno propoziciji 1.2.7) na prvih n stanja oblika

$$\rho_n(f) = \underset{\mathsf{X}^{n+1}}{\int \cdots \int} f(x_{0:n}) \rho_0(dx_0) \prod_{i=0}^{n-1} R_i(x_i, dx_{i+1}) \tag{3.5}$$

i iduća propozicija pokazuje da je u opisanom slučaju $\phi_{0:n|n}$ apsolutno neprekidna u odnosu na ρ_n.

Propozicija 3.1.1. *Neka je $\{X_n, Y_n\}_{n\geq 0}$ parcijalno dominiran skriven Markovljev proces s vrijednostima u izmjerivom prostoru $(\mathsf{X} \times \mathsf{Y}, \mathcal{X} \otimes \mathcal{Y})$ kao u 1.3.3, gdje je ν inicijalna distribucija od X_0, a Q pripadna Markovljeva prijelazna jezgra. Također, neka je $\{R_k\}_{k\geq 0}$ niz Markovljevih prijelaznih jezgri na X, a ρ_0 distribucija nad X, tako da vrijedi*

- $\nu \ll \rho_0$,

- *za svaki indeks k i svaki $x \in \mathsf{X}$, za mjeru $T_k^u(x,\cdot)$ definiranu kao u* (3.3) *vrijedi* $T_k^u(x,\cdot) \ll R_k(x,\cdot)$.

Tada za svaki indeks k vrijedi da je $\phi_{0:k|k} \ll \rho_k$, gdje je ρ_k vjerojatnosna mjera definirana kao u (3.5). *Za verzije Radon-Nikodymove derivacije tada vrijedi*

$$\frac{d\phi_{0:k|k}}{d\rho_k}(x_{0:k}) = \frac{1}{L_k} g_0(x_0) \frac{d\nu}{d\rho_0}(x_0) \frac{dT_0^u(x_0,\cdot)}{dR_0(x_0,\cdot)}(x_1) \cdots \frac{dT_{k-1}^u(x_{k-1},\cdot)}{dR_{k-1}(x_{k-1},\cdot)}(x_k) \; (g.s.) \tag{3.6}$$

Dokaz. Tvrdnju dokazujemo indukcijom. Iz

$$\phi_0(f) = \frac{1}{L_0} \int_{\mathsf{X}} f(x_0) \nu(dx_0) g_0(x_0)$$

se jasno vidi da vrijedi

$$\frac{d\phi_0}{d\rho_0}(x) = \frac{1}{L_0} g_0(x) \frac{d\nu}{d\rho_0}(x) \tag{3.7}$$

jer za inicijalnu distribuciju ν vrijedi $\nu \ll \rho_0$. To dokazuje bazu indukcije.

Za dokazivanje koraka indukcije, u jednadžbu (3.4) uvrstimo pretpostavku indukcije $\phi_{0:k|k} \ll \rho_k$, i dobivamo

$$\phi_{0:k+1|k+1}(f) = \frac{L_k}{L_{k+1}} \underbrace{\int \cdots \int}_{\mathsf{X}^{k+2}} f(x_{0:k+1}) \frac{d\phi_{0:k|k}}{d\rho_k}(x_{0:k}) \rho_k(dx_{0:k}) T_k^u(x_k, dx_{k+1}).$$

Ako u taj izraz uvrstimo $T_k^u(x_k,\cdot) \ll R_k(x_k,\cdot)$ dobivamo

$$\phi_{0:k+1|k+1}(f) = \frac{L_k}{L_{k+1}} \underbrace{\int \cdots \int}_{\mathsf{X}^{k+2}} f(x_{0:k+1}) \frac{d\phi_{0:k|k}}{d\rho_k}(x_{0:k}) \rho_k(dx_{0:k}) R_k(x_k, dx_{k+1}) \frac{dT_k^u(x_k,\cdot)}{dR_k(x_k,\cdot)}(x_{k+1})$$

što primjećujući da je $\rho_k(dx_{0:k}) R_k(x_k, dx_{k+1})$ zapravo mjera ρ_{k+1} i uvrštavajući $\frac{1}{L_k}$ iz (3.6) dovršava dokaz koraka indukcije. □

Već spomenuta nemogućnost analitičkog izračuna L_i-eva prisutnih u (3.6) nas odmah navodi na korištenje samo-normalizirajuće verzije uzorkovanja po važnosti u kojoj možemo ignorirati konstante.

Težine u algoritmu uzorkovanja po važnosti su vrijednosti Radon-Nikodymove derivacije, te za njihov izračun koristimo jednakost (3.6). Primijetimo prisutnost sekvencijalne strukture - za izračun težine $\frac{d\phi_{0:k+1|k+1}}{d\rho_{k+1}}(\xi_{0:k+1})$ vektora $\xi_{0:k+1}$ distribuiranog po ρ_{k+1}, dovoljno je pomnožiti težinu $\frac{d\phi_{0:k|k}}{d\rho_k}(\xi_{0:k+1})$ vektora $\xi_{0:k}$ distribuiranog po ρ_k s vrijednošću $\frac{dT_k^u(\xi_k,\cdot)}{dR_k(\xi_k,\cdot)}(\xi_{k+1})$, koju nazivamo *inkrementalna težina*. U uzorkovanju po važnosti takvih vektora (varijabli) imamo M, i svaki od njih u ovom sekvencijalnom kontekstu nazivamo *čestica*. Niz vrijednosti koje je jedna čestica poprimila nazivamo njenom *putanjom*.

Ako za svaku česticu produljimo putanju i izračunamo iduću težinu pomoću inkrementalnih težina, i to ponavljamo redom za dijelove instrumentalnog lanca $\rho_0, R_0, R_1, \dots,$ dobivamo algoritam *sekvencijalnog uzorkovanja po važnosti*. Broj skrivenih stanja koje procjenjujemo, odnosno broj opservacija koje imamo, neka bude $S+1$. Stoga će indeks stanja s ići od 0 do S.

Algoritam 3.1.2 (Sekvencijalno uzorkovanje po važnosti)**.**
Inicijalizacija:

Generirajmo nezavisan i jednako distribuiran slučajan uzorak $\xi_1^0, \xi_2^0, \dots, \xi_M^0$ *tako da* $\xi_i^0 \sim \rho_0$. *Za svaki i takav da* $1 \le i \le M$ *izračunajmo inicijalne težine*

$$\omega_i^0 := g_0(\xi_i^0)\frac{d\nu}{d\rho_0}(\xi_i^0). \tag{3.8}$$

Mutacijski korak:

Za svaki vremenski indeks s koji redom ide od 1 *do* S,
nezavisno za svaki i takav da $1 \le i \le M$ *uzorkujemo* $\xi_i^s \sim R_{s-1}(\xi_i^{s-1}, \cdot)$,
računamo novu težinu i-te čestice

$$\omega_i^s := \omega_i^{s-1}\frac{dT_{s-1}^u(\xi_i^{s-1}, \cdot)}{dR_{s-1}(\xi_i^{s-1}, \cdot)}(\xi_i^s) = \omega_i^{s-1} g_s(\xi_i^s)\frac{dQ(\xi_i^{s-1}, \cdot)}{dR_{s-1}(\xi_i^{s-1}, \cdot)}(\xi_i^s). \tag{3.9}$$

i produljujemo putanju i-te čestice sa $\xi_i^{0:s} := (\xi_i^{0:s-1}, \xi_i^s)$.

Evaluacija:

Za funkciju $f \in \mathcal{F}^b(\mathsf{X}^{k+1})$ *nad prvih* $k+1$ *stanja skrivenog lanca, gdje je* $k \le S$ *procjenjujemo*

$$\phi_{0:k|k}(f) \approx \hat{\phi}_{0:k|k}^{SUV}(f) = \sum_{i=1}^{M} \frac{\omega_i^k f(\xi_i^{0:k})}{\sum_{l=1}^{M} \omega_l^k}.$$

Izbor instrumentalnih jezgri

Za potpunu specifikaciju algoritma 3.1.2 sekvencijalnog uzorkovanja po važnosti još trebamo odrediti kako izabrati instrumentalne jezgre R_k i inicijalnu instrumentalnu distribuciju ρ_0 koje zajedno definiraju instrumentalni lanac.

Izbor instrumentalne inicijalne distribucije ρ_0 često ćemo napraviti prema potrebi od primjene do primjene, dijelom jer odabir te distribucije možemo napraviti na mnogo različitih načina, a dijelom jer uz svojstva zaboravljanja instrumentalnog lanca, različiti odabiri inicijalnih distribucija rezultiraju sličnim ponašanjem lanca.

Promotrimo ponovno definiciju nenormalizirane prijelazne jezgre T_k^u,

$$T_k^u(x, f) = \int_{\mathsf{X}} f(x')Q(x, dx')g_{k+1}(x'). \tag{3.10}$$

Želimo niz prijelaznih jezgri $\{R_k\}$ tako da za svaki k i $x \in \mathsf{X}$ vrijedi $T_u^k(x, \cdot) \ll R_k(x, \cdot)$. Jedan izbor instrumentalne jezgre, koji direktno sugerira definicija od T_k^u, je $R_k = Q$ za svaki k. Taj izbor nazivamo *apriorna jezgra*.

Apriorna jezgra

Neka je instrumentalna jezgra R_k jednaka Q. Instrumentalni lanac tada slijedi isti proces kao i skriveni lanac, i za svaki $x \in \mathsf{X}$ vrijedi $R_k(x, \cdot) \sim Q(x, \cdot)$. Za inkrementalnu težinu tada vrijedi

$$\frac{dT_k^u(\xi^k, \cdot)}{dR_k(\xi^k, \cdot)}(\xi^k) = \frac{dT_k^u(\xi^k, \cdot)}{dQ(\xi^k, \cdot)}(\xi^{k+1}) = g_{k+1}(\xi^{k+1})$$

što znači da vrijednost težine ne ovisi o prošloj vrijednosti putanje ξ^k, nego samo o novoj vrijednosti ξ^{k+1}. S obzirom na način upotrebe skrivenih Markovljevih modela, jezgra Q i funkcije g_k često će nam biti poznati. Stoga će implementacija algoritma s apriornom jezgrom biti neposredna i jednostavna. S druge strane, primjeri 3.1.3 i 3.1.4 nam pokazuju kako ta jednostavnost ima svoju cijenu, pa ipak ponekad želimo drukčiji odabir instrumentalne jezgre.

Svi primjeri u ostatku teksta računaju filtraciju u modelima gdje skrivena stanja poprimaju vrijednosti u $\mathbb{R}$. Uvodimo naziv za funkciju identitete, $f_{id} : \mathbb{R} \to \mathbb{R}$ takvu da $f_{id}(x) = x$ za svaki $x \in \mathbb{R}$, jer nam je ona najjednostavnija za korištenje u takvim primjerima. Idući primjer je replikacija rezultata iz potpoglavlja 7.2 u [3].

Primjer 3.1.3 (Autoregresijski model, apriorna jezgra)**.** Proučimo ponašanje algoritma sekvencijalnog uzorkovanja po važnosti s apriornom jezgrom na jednodimenzionalnom autoregresijskom modelu sa šumom. Za taj model neka vrijedi

$$\begin{aligned} X_{k+1} &= \phi X_k + \sigma_U U_k, \\ Y_{k+1} &= X_k + V_k, \end{aligned} \tag{3.11}$$

gdje su $\{U_k\}_{k\geq 0}, \{V_k\}_{k\geq 0}$ nezavisne varijable jednako distribuirane po $N(0, 1)$, a X_0 ima distribuciju koja je stacionarna za skriveni lanac, $N(0, \sigma_U^2/(1 - \phi^2))$. Konstanta autokorelacije ϕ neka je 0.9, a konstanta skrivenog šuma σ_U neka je 0.1.

Promotrimo niz opservacija

$$y_{0:5} = [-0.652, -0.345, -0.676, 1.142, 0.721, 20]$$

gdje je prvih 5 vrijednosti dobiveno simulacijom, a zadnja je umjetno postavljena na netipično veliku vrijednost od 20.

Model definiran sa (3.11) spada u linearan gaussovski, stoga možemo egzaktno izračunati pripadno izglađivanje. Zbog svojstva linearnih gaussovskih modela, uvjetovanje također daje varijable s normalnom distribucijom koje možemo u potpunosti opisati s očekivanjem i varijancom.

Korištenjem Kalmanovog filtra (poglavlje 5 u [3]) dobivamo da za filtriranje ϕ_5 vrijedi $\phi_5(f_{id}) = \mathbb{E}[X_5|Y_{0:5} = y_{0:5}] = 0.907$. S tom egzaktnom vrijednošću uspoređujemo procjene koje daje sekvencijalno uzorkovanje po važnosti s apriornom jezgrom.

Na slici 3.1 možemo primijetiti kako procjene algoritma podcjenjuju vrijednost $\phi_5(f_{id})$. Uzorkovanjem pomoću apriorne jezgre nikada ne uzimamo u obzir vrijednosti opservacija. Zato često vrijednosti čestica nisu blizu onih vrijednosti stanja koja bi s velikom vjerojatnošću proizvela opservacije slične našim podacima.

Dakle, algoritam s apriornom jezgrom ignorira značajnost opservacija poput $y_5 = 20$.

Sve to nas navodi na proučavanje jezgri koje uzimaju u obzir opservacije tijekom uzorkovanja.

Optimalna jezgra

Promotrimo rekurziju za izglađivanje (3.4)

$$\phi_{0:k+1|k+1}(f) = \frac{L_k}{L_{k+1}} \int \cdots \int_{\mathsf{X}^{k+2}} f(x_{0:k+1}) \phi_{0:k|k}(dx_{0:k}) T_k^u(x_k, dx_{k+1}).$$

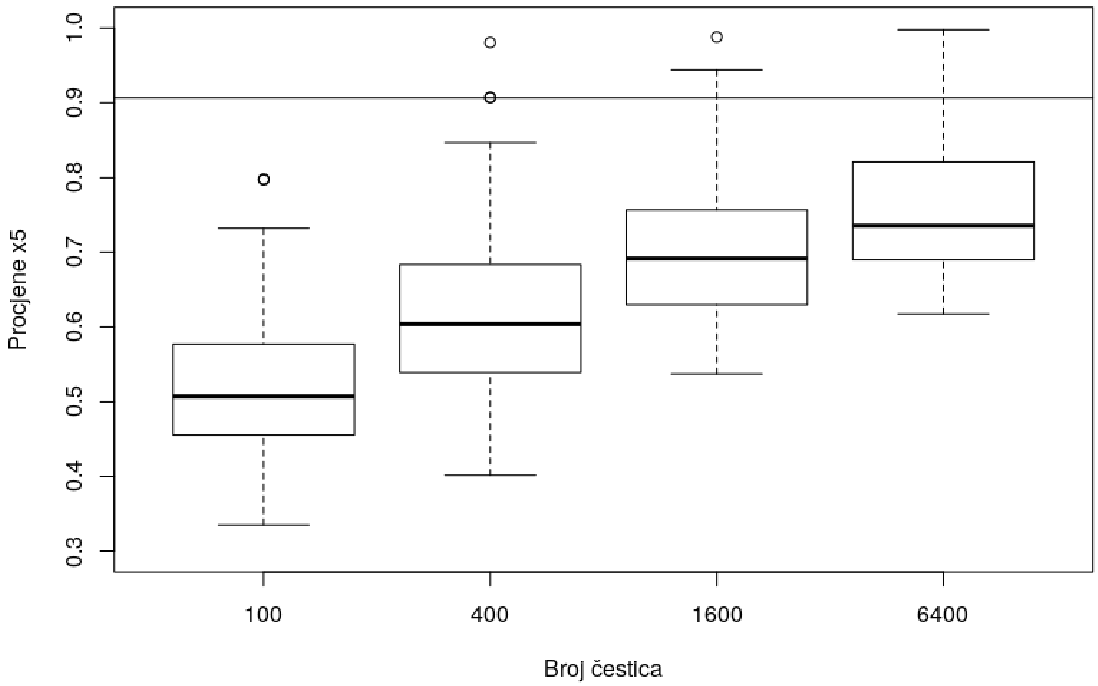
Kada bi za svaki x_k, $T_k^u(x_k, \cdot)$ bila vjerojatnosna distribucija iz koje znamo uzorkovati, putanju neke čestice $\xi^{0:k}$ bismo direktno mogli produžiti uzorkovanjem iz $T_k^u(\xi^k, \cdot)$ bez mijenjanja pripadne težine. To nije nužno istina, ali daje cilj kojemu možemo težiti. Taj cilj je distribucija koju dobijemo normaliziranjem $T_k(x_k, \cdot) = \frac{T_k^u(x_k, \cdot)}{T_k^u(x_k, \mathbb{1})}$, i nazivamo ju *optimalna jezgra*.

Postavljanjem $R_k = T_k$ vrijedi

$$\frac{dT_k^u(\xi^k, \cdot)}{dR_k(\xi^k, \cdot)}(\xi^{k+1}) = \frac{dT_k^u(\xi^k, \cdot)}{dT_k(\xi^k, \cdot)}(\xi^{k+1}) = T_k^u(\xi^k, \mathbb{1}) = \int_{\mathsf{X}} Q(\xi^k, dx_{k+1}) g_{k+1}(x_{k+1})$$

što nam daje potpuno drukčiju situaciju nego s apriornom jezgrom jer sada inkrementalna težina u potpunosti ovisi o prijašnjoj vrijednosti ξ^k putanje jedne čestice. U apriornoj jezgri je inkrementalna težina ovisila samo o novoj vrijednosti ξ^{k+1} putanje.

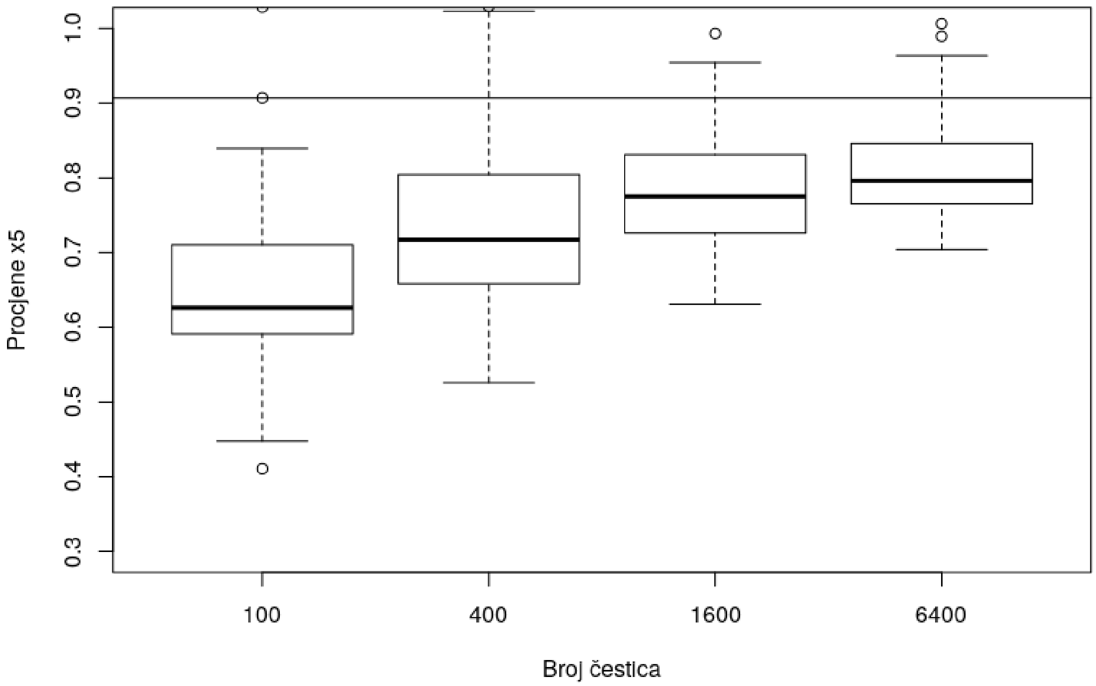
Uvjetno na dosadašnju putanju čestice $\xi^{0:k}$, inkrementalna težina je konstantna, odnosno varijanca inkrementalne težine je 0. Optimalna jezgra je nazvana optimalnom upravo zato što minimizira varijancu težina.



Slika 3.1: Procjene skrivene vrijednosti X_5 modela (3.11) *s opservacijama $y_{0:5}$ pomoću algoritma sekvencijalnog uzorkovanja po važnosti korištenjem **apriorne jezgre**. Svaki boxplot je dobiven sa 125 ponavljanja algoritma za procjenu, i navedenim brojem čestica. Horizontalna crta prikazuje točnu vrijednost* 0.907.

Implementacija ovog slučaja nije jednostavna kao u slučaju apriorne jezgre. Prvi problem vezan je uz uzorkovanje iz T_k, koje nije nužno moguće direktno, a drugi problem je izračun inkrementalne težine $T_k^u(\xi^k, \mathbb{1})$ koji nije uvijek moguć analitički. U slučajevima modela konačnoga broja stanja i linearnih gaussovskih modela nema takvih problema (kao na primjer u primjeru 3.1.4). S druge strane, za sve ostale slučajeve postoji niz potencijalnih rješenja ili ideja koje nam mogu pomoći u **aproksimaciji** optimalne jezgre, od kojih opisujemo dvije.

Primjer 3.1.4 (Autoregresijski model, optimalna jezgra)**.** Ovaj primjer je nastavak na primjer 3.1.3. Koristimo isti model i isti niz opservacija, ali ovoga puta koristimo optimalnu jezgru kao instrumentalnu. Koristeći definiciju od $T_k(\xi^k, \cdot)$ i specifikaciju modela (3.11)



Slika 3.2: Procjene skrivene vrijednosti X_5 modela (3.11) *s opservacijama $y_{0:5}$ pomoću algoritma sekvencijalnog uzorkovanja po važnosti korištenjem* ***optimalne jezgre****. Svaki boxplot je dobiven sa 125 ponavljanja algoritma za procjenu, i navedenim brojem čestica. Horizontalna crta prikazuje točnu vrijednost* 0.907.

imamo

$$T_k(\xi^k, f) = \frac{T_k^u(\xi^k, f)}{T_k^u(\xi^k, \mathbb{1})} = \frac{\int_{\mathsf{X}} f(x) Q(\xi^k, dx) g_{k+1}(x)}{\int_{\mathsf{X}} Q(\xi^k, dx) g_{k+1}(x)}.$$

Uvrštavanjem $Q(\xi_k, f) = \int_{\mathsf{X}} f(x) \exp(\frac{-(\xi_k - \phi x)^2}{2\sigma_U^2}) dx$, $g_k(x) = \exp(\frac{-(x-y_k)^2}{2})$ i algebarskim manipulacijama dobivamo

$$T_k(\xi^k, \cdot) \sim N\left(\frac{\xi^k \phi + \sigma_U^2 y_{k+1}}{1 + \sigma_U^2}, \frac{\sigma_U^2}{1 + \sigma_U^2}\right).$$

Slično, za inkrementalnu težinu vrijedi

$$\frac{dT_k^u(\xi^k, \cdot)}{dR_k(\xi^k, \cdot)} = T_k^u(\xi^k, \mathbb{1}) = exp\left(\frac{-(\phi \xi^k - y_{k+1})^2}{2\sigma_U^2}\right).$$

Primjenom optimalne jezgre za procjenu filtracije $\phi_5(f_{id})$ na slici 3.2, i usporedbom s rezultatima na slici 3.1, vidimo da algoritam s optimalnom jezgrom ima bolje procjene od algoritma s apriornom jezgrom.

Sada opisujemo dvije metode aproksimacije optimalne jezgre. U obje metode, pretpostavljamo da skriveni lanac poprima vrijednosti na realnom prostoru $\mathsf{X} = \mathbb{R}^n$ te je jezgra prijelaza skrivenog lanca Q apsolutno neprekidna u odnosu na Lebesgueovu mjeru, odnosno postoji izmjeriva funkcija q takva da $Q(x', f) = \int_{\mathsf{X}} f(x)q(x', x)\lambda(dx)$ za svaki $x' \in \mathsf{X}$, $f \in \mathcal{F}^b(\mathsf{X})$ i Lebesgueovu mjeru λ.

Metoda prihvaćanja-odbijanja za optimalnu jezgru

Zato što su nam poznate vrijednosti funkcije gustoće od $T_k(\xi_k, \cdot)$ do na konstantu, za uzorkovanje iz nje možemo koristiti metodu prihvaćanja-odbijanja (engl. *accept-reject*). Ona se sastoji od odabira neke druge neprekidne slučajne varijable $\mathcal{D}$ nad $(\mathsf{X}, \mathcal{X})$ s funkcijom gustoće f, takve da postoji konstanta C tako da za svaki $x \in \mathsf{X}$ vrijedi

$$\frac{1}{T_k^u(\xi^k, \mathbb{1})} q(\xi^k, x) g_{k+1}(x) \leq C f(x).$$

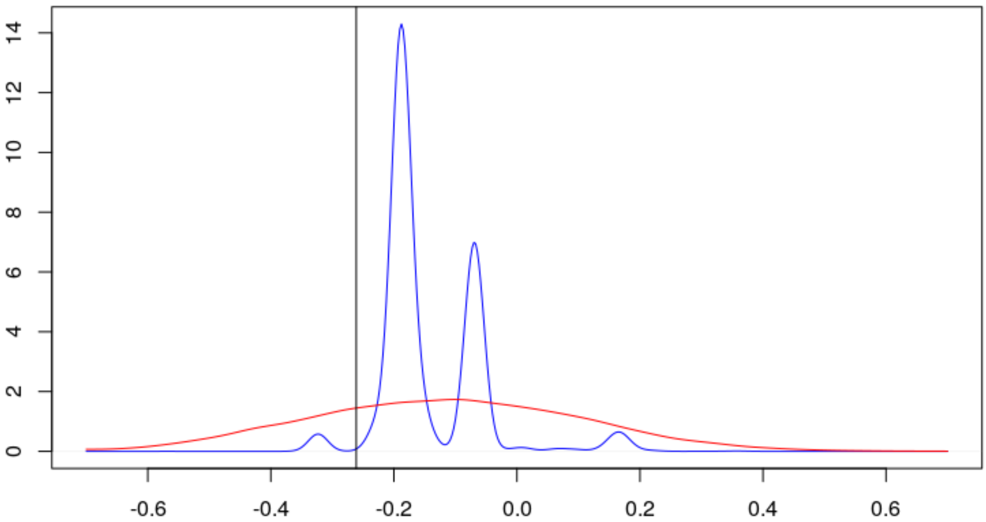
Lijeva strana nejednakosti je zapravo funkcija gustoće od $T_k(\xi^k, \cdot)$. Metoda uniformno uzorkuje točke ispod grafa funkcije $Cf(x)$, i prvu točku koja je i ispod grafa funkcije $\frac{1}{T_k^u(\xi^k, \mathbb{1})} Q(\xi^k, x) g_{k+1}(x)$ algoritam prihvaća. Ta prihvaćena točka predstavlja slučajan uzorak iz $T_k(\xi_k, \cdot)$. Detaljniji opis metode i dokaz korektnosti čitatelj može pronaći u [18] ili [3].

Lokalna linearna aproksimacija optimalne jezgre

Ako je distribucija $T_k(\xi_k, \cdot)$ unimodalna, tada za njenu aproksimaciju ima smisla koristiti neku dobro poznatu unimodalnu distribuciju, primjerice normalnu ili t-distribuciju. U primjeru 3.1.5 detaljnije razrađujemo ovu metodu aproksimacije za model stohastičke volatilnosti. Sličan račun kao u primjeru 3.1.5 može se napraviti i u slučaju svakog modela u višim dimenzijama takvog da je funkcija $\ln(q(\xi_k, x) g_k(x))$ konkavna i dva puta neprekidno diferencijabilna.

Primjer 3.1.5 (Stohastička volatilnost, aproksimacija optimalne jezgre)**.** Podsjetimo se modela za stohastičku volatilnost iz primjera 1.1.4

$$\begin{aligned} X_k &= \delta X_{k-1} + \sigma_U U_k, & U_k &\sim N(0,1), \\ Y_k &= \exp(X_k/2) V_k, & V_k &\sim N(0,1). \end{aligned}$$



Slika 3.3: Procjene distribucije ϕ_{10} sekvencijalnim uzorkovanjem po važnosti za model stohastičke volatilnosti, primjer 3.1.5. Plavo je rezultat korištenja metode aproksimacije optimalne jezgre iz primjera 3.1.5, a crveno uz korištenje apriorne jezgre. Procjene funkcije gustoće su dobivene pomoću čestica i njihovih težina koje predstavljaju uzorke iz distribucije ϕ_{10} i gaussovskih jezgara za procjenu gustoće. U algoritmu je upotrebljeno 10000 čestica. Točna simulirana vrijednost je vertikalna crna crta.

Za funkcije gustoće njegovih jezgri prijelaza vrijedi

$$\begin{aligned} q(\xi_k, x) &= C_q \exp(-(x - \delta\xi_k)^2 / 2\sigma_U^2), \\ g_k(x) &= C_g \exp(-y_k^2 \exp(-x)/2 - x/2). \end{aligned}$$

Slijedi da je funkcija $\log(q(\xi_k, x)g_k(x))$ konkavna po x jer je linearna kombinacija konkavnih funkcija. Jedinstvena stacionarna točka te funkcije tada ujedno mora biti maksimum funkcije gustoće $q(\xi_k, x)g_k(x)$, što znači da je distribucija $T_k(\xi_k, \cdot)$ unimodalna. Sada Taylorovim polinomom drugog reda oko proizvoljne točke x_0 imamo aproksimaciju

$$\log(q(\xi_k, x)g_k(x)) \approx C_0 + (x - x_0)d_1(x_0) + \frac{(x - x_0)^2}{2} d_2(x_0) \tag{3.12}$$

gdje su derivacije

$$d_1(x) = -(x - \delta\xi_k)/\sigma_U^2 + y_k^2 \exp(-x)/2 - 1/2,$$
$$d_2(x) = -1/\sigma_U^2 - y_k^2 \exp(-x)/2.$$

Funkcija $d_1(x)$ ima jedinstvenu stacionarnu točku koja je točka maksimuma funkcije gustoće. Možemo je pronaći Newtonovom metodom. Tu točku maksimuma označimo s x_m. Jednadžba (3.12) sugerira da normalna distribucija $N(x_m, s_m^2)$, gdje je $s_m = 1/\sqrt{-d_2(x_m)}$, može aproksimirati distribuciju $T_k(\xi_k, \cdot)$. S druge strane, ako želimo teže repove, a slično lokalno ponašanje oko maksimuma, možemo upotrijebiti t-distribuciju s istim očekivanjem i varijancom kao navedena normalna distribucija.

Na slici 3.3 uspoređujemo dvije jezgre za procjenu filtracijske distribucije ϕ_{10} u modelu stohastičke volatilnosti. Jedna je jezgra gornje navedena metoda za aproksimaciju optimalne jezgre s t-distribucijom s 5 stupnjeva slobode. Druga jezgra je apriorna. Podaci koje koristimo simulirani su s $\delta = 0.9, \sigma_U = 0.1$.

3.2 Sekvencijalno uzorkovanje s reuzorkovanjem

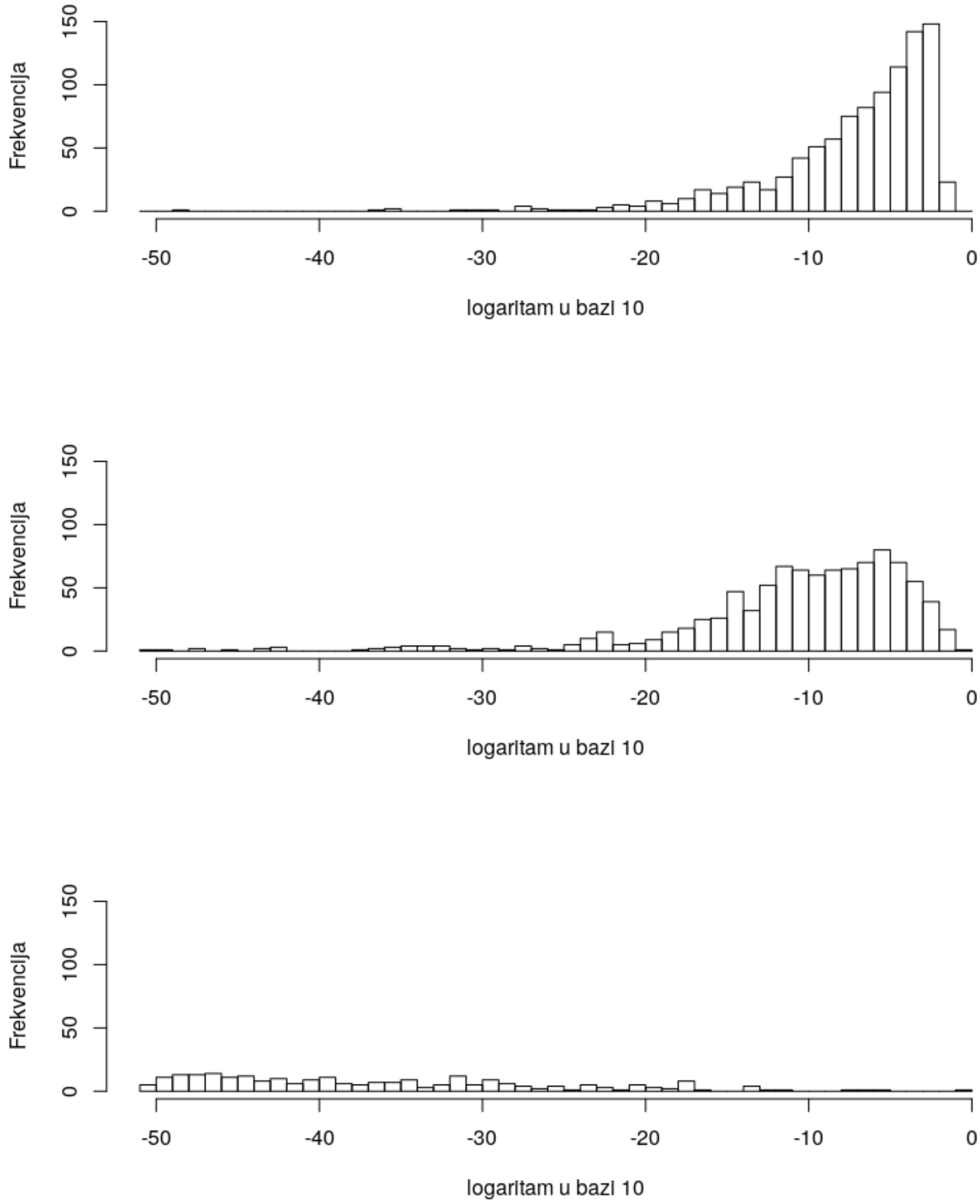
Algoritam sekvencijalnog uzorkovanja po važnosti ima veliku manu, a ona je da težine degeneriraju s rastućim brojem stanja skrivenog Markovljevog modela. U slučaju broja stanja većeg od 10, za većinu čestica njihova težina postaje beznačajna u ukupnoj sumi kojom procjenjujemo $\phi_{0:k|k}(f)$. Odnosno, ukupna masa sume $\sum_{i=1}^{M} \omega_i^k$ postane koncentrirana u tek nekoliko vrijednosti, a ostale čestice troše procesorsko vrijeme. U [12] je dokazano da slučajni proces težine jedne čestice zadovoljava martingalno svojstvo, što znači da je varijanca rastuća, a u [3] dokazuju da ta varijanca raste eksponencijalno za specifičan umjetni model.

Primjer 3.2.1 (Težine čestica u modelu stohastičke volatilnosti)**.** Kao nastavak na primjer 3.1.5, proučavamo težine čestica u algoritmu sekvencijalnog uzorkovanja na podacima i modelu opisanom u primjeru. Nakon 5, 15, 100 koraka gradimo histogram normaliziranih težina, promatrajući logaritme u bazi 10 njihovih vrijednosti. Kao što vidimo na slici 3.4, čestice su u potpunosti neupotrebljive nakon 100-tog koraka, jer je cijela suma težina sadržana u tek nekoliko čestica.

Zato kao jednu od mjera koliko dobro algoritam radi koristimo mjere degeneracije težina.

Mjerenje degeneracije težina

Pretpostavimo da želimo procijeniti degeneraciju nekog niza težina $\omega_1, \omega_2, \ldots, \omega_n$. Cilj nam je kvantificirati koliko je suma težina koncentrirana u malome broju njih. Jedna



Slika 3.4: Histogrami težina čestica algoritma sekvencijalnog uzorkovanja s optimalnom jezgrom iz primjera 3.1.5 nakon 5, 15 i 100 koraka, s 1000 čestica. Sve težine koje imaju $\log_{10}$ *manji od -50 nisu prikazane.*

klasična mjera koncentracije je statistička, poznata i kao informacijska ili Shannonova, entropija

$$E(\omega_{1:n}) = -\sum_{i=1}^{n} \frac{\omega_i}{\sum_{j=1}^{n} \omega_j} \log(\frac{\omega_i}{\sum_{j=1}^{n} \omega_j}).$$

Ako je sva suma koncentrirana u jednoj težini, entropija poprima vrijednost 0. Većim raspršenjem mase po različitim vrijednostima, entropija raste, sve do $\log(n)$ u slučaju jednakih težina. U našem slučaju želimo imati što veću entropiju.

Neka je $(\xi^{0:k})$ slučajni vektor distribuiran kao putanja instrumentalnog lanca definiranog sa ρ_0 i $\{R_i\}_{0\leq i\leq k}$. Tada slučajnu varijablu težine možemo definirati sa

$$\omega^* = g_0(\xi_0)\frac{d\nu}{d\rho_0}(\xi_0)\frac{dT_0^u(\xi_0,\cdot)}{dR_0(\xi_0,\cdot)}(\xi_1)\cdots\frac{dT_{k-1}^u(\xi_{k-1},\cdot)}{dR_{k-1}(\xi_{k-1},\cdot)}(\xi_k).$$

Težine čestica $\omega_1^k, \omega_2^k, \ldots, \omega_M^k$ (3.9) u algoritmu sekvencijalnog uzorkovanja po važnosti čine uzorak slučajne varijable ω^*. Koeficijent varijacije uzorka te varijable, odnosno devijacija podijeljena s aritmetičkom sredinom, tada je

$$CV_{\omega^*} = n\left(\sum_{i=1}^{n} \frac{1}{n}\left(\frac{\omega_i}{\sum_{j=1}^{n} \omega_j} - \frac{1}{n}\right)^2\right)^{1/2}. \tag{3.13}$$

Pomoću koeficijenta varijacije, uvodimo heurističku mjeru zvanu *efektivna veličina uzorka*

$$N_{ef} = \frac{n}{1 + CV_{\omega^*}^2} = \left(\sum_{i=1}^{n}\left(\frac{\omega_i}{\sum_{j=1}^{n} \omega_j}\right)^2\right)^{-1} \tag{3.14}$$

čiju intuiciju prenosimo iz [15]. Naime, ako od danih n težina imamo m njih koje su jednake $\frac{1}{m}$, a ostalih $n-m$ je jednako 0, tada za m, n i koeficijent varijacije vrijedi

$$m = \frac{n}{1 + CV_{\omega^*}^2}$$

gdje jednakost dobivamo uvrštavajući m, n u (3.13). Broj m na neki način predstavlja broj težina koje su nam značajne u uzorku. Stoga formulu za m postavljamo kao formulu od N_{ef} u općenitijem slučaju. Drugu jednakost u (3.14) dobivamo direktnim uvrštavanjem vrijednosti $CV_{\omega^*}^2$. Minimum efektivne veličine uzorka je 1, a postiže se kada je suma težina koncentrirana u jednoj vrijednosti. Maksimum je n, a on se postiže kada su sve težine jednake.

Dodavanje reuzorkovanja

Kao što smo spomenuli u opisu algoritma uzorkovanja po važnosti s reuzorkovanjem, reuzorkovanje pretvara uzorak s nekim težinama u uzorak s težinama 1. Mi ćemo se dodatkom reuzorkovanja boriti protiv degeneracije težina.

Naravno, korištenje reuzorkovanja ima svoju cijenu, a to je povećanje varijance procjene. To vidimo iz usporedbe granične varijance uzorkovanja po važnosti $\hat{\sigma}^2(f)$ s graničnom varijancom uzorkovanja po važnosti s reuzorkovanjem $\hat{\sigma}^2(f) + Var_\mu(f)$, (lema 2.3.7 i teorem 2.3.9, $\alpha = 1$).

Stoga, ne želimo reuzorkovati nakon svakog stanja, odnosno produljivanja putanje čestice. Opcija koju odabiremo jest da reuzorkujemo svaki puta kada efektivna veličina uzorka (3.14) padne ispod neke fiksne granice. Reuzorkovanjem mjeru efektivne veličine uzorka resetiramo na njen maksimum. Rad [15] pokazuje kako nije posebno bitno namještati tu fiksnu granicu, sve dok je granica prirodni broj veći od 1.

Po uzoru na algoritam 2.3.6, u opisu ćemo odmah zapisati σ-algebre uz koje su nizovi uvjetno nezavisni, a algoritam će biti napisan pomoću dvostrukih nizova. Prisjetimo se kako n označava veličinu niza M_n kojim aproksimiramo integrale, pa indeks čestica i ide od 1 do M_n. Izuzev dodatka notacije i koraka reuzorkovanja, ovaj algoritam je isti kao algoritam sekvencijalnog uzorkovanja po važnosti.

Algoritam 3.2.2 (Sekvencijalno uzorkovanje po važnosti s reuzorkovanjem)**.**
Inicijalizacija:

Generirajmo nezavisan i jednako distribuiran slučajan uzorak $\xi^0_{n,1}, \xi^0_{n,2}, \dots, \xi^0_{n,M_n}$ tako da $\xi^0_{n,i} \sim \rho_0$. Za svaki i takav da $1 \le i \le M_n$ izračunajmo inicijalne težine

$$\omega^0_{n,i} := g_0(\xi^0_{n,i}) \frac{d\nu}{d\rho_0}(\xi^0_{n,i}).$$

Postavljamo $\mathcal{F}^1_n = \sigma(\xi^0_{n,1} \dots, \xi^0_{n,M_n})$.

Mutacijski korak:

Za svaki vremenski indeks s koji redom ide od 1 do S,
nezavisno za svaki i takav da $1 \le i \le M_n$ uzorkujemo $\tilde{\xi}^s_{n,i} \sim R_{s-1}(\xi^{s-1}_{n,i}, \cdot)$,
računamo novu težinu i-te čestice

$$\omega^s_{n,i} := \omega^{s-1}_i \frac{dT^u_{s-1}(\xi^{s-1}_{n,i}, \cdot)}{dR_{s-1}(\xi^{s-1}_{n,i}, \cdot)}(\tilde{\xi}^s_{n,i}) = \omega^{s-1}_{n,i} g_s(\tilde{\xi}^s_{n,i}) \frac{dQ(\xi^{s-1}_{n,i}, \cdot)}{dR_{s-1}(\xi^{s-1}_{n,i}, \cdot)}(\tilde{\xi}^s_{n,i}).$$

Produljujemo putanju i-te čestice sa $\xi^{0:s}_{n,i} := (\xi^{0:s-1}_{n,i}, \tilde{\xi}^s_{n,i})$. Putanje $\xi^{0:s}_{n,1}, \dots, \xi^{0:s}_{n,M_n}$ su uvjetno nezavisne s obzirom na $\mathcal{G}^s_n = \sigma(\mathcal{F}^s_n \cup \sigma(\xi^{s-1}_{n,1}, \dots, \xi^{s-1}_{n,M_n}))$

(Opcionalno) Reuzorkovanje:

Generirajmo n.j.d prirodne brojeve $I^s_{n,1}, I^s_{n,2}, \dots, I^s_{n,M_n}$ tako da za svaki par (i, j) takav da $1 \le i, j \le M_n$ vrijedi

$$P(I^s_{n,j} = i | \xi^{0:s}_{n,1}, \xi^{0:s}_{n,2}, \dots, \xi^{0:s}_{n,M_n}) = \frac{\omega^s_{n,i}}{\sum_{k=1}^{M_n} \omega^s_{n,k}}$$

te za svaki j definirajmo

$$\hat{\xi}^{0:s}_{n,j} = \xi^{0:s}_{I_i}.$$

Putanje $\hat{\xi}^{0:s}_{n,1}, \dots, \hat{\xi}^{0:s}_{n,M_n}$ su uvjetno nezavisne s obzirom na $\mathcal{H}^s_n = \sigma(\mathcal{G}^s_n \cup \sigma(\tilde{\xi}^s_{n,1}, \dots \tilde{\xi}^s_{n,M_n}))$. Ako se uzorkovanje dogodilo, postavljamo $\mathcal{F}^{s+1}_n = \mathcal{H}^s_n$ i redefiniramo putanje čestica $\xi^{0:s}_{n,i} := \hat{\xi}^{0:s}_{n,i}$ i težine $\omega^s_{n,i} := 1$ za svaki i takav da $1 \le i \le M_n$.

Ako se reuzorkovanje nije dogodilo, postavljamo $\mathcal{F}^{s+1}_n = \mathcal{G}^s_n$.

Evaluacija:

Za funkciju $f \in \mathcal{F}^b(\mathsf{X}^{k+1})$ nad prvih $k+1$ stanja skrivenog lanca, gdje je $k \le S$ procjenjujemo

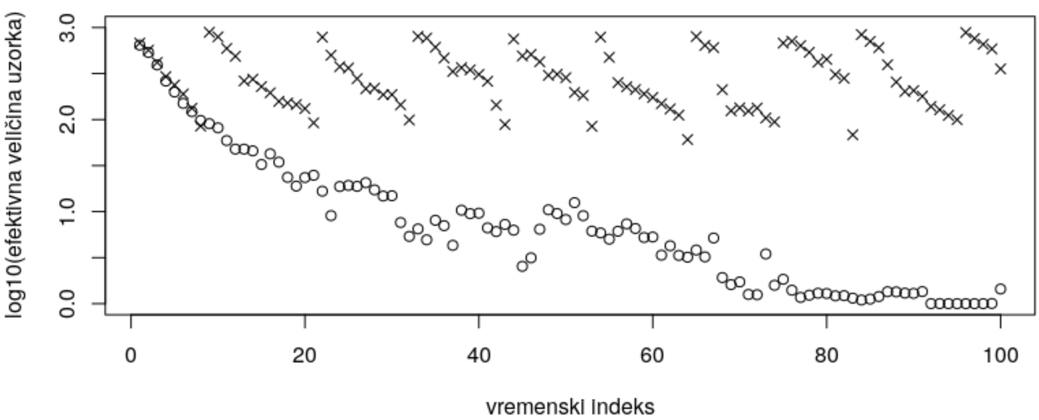
$$\phi_{0:k|k}(f) \approx \hat{\phi}^{SUVR}_{0:k|k}(f) = \sum_{i=1}^{M} \frac{\omega^k_i f(\xi^{0:k}_i)}{\sum_{l=1}^{M} \omega^k_l}.$$

Napomena 3.2.3 (O izglađivanju)**.** *Iako smo definirali algoritam 3.2.2 pomoću funkcija nad cijelom putanjom čestice, u tom obliku ga nikada nećemo iskoristiti. Za razliku od algoritma 3.1.2, čestice više nisu međusobno nezavisne, i mnoge putanje se ili gube ili dupliciraju kroz reuzorkovanje. Zbog toga je informacija o inicijalnim stanjima sadržana u česticama nepouzdana.*

Najčešći slučaj korištenja algoritma 3.2.2 je u slučaju filtracije (definicija 1.3.6). To objašnjava njegov drugi naziv, čestični filter*, koji se koristi čak i onda kada nije u pitanju* filtracija.

Korištenjem reuzorkovanja, mnoge čestice više uopće neće imati informacije o inicijalnim stanjima putanja jer se informacije o manje vjerojatnim događajima nad početnim stanjima gube kroz reuzorkovanje. Za probleme računanja integrala nad fiksnim stanjima skrivenog lanca, na primjer $\phi_0(f)$, [3] opisuju drukčije prikladne metode pod nazivom statistika fiksne dimenzije.

Primjer 3.2.4 (Usporedba težine čestica u sekvencijalnog uzorkovanju)**.** Uspoređujemo vrijednosti efektivne veličine uzorka između sekvencijalnog uzorkovanja s optimalnom



Slika 3.5: Prikaz efektivne veličine uzorka na težinama u sekvencijalnom uzorkovanju i sekvencijalnom uzorkovanju s reuzorkovanjem. Algoritmi s 1000 *čestica procjenjuju filtriranu distribuciju* ϕ_k*, gdje k ide od* 0 *do* 100*.*

jezgrom i sekvencijalnog uzorkovanja s reuzorkovanjem s optimalnom jezgrom. Za reuzorkovanje koristimo strategiju fiksne granice. Model i podaci koje koristimo su isti kao u 3.2.1. Na slici 3.5 promatramo efektivnu veličinu uzorka N_ef (3.14) kroz jedno izvođenje algoritma s i bez reuzorkovanja. Vidimo kako N_ef brzo pokazuje prisutnost degeneracije težina u algoritmu bez reuzorkovanja, dok reuzorkovanje resetira vrijednost N_ef na maksimum.

3.3 Asimptotika sekvencijalnih metoda

Nakon što pokažemo dovoljne uvjete u kojima mutacijski korak čuva asimptotsku normalnost, kombinirat ćemo ga s korakom reuzorkovanja iz drugog poglavlja i induktivno primijeniti na dokaz asimptotske normalnosti cijelog algoritma sekvencijalnog uzorkovanja s reuzorkovanjem.

Promatramo samo slučaj **filtriranja**, u kojem računamo distribucije stanja zadnjeg skrivenog stanja. Prisjetimo se da filtraciju $\phi_{k:k|k}$ skraćeno zapisujemo ϕ_k. Ako imamo niz od $S + 1$ stanja, algoritam redom prolazi kroz aproksimacije distribucija $\phi_0, \phi_1, \ldots, \phi_S$, gdje počnemo od procjene klasičnog uzorkovanja po važnosti za ϕ_0.

Iz distribucije ϕ_k prelazimo u distribuciju ϕ_{k+1} pomoću prijelazne jezgre T_k^u, instrumentalne prijelazne jezgre R_k i njihove Radon-Nikodymove derivacije $\frac{dT_k^u(x,\cdot)}{dR_k(x,\cdot)}$ za svaki $x \in \mathsf{X}$. U idućem teoremu ν predstavlja ϕ_k, a L, R redom predstavljaju T_k^u, R_k, te s obzirom na to da radimo za proizvoljni korak po stanjima, imamo $\xi_{n,i}, \tilde{\xi}_{n,i}.\omega_{n,i}$ umjesto $\xi_{n,i}^s, \tilde{\xi}_{n,i}^s, \omega_{n,i}^s$.

Distribuciju μ, koja predstavlja ϕ_{k+1}, definiramo s

$$\mu(f) := \frac{\int_{\mathsf{X}} \nu(dx)T(x, f)}{\int_{\mathsf{X}} \nu(dx)T(x, \mathsf{X})}.$$

Mjeru iz brojnika označimo sa $\nu T(f) = \int_{\mathsf{X}} \nu(dx)T(x, f)$, i normalizirajuća konstanta iz nazivnika će tada biti $\nu T(\mathsf{X})$.

Teorem 3.3.1 (Mutacijski korak čuva asimptotsku normalnost)**.** *Neka je $\{\xi_{n,i}, 1\}_{1\leq i\leq M_n}$ asimptotski normalan za $(\nu, A, \sigma, \{M_n^{1/2}\})$ i konzistentan za (ν, C).*

1. Neka je $0 < \nu T(\mathsf{X}) < \infty$.

2. Neka vrijedi $x \mapsto T(x, \mathsf{X}) \in C$ i $x \mapsto T(x, \mathsf{X}) \in A$

3. Neka za svaki $x \in \mathsf{X}$ vrijedi $T(x, \cdot) \ll R(x, \cdot)$, i postoji pozitivna verzija od $\frac{dT(x,\cdot)}{dR(x,\cdot)}$.

Tada je $\{\tilde{\xi}_{n,i}, \omega_{n,i}\}_{1\leq i\leq M_n}$ definirano kao u algoritmu 3.2.2 asimptotski normalan za $(\mu, \tilde{A}, \tilde{\sigma}, \{M_n^{1/2}\})$, gdje vrijedi

$$\tilde{A} = \{f \in \mathcal{L}^2(\mathsf{X}, \mu) : x \mapsto T(x, f) \in A, x \mapsto \int_{\mathsf{X}} R(x, dx') \left(\frac{dT(x, \cdot)}{dR(x, \cdot)}(x')\right)^2 f(x')^2 \in C\}$$

$$\tilde{\sigma}^2(f) = \frac{\sigma^2(\nu T[f - \mu(f)]) + \eta^2(f - \mu(f))}{\nu T(\mathsf{X})^2}$$

$$\eta^2(f) = \iint\limits_{\mathsf{X}\times\mathsf{X}} \nu(dx)R(x, dx') \left(\frac{dT(x, \cdot)}{dR(x, \cdot)}(x')f(x')\right)^2 - \int_{\mathsf{X}} \nu(dx)T(x, f)^2.$$

Dokaz. Cjelovitost skupa $\tilde{A}$ slijedi korištenjem Jensenove nejednakosti, linearnosti integriranja i iz cjelovitosti od A i C. Neka je $f \in \tilde{A}$ tako da bez smanjenja općenitosti vrijedi $\mu(f) = 0$. Želimo dokazati

$$M_n^{1/2} \sum_{i=1}^{M_n} \frac{\omega_{n,i}}{\sum_{j=1}^{M_n} \omega_{n,j}} f(\tilde{\xi}_{n,i}) \xrightarrow[n\to\infty]{\mathcal{D}} N(0, \tilde{\sigma}^2(f)). \tag{3.15}$$

Dokaz rastavljamo na nekoliko koraka.

Korak 1. U prvom koraku treba dokazati da vrijedi konvergencija

$$\sum_{i=1}^{M_n} \frac{\omega_{n,i}}{M_n} \xrightarrow[n\to\infty]{\mathcal{P}} \nu T(\mathsf{X})$$

iz čega je onda, po Slutskyevom teoremu A.7 za dijeljenje, za (3.15) dovoljno dokazati da vrijedi

$$M_n^{1/2} \sum_{i=1}^{M_n} \frac{\omega_{n,i}}{M_n} f(\tilde{\xi}_{n,i}) \xrightarrow[n\to\infty]{\mathcal{D}} N(0, \nu T(\mathsf{X})^2 \tilde{\sigma}^2(f)). \tag{3.16}$$

Dokaz *koraka 1.* preskačemo zbog konciznosti, a čitatelja upućujemo na 9.3 u [3].

Raspišimo izraz u (3.16)

$$M_n^{1/2} \sum_{i=1}^{M_n} \frac{\omega_{n,i}}{M_n} f(\tilde{\xi}_{n,i}) = M^{1/2}(A_n + B_n) \ (g.s.)$$

gdje je

$$A_n = \frac{1}{M_n} \sum_{i=1}^{M_n} \mathbb{E}[\omega_{n,j} f(\tilde{\xi}_{n,i}) | \mathcal{F}_n] \ (g.s.)$$

$$B_n = \frac{1}{M_n} \sum_{i=1}^{M_n} \omega_{n,i} f(\tilde{\xi}_{n,i}) - \mathbb{E}[\omega_{n,i} f(\tilde{\xi}_{n,i}) | \mathcal{F}_n] \ (g.s.)$$

Za tvrdnju zadatka je sada dovoljno dokazati asimptotsku normalnost od A_n, a na B_n primijeniti lemu A.5, i iskoristiti lemu A.6 za kombiniranje konvergencija.

Korak 2. Dokazujemo asimptotsku normalnost od A_n. Po konstrukciji mutacijskog koraka imamo da

$$\begin{aligned}
\mathbb{E}[\omega_{n,i} f(\tilde{\xi}_{n,i}) | \mathcal{F}_n] &= \int_{\mathsf{X}} R(\xi_{n,i}, dx) \omega_{n,i}(x) f(x) \\
&= \int_{\mathsf{X}} R(\xi_{n,i}, dx) \frac{dT(\xi_{n,i}, \cdot)}{dR(\xi_{n,i}, \cdot)}(x) f(x) \\
&= T(\xi_{n,i}, f) \ (g.s.)
\end{aligned} \tag{3.17}$$

Tada za A_n vrijedi

$$A_n = \frac{1}{M_n} \sum_{i=1}^{M_n} \mathbb{E}[\omega_{n,i} f(\tilde{\xi}_{n,i}) | \mathcal{F}_n] = \frac{1}{M_n} \sum_{i=1}^{M_n} T(\xi_{n,i}, f) \ (g.s.)$$

i zbog $f \in \tilde{A}$ imamo po definiciji da je $x \mapsto T(x, f) \in A$ pa vrijedi asimptotska normalnost

$$M_n^{1/2} A_n \xrightarrow[n\to\infty]{\mathcal{D}} N(0, \sigma^2(\nu T f)). \tag{3.18}$$

Korak 3. Za asimptotsku normalnost od B_n, prvo označimo $V_{n,i} = M_n^{-1/2} \omega_{n,i} f(\tilde{\xi}_{n,i}) \ (g.s.)$.

Varijable $V_{n,i}$ su po konstrukciji uvjetno nezavisne s obzirom na $\mathcal{G}_n$ kao u algoritmu 3.2.2, i dokazat ćemo sva tri svojstva koja su potrebna za primjenu teorema A.5. Analogno kao u (3.17), za $V_{n,i}^2$

$$\mathbb{E}[V_{n,i}^2|\mathcal{F}_n] = \frac{1}{M_n}\int_{\mathsf{X}} R(\xi_{n,i},dx)\left(\frac{dT(\xi_{n,i},\cdot)}{dR(\xi_{n,i},\cdot)}(x)f(x)\right)^2, \tag{3.19}$$

$$\sum_{i=1}^{M_n}\mathbb{E}[V_{n,i}^2|\mathcal{F}_n] = \frac{1}{M_n}\sum_{i=1}^{M_n}\int_{\mathsf{X}} R(\xi_{n,i},dx)\left(\frac{dT(\xi_{n,i},\cdot)}{dR(\xi_{n,i},\cdot)}(x)f(x)\right)^2$$

$$\xrightarrow[n\to\infty]{\mathcal{P}} \iint\limits_{\mathsf{X}\times\mathsf{X}} \nu(dx')R(x',dx)\left(\frac{dT(\xi_{n,i},\cdot)}{dR(\xi_{n,i},\cdot)}(x)f(x)\right)^2$$

gdje imamo konvergenciju jer je funkcija $x \mapsto \int_{\mathsf{X}} R(\xi_{n,i},dx)\left(\frac{dT(\xi_{n,i},\cdot)}{dR(\xi_{n,i},\cdot)}(x)f(x)\right)^2$ po definiciji u skupu C. Iz te pripadnosti slijedi i integrabilnost funkcije, stoga $V_{n,i}$ ima konačan drugi uvjetni moment (3.19), što je upravo svojstvo A.5.i.

Zbog uvjetne Jensenove nejednakosti imamo $\mathbb{E}[V_{n,i}^2|\mathcal{F}_n] \geq \mathbb{E}[V_{n,i}|\mathcal{F}_n]^2$ $(g.s.)$, što s cjelovitosti skupa C pokazuje da funkcija $t \mapsto \left(\int_{\mathsf{X}} R(t,dx)\left(\frac{dT(t,\cdot)}{dR(t,\cdot)}(x)f(x)\right)\right)^2$ pripada C. Sada

$$\mathbb{E}[V_{n,i}|\mathcal{F}_n]^2 = \frac{1}{M_n}\left(\int_{\mathsf{X}} R(\xi_{n,i},dx)\left(\frac{dT(\xi_{n,i},\cdot)}{dR(\xi_{n,i},\cdot)}(x)f(x)\right)\right)^2 = \frac{1}{M_n}T(\xi_{n,i},f)^2 \ (g.s.),$$

$$\sum_{i=1}^{M_n}\mathbb{E}[V_{n,i}|\mathcal{F}_n]^2 = \frac{1}{M_n}\sum_{i=1}^{M_n}T(\xi_{n,i},f)^2 \xrightarrow[n\to\infty]{\mathcal{P}} \int_{\mathsf{X}}\nu(dx)T(x,f)^2. \tag{3.20}$$

Oduzimanjem konvergencija (3.19) i (3.20) dobivamo uvjet A.5.ii.,

$$\sum_{i=1}^{M_n}\mathbb{E}[V_{n,i}^2|\mathcal{F}_n] - \mathbb{E}[V_{n,i}|\mathcal{F}_n]^2 \xrightarrow[n\to\infty]{\mathcal{P}}$$

$$\iint\limits_{\mathsf{X}\times\mathsf{X}} \nu(dx')R(x',dx)\left(\frac{dT(\xi_{n,i},\cdot)}{dR(\xi_{n,i},\cdot)}(x)f(x)\right)^2 - \int_{\mathsf{X}}\nu(dx)T(x,f)^2 = \eta^2(f),$$

što nam daje izraz za graničnu varijancu od B_n.

Za A.5.iii., uzmimo proizvoljan pozitivan realan ϵ. Vrijedi

$$\mathbb{E}[V_{n,i}^2\mathbb{1}_{\{|V_{n,i}|\geq\epsilon\}}|\mathcal{F}_n] = \frac{1}{M_n}\int_{\mathsf{X}} R(\xi_{n,i},dx)\left(\frac{dT(\xi_{n,i},\cdot)}{dR(\xi_{n,i},\cdot)}(x)f(x)\right)^2\mathbb{1}_{\left\{\left|\frac{dT(\xi_{n,i},\cdot)}{dR(\xi_{n,i},\cdot)}(x)f(x)\right|\geq\epsilon M_n^2\right\}}.$$

Slično kao u dokazu 2.3.8, možemo uzeti pozitivan realan broj K, i za dovoljno velik n će tada biti

$$\sum_{i=1}^{M_n} \mathbb{E}[V_{n,i}^2 \mathbb{1}_{\{|V_{n,i}|\geq\epsilon\}}|\mathcal{F}_n] \leq$$
$$\frac{1}{M_n}\sum_{i=1}^{M_n}\int_{\mathsf{X}} R(\xi_{n,i},dx)\left(\frac{dT(\xi_{n,i},\cdot)}{dR(\xi_{n,i},\cdot)}(x)f(x)\right)^2 \mathbb{1}_{\left\{\left|\frac{dT(\xi_{n,i},\cdot)}{dR(\xi_{n,i},\cdot)}(x)f(x)\right|\geq K\right\}} \xrightarrow[n\to\infty]{\mathcal{P}}$$
$$\iint_{\mathsf{X}^2} \nu(dx')R(x',dx)\left(\frac{dT(x',\cdot)}{dR(x',\cdot)}(x)f(x)\right)^2 \mathbb{1}_{\left\{\left|\frac{dT(x',\cdot)}{dR(x',\cdot)}(x)f(x)\right|\geq K\right\}}. \tag{3.21}$$

gdje konvergencija slijedi zbog cjelovitosti od C. Zadnju liniju (3.21) možemo napraviti proizvoljno malom puštanjem $K \to \infty$. Sada korištenje leme A.2 dovršava dokaz svojstva A.5.iii.

Korak 4. Dokazali smo

$$M_n^{1/2}A_n \xrightarrow[n\to\infty]{\mathcal{D}} N(0,\sigma^2(Tf)),$$
$$\exp(ivM_n^{1/2}B_n) \xrightarrow[n\to\infty]{\mathcal{P}} exp(-v^2\eta^2(f)/2).$$

Jednako kao u završetku dokaza 2.3.10, primjenom leme A.6 dobivamo bivarijatnu normalnu distribuciju i uzimanjem linearne kombinacije $M_n^{1/2}A_n + (M_n^{1/2}M_n^{-1/2})M_n^{1/2}B_n$ dobivamo željenu asimptotsku normalnost. □

Prije teorema o asimptotskoj normalnosti, navedimo teorem o uvjetima u kojima se čuva konzistentnost. Tehnike dokazivanja konzistentnosti su slične onima kojima se dokazuje asimptotska normalnost, i dokaz teorema se može pronaći u [3], potpoglavlje 9.4.

Teorem 3.3.2 (Konzistentnost sekvencijalnog uzorkovanja po važnosti s reuzorkovanjem)**.** *Neka vrijedi*

1. $\int_{\mathsf{X}} Q(x,dx')g_k(x') > 0$ *za svaki* $x \in \mathsf{X}$ *i indeks* k.
2. $\sup_{x\in\mathsf{X}} g_k(x) < \infty$ *za svaki indeks* k.
3. *Za svaki* $x \in \mathsf{X}$ *i indeks* k *jezgra* $T_k^u(x,\cdot)$ *je apsolutno neprekidna u odnosu na* $R_k(x,\cdot)$ *s pozitivnom verzijom Radon-Nikodymove derivacije* $\frac{dT_k^u(x,\cdot)}{dR_k(x,\cdot)}$ *tako da vrijedi i* $\sup_{(x,x')\in\mathsf{X}\times\mathsf{X}} \frac{dT_k^u(x,\cdot)}{dR_k(x,\cdot)}(x') < \infty$.

Ako je $\{\xi_{n,i}^0, 1\}_{1\leq i\leq M_n}$ *konzistentan dvostruki niz za* $(\phi_0, \mathcal{L}^1(\mathsf{X},\phi_0))$ *onda je za svaki* $k > 0$ *dvostruki niz* $\{\xi_{n,i}^k, 1\}_{1\leq i\leq M_n}$ *konzistentan za* $(\phi_k, \mathcal{L}^1(\mathsf{X},\phi_k))$.

Uz izbor apriorne jezgre $R_k = Q$, uvjet 3. slijedi iz uvjeta $g_k(x) > 0$ za svaki $x \in \mathsf{X}$ i uvjeta 2. S druge strane, uz izbor optimalne jezgre $R_k = T_k$ zbog

$$\frac{dT_k^u(x,\cdot)}{dT_k(x,\cdot)}(x') = T_k^u(x,\mathsf{X}) = \int_{\mathsf{X}} Q(x,dx')g_k(x')$$

pa pozitivnost uvjeta 3. slijedi iz uvjeta 1., i ograničenost po (x,x') slijedi iz uvjeta 2. i konačnosti vjerojatnosne mjere $Q(x,\cdot)$ za svaki x.

Idući teorem govori o asimptotskoj normalnosti sekvencijalnog algoritma s reuzorkovanjem ako u svakom koraku primijenimo reuzorkovanje. Teorem nećemo koristiti u praksi, ali nam daje nekoliko uvida. Svi sekvencijalni Monte Carlo algoritmi koriste slične korake s modifikacijama i usporedba graničnih vrijednosti varijanci tih algoritama nam daje do znanja jesu li neke metode efikasnije od drugih. S druge strane, osim za dokaz teorema, izveli smo alate koji nam mogu koristiti u raznim situacijama proučavajući sekvencijalne Monte Carlo algoritme. Također, provjeravanje uvjeta idućeg teorema nam u praksi može naznačiti koji dijelovi modela stvaraju problem u primjenama u kojima koristimo opisane algoritme.

Teorem 3.3.3 (Asimptotska normalnost sekvencijalnog uzorkovanja po važnosti s reuzorkovanjem)**.** *Neka vrijede svi uvjeti teorema 3.3.2, i neka je* $\{\xi^0_{n,i}, 1\}$ *konzistentan za* $(\phi_0, \mathcal{L}^1(\mathsf{X},\phi_0))$. *Tada vrijedi, ako je* $\{\xi^0_{n,i}, 1\}_{1\le i\le M_n}$ *asimptotski normalan za* $(\phi_0, \mathcal{L}^2(\mathsf{X},\phi_0), \sigma_0, \{M_n^{1/2}\})$*, onda je za svaki* $k > 0$ *dvostruki niz* $\{\xi^k_{n,i}, 1\}$ *asimptotski normalan za* $(\phi_k, \mathcal{L}^2(\mathsf{X},\phi_k), \sigma_k, \{M_n^{1/2}\})$*, gdje je* σ_k *definiran rekurzivno sa*

$$\sigma_k^2(f) = \frac{\sigma_{k-1}^2(\phi_{k-1}T_k^u[f-\phi_k(f)]) + \eta_k^2(f-\phi_k(f))}{\phi_{k-1}T_k^u(\mathsf{X})} + Var_{\phi_k}(f),$$

$$\eta_k^2(f) = \iint_{\mathsf{X}\times\mathsf{X}} \phi_{k-1}(dx)R_k(x,dx')\left(\frac{dT_k^u(x,\cdot)}{dR_k(x,\cdot)}(x')f(x')\right)^2 - \int_{\mathsf{X}} \phi_{k-1}(dx)T_k^u(x,f)^2.$$

Dokaz. Prvo na dvostruki niz $\{\xi^{k-1}_{n,i}, 1\}$ primjenjujemo mutacijski korak. Stoga moramo provjeriti zadovoljava li on pretpostavke teorema o mutacijskom koraku 3.3.1. Zbog teorema 3.3.2 imamo konzistentnost $\{\xi^{k-1}_{n,i}, 1\}$ za $\mathcal{L}^1(\mathsf{X},\phi_{k-1})$ i konzistentnost $\{\xi^k_{n,i}, 1\}$ za $\mathcal{L}^1(\mathsf{X},\phi_k)$, a zbog induktivne pretpostavke imamo asimptotsku normalnost za $\mathcal{L}^2(\mathsf{X},\phi_{k-1})$.

Prvi uvjet 3.3.1 jest

$$0 < \phi_{k-1}T_{k-1}^u(\mathsf{X}) = \int_{\mathsf{X}} \phi_{k-1}(dx)T_{k-1}^u(x,\mathsf{X}) = \iint_{\mathsf{X}\times\mathsf{X}} \phi_{k-1}(dx)Q(x,dx')g_k(x') < \infty.$$

Pozitivnost slijedi iz uvjeta $\int_{\mathsf{X}} Q(x,dx')g_k(x') > 0$ za svaki x i jer je ϕ_{k-1} vjerojatnosna mjera. Konačnost slijedi iz $\sup_{x\in\mathsf{X}} g_k(x) < \infty$ i jer su ϕ_{k-1} i $Q(x,\cdot)$ vjerojatnosne mjere.

Drugi uvjet teorema 3.3.1 govori o pripadnosti funkcije $x \mapsto T^u_{k-1}(x, \mathsf{X})$ prostorima $\mathcal{L}^2(\mathsf{X}, \phi_{k-1})$ i $\mathcal{L}^1(\mathsf{X}, \phi_{k-1})$. Dovoljno je

$$\int_{\mathsf{X}} \phi_{k-1}(dx) T^u_{k-1}(x, \mathsf{X})^2 = \int_{\mathsf{X}} \phi_{k-1}(dx) \left(\int_{\mathsf{X}} Q(x, dx') g_k(x') \right)^2 \leq \iint\limits_{\mathsf{X}\times\mathsf{X}} \phi_{k-1}(dx) Q(x, dx') g_k(x')^2 < \infty.$$

gdje nejednakost slijedi prema Jensenovoj nejednakosti, a konačnost iz uvjeta $\sup_{x \in \mathsf{X}} g_k(x) < \infty$. Treći uvjet teorema 3.3.1 je direktna posljedica uvjeta o konačnosti i pozitivnosti Radon-Nikodymovih derivacija.

Ostaje provjeriti obuhvaća li skup funkcija $\tilde{A}$ iz teorema 3.3.1 skup $\mathcal{L}^2(\mathsf{X}, \phi_k)$ te primijeniti korak reuzorkovanja. Uzmimo funkciju $f \in \mathcal{L}^2(\mathsf{X}, \phi_k)$. Prvo trebamo dokazati da vrijedi $x \mapsto T^u_{k-1}(x, f) \in \mathcal{L}^2(\mathsf{X}, \phi_{k-1})$. Imamo, koristeći Jensenovu nejednakost i uvjete konačnosti,

$$\int_{\mathsf{X}} \phi_{k-1}(dx) T(x, f)^2 = \int_{\mathsf{X}} \phi_{k-1}(dx) \left(\int_{\mathsf{X}} Q(x, dx') g_k(x') f(x') \right)^2 \leq$$
$$\iint\limits_{\mathsf{X}\times\mathsf{X}} \phi_{k-1}(dx) Q(x, dx') g_k(x')^2 f(x')^2 \leq \|g_k\|_\infty \iint\limits_{\mathsf{X}\times\mathsf{X}} \phi_{k-1}(dx) Q(x, dx') g_k(x') f(x')^2 =$$
$$\|g_k\|_\infty \phi_{k-1} T^u_{k-1}(\mathsf{X}) \int_{\mathsf{X}} \phi_k(dx') f(x')^2 < \infty.$$

Drugo, trebamo dokazati da $x \mapsto \int_{\mathsf{X}} R_{k-1}(x, dx') \left(\frac{dT^u_{k-1}(x,\cdot)}{dR_{k-1}(x,\cdot)}(x') \right)^2 f(x')^2 \in \mathcal{L}^1(\mathsf{X}, \phi_{k-1})$. Analogno, koristeći konačnost Radon-Nikodymovih derivacija, vrijedi

$$\begin{aligned}
\int_{\mathsf{X}} \phi_{k-1}(dx) \int_{\mathsf{X}} R_{k-1}(x, dx') \left(\frac{dT^u_{k-1}(x,\cdot)}{dR_{k-1}(x,\cdot)}(x') \right)^2 f(x')^2 &= \\
\int_{\mathsf{X}} \phi_{k-1}(dx) \int_{\mathsf{X}} T^u_{k-1}(x, dx') \frac{dT^u_{k-1}(x,\cdot)}{dR_{k-1}(x,\cdot)}(x') f(x')^2 &\leq \\
\phi_{k-1} T^u_{k-1}(\mathsf{X}) \int_{\mathsf{X}} \phi_k(dx') \frac{dT^u_{k-1}(x,\cdot)}{dR_{k-1}(x,\cdot)}(x') f(x')^2 &\leq \\
\left\| \frac{dT^u_{k-1}(x,\cdot)}{dR_{k-1}(x,\cdot)}(x') \right\|_\infty \phi_{k-1} T^u_{k-1}(\mathsf{X}) \int_{\mathsf{X}} \phi_k(dx') f(x')^2 &< \infty.
\end{aligned} \tag{3.22}$$

što dokazuje korektnost primjene mutacijskog koraka. Nejednakost (3.22) pokazuje dovoljne uvjete za pripadnost funkcija $\frac{dT^u_{k-1}(x,\cdot)}{dR_{k-1}(x,\cdot)}(x')$, $\frac{dT^u_{k-1}(x,\cdot)}{dR_{k-1}(x,\cdot)}(x') f(x')^2$ skupu $\in \mathcal{L}^1(\mathsf{X}, \phi_k)$. To je dovoljno za primjenu leme o reuzorkovanju. Potpuno analogno kao u dokazu teorema 2.3.10 primjenom reuzorkovanja dobivamo pribrojnik $Var_{\phi_k}(f)$ u rekurzivnoj definiciji varijance σ_k^2. $\square$

Dodatak

A Asimptotski rezultati o uvjetno nezavisnim dvostrukim nizovima

U ovom dodatku navodimo i dokazujemo pomoćne rezultate vezane uz asimptotiku dvostrukih nizova koji su uvjetno nezavisni s obzirom na neki niz σ-algebri $\{\mathcal{F}_n\}$. Dvije glavne tvrdnje, propozicija A.3 i teorem A.5 su iz 9. poglavlja u [3]. Dokaz teorema A.5 ne slijedi iste argumente, nego koristi lemu A.4 koja je iz [4].

Počinjemo s dvije leme koje koristimo u dokazu propozicije A.3.

Lema A.1. *Neka su $z_1, z_2, \dots, z_n$ i $z_1', z_2', \dots, z_n'$ kompleksni brojevi takvi da za svaki od njih vrijedi $|z| \leq 1$. Tada vrijedi*

$$|z_1 z_2 \cdots z_n - z_1' z_2' \cdots z_n'| \leq \sum_{i=1}^{n} |z_i - z_i'|.$$

Dokaz. Dokaz indukcijom po $k \in \mathbb{N}$. Baza indukcije očito vrijedi. Promotrimo sada $|z_1 z_2 \cdots z_{k+1} - z_1' z_2' \cdots z_{k+1}'|$. Imamo

$$\begin{aligned} |z_1 z_2 \cdots z_{k+1} - z_1' z_2' \cdots z_{k+1}'| &= |z_1 z_2 \cdots z_{k+1} - z_1' z_2' \cdots z_{k+1} + z_1' z_2' \cdots z_k' (z_{k+1} - z_{k+1}')| \\ &\leq |z_{k+1}| |z_1 z_2 \cdots z_k - z_1' z_2' \cdots z_k'| + |z_1' z_2' \cdots z_k'| |z_{k+1} - z_{k+1}'| \\ &\overset{pretp.}{\leq} \sum_{i=1}^{k} |z_i - z_i'| + |z_{k+1} - z_{k+1}'|. \end{aligned}$$

$\square$

Lema A.2. *Neka je $\{X_n\}_{n\geq 0}$ niz slučajnih varijabli takvih da vrijedi $X_n \geq 0$ $(g.s.)$ za svaki n. Neka za svaki $\epsilon > 0$ postoji niz $\{A_n^\epsilon\}_{n\geq 0}$ slučajnih varijabli takvih da za svaki n vrijedi $A_n^\epsilon \geq X_n$ $(g.s.)$ i taj niz konvergira po vjerojatnosti u ϵ, to jest $A_n^\epsilon \xrightarrow[n\to\infty]{\mathcal{P}} \epsilon$. Tada vrijedi $X_n \xrightarrow[n\to\infty]{\mathcal{P}} 0$.*

Dokaz. Ako pretpostavimo suprotno, postoje $\epsilon_0 > 0$ i $\epsilon_1 > 0$ takvi da postoji niz n_k i vrijedi $P(|X_{n_k}|> \epsilon_0) > \epsilon_1$ za svaki $k \in \mathbb{N}$. Promatrajmo niz $\{A_n^{\frac{\epsilon_0}{4}}\}$. Iz $A_{n_k}^{\frac{\epsilon_0}{4}} \geq X_{n_k}$ $(g.s.)$ i zbog $X_{n_k} = |X_{n_k}|$ slijedi da $\{|A_{n_k}^{\frac{\epsilon_0}{4}} - \epsilon_0/4|> \epsilon_0/4\} \supseteq \{|X_n|> \epsilon_0\}$ do na skup mjere 0. Stoga za svaki k vrijedi $P(|A_{n_k}^{\frac{\epsilon_0}{4}} - \epsilon_0/4|> \epsilon_0/4) > \epsilon_1$, što je kontradikcija s konvergencijom po vjerojatnosti niza $\{A_n^{\frac{\epsilon_0}{4}}\}$. □

Pomoću sljedeće propozicije dokazujemo glavni pomoćni rezultat, teorem A.5, ali u propoziciji zapravo obavljamo većinu "posla" dokazivanja tog teorema.

Imaginarnu jedinicu označavamo sa i.

Propozicija A.3. *Neka je $\{U_{n,i}\}_{1\leq i\leq M_n}$ dvostruki niz koji je uvjetno nezavisan s obzirom na $\{\mathcal{F}_n\}_{n\geq 0}$, niz σ-algebri koje su podskupovi od $\mathcal{F}$. Neka redom vrijede i svojstva:*

i. *Za svaki par (n, i) takav da je $1 \leq i \leq M_n$, vrijedi $\mathbb{E}[U_{n,i}|\mathcal{F}_n] = 0$ $(g.s.)$ te $\mathbb{E}[U_{n,i}^2|\mathcal{F}_n] < \infty$ $(g.s.)$.*

ii. *Uz oznake $\sigma_{n,i}^2 = \mathbb{E}[U_{n,i}^2|\mathcal{F}_n]$ $(g.s.)$, postoji pozitivan realan broj σ takav da vrijedi $\sum_{i=1}^{M_n} \sigma_{n,i}^2 \xrightarrow[n\to\infty]{\mathcal{P}} \sigma^2$.*

iii. *Za svaki $\epsilon > 0$ vrijedi*

$$\sum_{i=1}^{M_n} \mathbb{E}[U_{n,i}^2 \mathbb{1}_{\{|U_{n,i}|\geq\epsilon\}}|\mathcal{F}_n] \xrightarrow[n\to\infty]{\mathcal{P}} 0.$$

Tada za svaki $u \in \mathbb{R}$ vrijedi

$$\mathbb{E}[exp(\mathrm{i}u \sum_{i=1}^{M_n} U_{n,i})|\mathcal{F}_n] \xrightarrow[n\to\infty]{\mathcal{P}} exp(-u^2\sigma^2/2).$$

Dokaz. Fiksirajmo proizvoljan $u \in \mathbb{R}$. Iz svojstva ii. i zbog neprekidnosti funkcije eksponenciranja slijedi

$$exp(-\frac{u^2}{2} \sum_{i=1}^{M_n} \sigma_{n,i}^2) \xrightarrow[n\to\infty]{\mathcal{P}} exp(-u^2\sigma^2/2).$$

Dakle dovoljno je dokazati da vrijedi

$$\mathbb{E}[exp(\mathrm{i}u \sum_{i=1}^{M_n} U_{n,i})|\mathcal{F}_n] - exp(-\frac{u^2}{2} \sum_{i=1}^{M_n} \sigma_{n,i}^2) \xrightarrow[n\to\infty]{\mathcal{P}} 0, \tag{A1}$$

a iz leme A.1 i uvjetne nezavisnosti dvostrukog niza $\{U_{n,i}\}$ slijedi

$$\left|\mathbb{E}[exp(\mathrm{i}u \sum_{i=1}^{M_n} U_{n,i})|\mathcal{F}_n] - exp(-\frac{u^2}{2} \sum_{i=1}^{M_n} \sigma_{n,i}^2)\right| \leq \sum_{1=1}^{M_n} \left|\mathbb{E}[exp(\mathrm{i}uU_{n,i})|\mathcal{F}_n] - exp(-u^2\sigma_{n,i}^2/2)\right| \ (g.s.)$$

Sada nastavljamo dokazujući da vrijedi

$$\text{za } A_n \overset{def.}{=} \sum_{1=1}^{M_n} \left|\mathbb{E}[exp(\text{i}uU_{n,i})|\mathcal{F}_n] - (1 - u^2\sigma_{n,i}^2/2)\right|, \qquad A_n \xrightarrow[n\to\infty]{\mathcal{P}} 0,$$

$$\text{za } B_n \overset{def.}{=} \sum_{1=1}^{M_n} |1 - u^2\sigma_{n,i}^2/2 - exp(-u^2\sigma_{n,i}^2/2)|, \qquad B_n \xrightarrow[n\to\infty]{\mathcal{P}} 0.$$

Prema nejednakosti dobivenoj preko Taylorovog raspisa kompleksne eksponencijalne funkcije (26.4. u [1]) za realne u, x vrijedi

$$|exp(\text{i}ux) - (1 + \text{i}ux - (ux)^2/2)| \leq min\{|ux|^2, |ux|^3\}$$

pa uvrštavanjem $x = U_{n,i}$, i korištenjem monotonosti i Jensenove nejednakosti za uvjetno očekivanje imamo

$$\left|\mathbb{E}[exp(iuU_{n,i})|\mathcal{F}_n] - (1 + \mathbb{E}[iuU_{n,i}|\mathcal{F}_n] - \mathbb{E}\left[(uU_{n,i})^2/2|\mathcal{F}_n\right])\right| \overset{Jen.}{\leq}$$
$$\mathbb{E}\left[\left|exp(iuU_{n,i}) - (1 + iuU_{n,i} - (uU_{n,i})^2/2)\right| |\mathcal{F}_n\right] \overset{mon.}{\leq} \mathbb{E}[min\{|uU_{n,i}|^2, |uU_{n,i}|^3\}|\mathcal{F}_n] \ (g.s.)$$

Koristeći svojstva minimuma, nadalje dobivamo nejednakosti

$$\begin{aligned}\mathbb{E}[min\{|uU_{n,i}|^2, |uU_{n,i}|^3\}|\mathcal{F}_n] &\leq \mathbb{E}[|uU_{n,i}|^2 \mathbb{1}_{\{|U_{n,i}|\geq\epsilon\}}|\mathcal{F}_n] + \mathbb{E}[|uU_{n,i}|^3 \mathbb{1}_{\{|U_{n,i}|<\epsilon\}}|\mathcal{F}_n] \\ &\leq u^2\mathbb{E}[U_{n,i}^2 \mathbb{1}_{\{|U_{n,i}|\geq\epsilon\}}|\mathcal{F}_n] + |u|^3\epsilon\sigma_{n,i}^2 \ (g.s.)\end{aligned}$$

Prvi redak u gornjem izrazu je jedan pribrojnik od A_n, a zbog svojstva i. jedan član postane 0, stoga vrijedi

$$A_n \leq \sum_{i=1}^{M_n} u^2\mathbb{E}[U_{n,i}^2 \mathbb{1}_{\{|U_{n,i}|\geq\epsilon\}}|\mathcal{F}_n] + |u|^3\epsilon\sigma_{n,i}^2 \ (g.s.) \tag{A2}$$

Za izraz pod sumom u (A2), prvi dio zbog iii. konvergira po vjerojatnosti u 0, a desni dio zbog ii. ima limes te vrijedi

$$\sum_{i=1}^{M_n} u^2\mathbb{E}[U_{n,i}^2 \mathbb{1}_{\{|U_{n,i}|\geq\epsilon\}}|\mathcal{F}_n] + |u|^3\epsilon\sigma_{n,i}^2 \xrightarrow[n\to\infty]{\mathcal{P}} 0 + |u|^3\epsilon\sigma^2.$$

Primjenom leme A.2 sada slijedi $A_n \xrightarrow[n\to\infty]{\mathcal{P}} 0$.

Za B_n slično kao gore primijenimo Taylorovu nejednakost $|e^{-x} + x - 1| \leq \frac{x^2}{2}$ sa $x = \frac{\sigma_{n,i}^2 u^2}{2}$ i dobivamo

$$B_n \leq \sum_{i=1}^{M_n} \frac{\sigma_{n,i}^4 u^4}{8} \ (g.s.)$$

Da bismo to analizirali, koristimo svojstva ii., iii. te rastavimo:

$$\sigma_{n,i}^2 = \mathbb{E}[U_{n,i}^2 \mathbb{1}_{\{|U_{n,i}|\geq\epsilon\}}|\mathcal{F}_n] + \mathbb{E}[U_{n,i}^2 \mathbb{1}_{\{|U_{n,i}|<\epsilon\}}|\mathcal{F}_n] \leq \epsilon^2 + \mathbb{E}[U_{n,i}^2 \mathbb{1}_{\{|U_{n,i}|\geq\epsilon\}}|\mathcal{F}_n] \ (g.s.)$$

Onda vrijedi

$$\max_{1\leq i\leq M_n} \sigma_{n,i}^2 \leq \epsilon^2 + \sum_{i=1}^{M_n} \mathbb{E}[U_{n,i}^2 \mathbb{1}_{\{|U_{n,i}|\geq\epsilon\}}|\mathcal{F}_n] \ (g.s.)$$

gdje desna strana konvergira po vjerojatnosti u ϵ^2 zbog iii.. Opet po lemi A.2 imamo da

$$\max_{1\leq i\leq M_n} \sigma_{n,i}^2 \xrightarrow[n\to\infty]{\mathcal{P}} 0. \tag{A3}$$

Na kraju,

$$B_n \overset{g.s.}{\leq} \sum_{i=1}^{M_n} \frac{\sigma_{n,i}^4 u^4}{8} \overset{g.s.}{\leq} \frac{u^4}{8} \max_{1\leq i\leq M_n} \sigma_{n,i}^2 \sum_{i=1}^{M_n} \sigma_{n,i}^2 \xrightarrow[n\to\infty]{\mathcal{P}} 0$$

te primjenom leme A.2 dobivamo da $B_n \xrightarrow[n\to\infty]{\mathcal{P}} 0$ stoga vrijedi i (A1), što dovršava dokaz. □

Primijetimo da je iii. zapravo *Lindebergov uvjet*, dok je (A3) jedna verzija uvjeta *uniformne asimptotske zanemarivosti*.

Spomenimo sada lemu koja uz prijašnju propoziciju, daje teorem A.5 kojeg upotrebljavamo u dokazima asimptotske normalnosti. Dokaz i tvrdnja su isti kao (A.4) iz [4].

Lema A.4. *Neka je $\mathcal{G}$ σ-algebra te X slučajna varijabla takva da vrijedi $\mathbb{E}[X^2|\mathcal{G}] < \infty$. Tada za svaki realan $\epsilon > 0$ vrijedi*

$$4\mathbb{E}\left[X^2 \mathbb{1}_{\{|X|\geq 2\epsilon\}}|\mathcal{G}\right] \geq \mathbb{E}\left[(X - \mathbb{E}[X|\mathcal{G}])^2 \, \mathbb{1}_{\{|X-\mathbb{E}[X|\mathcal{G}]|\geq\epsilon\}}|\mathcal{G}\right] \ (g.s.)$$

Dokaz. Označimo $Z = \mathbb{E}[X|\mathcal{G}]$ i $Y = X - \mathbb{E}[X|\mathcal{G}]$. Rastavljamo na dva slučaja, $\{|Z|\geq \epsilon\}$ i $\{|Z|< \epsilon\}$, a to smijemo jer se oni nalaze u $\mathcal{G}$.

1. Promatramo događaj $\{|Z|< \epsilon\}$. Zbog nejednakosti koje vrijede za varijable unutar očekivanja i jer je Z $\mathcal{G}$-izmjeriv na događaju $\{|Z|< \epsilon\}$, redom vrijedi

$$\begin{aligned}
\mathbb{E}\left[Y^2 \mathbb{1}_{\{|Y|\geq 2\epsilon\}}|\mathcal{G}\right] &\leq 2\mathbb{E}\left[((Y+Z)^2 + Z^2)\mathbb{1}_{\{|Y+Z|\geq\epsilon\}}|\mathcal{G}\right] \\
&\leq 2\left(1 + \frac{Z^2}{\epsilon^2}\right)\mathbb{E}\left[(Y+Z)^2 \mathbb{1}_{\{|Y+Z|\geq\epsilon\}}|\mathcal{G}\right] \\
&\leq 4\mathbb{E}\left[(Y+Z)^2 \mathbb{1}_{\{|Y+Z|\geq\epsilon\}}|\mathcal{G}\right] \ (g.s.)
\end{aligned}$$

2. Jer je Z $\mathcal{G}$-izmjeriv i jer uvjetno očekivanje iščezava na Y, redom vrijedi $\mathbb{E}[ZY|\mathcal{G}] = Z\mathbb{E}[Y|\mathcal{G}] = 0$ ($g.s.$). Sada, na događaju $\{|Z| \geq \epsilon\}$, imamo

$$\begin{aligned}\mathbb{E}\left[Y^2 \mathbb{1}_{\{|Y|\geq 2\epsilon\}}|\mathcal{G}\right] &\leq \mathbb{E}\left[Y^2 + Z^2 - \epsilon^2|\mathcal{G}\right] \\ &= \mathbb{E}\left[(Y+Z)^2 - \epsilon^2|\mathcal{G}\right] \\ &\leq \mathbb{E}\left[(Y+Z)^2 \mathbb{1}_{\{|Y+Z|\geq\epsilon\}}|\mathcal{G}\right] \ (g.s.)\end{aligned}$$

□

Zamjenom $U_{n,i} = V_{n,i} - \mathbb{E}[V_{n,i}|\mathcal{G}]$ i primjenom propozicije A.3 te prethodne leme za svojstvo iii., lako dobivamo idući teorem.

Teorem A.5. *Neka je $\{V_{n,i}\}_{1\leq i\leq M_n}$ dvostruki niz koji je uvjetno nezavisan s obzirom na $\{\mathcal{F}_n\}_{n\geq 0}$, niz σ-algebri koje su podskupovi od $\mathcal{F}$. Neka redom vrijede i svojstva:*

i. Za svaki par n, i takav da $1 \leq i \leq M_n$ vrijedi $\mathbb{E}[V_{n,i}^2|\mathcal{F}_n] < \infty$ ($g.s.$).

ii. Postoji pozitivan realan broj σ takav da vrijedi

$$\sum_{i=1}^{M_n} \mathbb{E}[V_{n,i}^2|\mathcal{F}_n] - \mathbb{E}[V_{n,i}|\mathcal{F}_n]^2 \xrightarrow[n\to\infty]{\mathcal{P}} \sigma^2.$$

iii. Za svaki $\epsilon > 0$ vrijedi

$$\sum_{i=1}^{M_n} \mathbb{E}[V_{n,i}^2 \mathbb{1}_{\{|V_{n,i}|\geq\epsilon\}}|\mathcal{F}_n] \xrightarrow[n\to\infty]{\mathcal{P}} 0.$$

Tada za svaki $u \in \mathbb{R}$ vrijedi

$$\mathbb{E}[exp(\mathrm{i}u \sum_{i=1}^{M_n} V_{n,i} - \mathbb{E}[V_{n,i}|\mathcal{F}_n])|\mathcal{F}_n] \xrightarrow[n\to\infty]{\mathcal{P}} exp(-u^2\sigma^2/2). \tag{A4}$$

Prijašnji teorem kombiniramo s asimptotskom normalnosti za dobivanje asimptotske normalnosti suma koje nisu nužno nezavisne.

Lema A.6 (Kombinacija konvergencija)**.** *Neka je $\mathcal{F}_n$ niz σ-algebri. Neka je A_n niz slučajnih varijabli takvih da vrijedi asimptotska normalnost*

$$A_n \xrightarrow[n\to\infty]{\mathcal{D}} N(0, \sigma_A^2)$$

i tako da je za svaki n, A_n $\mathcal{F}_n$-izmjeriv. Neka je B_n niz slučajnih varijabli takvih da

$$\mathbb{E}[\exp(\mathrm{i}vB_n)|\mathcal{F}_n] \xrightarrow[n\to\infty]{\mathcal{P}} \exp(-v^2\sigma_B^2/2).$$

Tada vrijedi

$$\begin{pmatrix} A_n \\ B_n \end{pmatrix} \xrightarrow[n\to\infty]{\mathcal{D}} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_B^2 \end{pmatrix}\right).$$

Dokaz. Koristimo karakteristične funkcije za slučajne vektore, primijenjene na niz vektora $(A_n, B_n)_{n\geq 0}$. Za svaki par realnih brojeva u, v i svaki n, zbog svojstva uvjetnog očekivanja imamo

$$\begin{aligned}\mathbb{E}[exp(\mathrm{i}(uA_n + vB_n))] &= \mathbb{E}[\mathbb{E}[exp(\mathrm{i}(uA_n + vB_n))|\mathcal{F}_n]] \\ = (\text{zbog } \sigma(A_n) \subset \mathcal{F}_n) &= \mathbb{E}[exp(\mathrm{i}u\alpha_n A_n)\mathbb{E}[exp(\mathrm{i}vB_n)|\mathcal{F}_n]]. \end{aligned} \tag{A5}$$

Zbog asimptotske normalnosti A_n, pomoću slabe konvergencije konačnih mjera (analogon Teorema 13.13. [20]) na konvergenciju možemo primijeniti neprekidnu i ograničenu funkciju $\exp(iu\cdot)$, stoga vrijedi

$$\exp(iuA_n) \xrightarrow[n\to\infty]{\mathcal{D}} \exp(iuA)$$

gdje A ima distribuciju $N(0, \sigma_A^2)$. Po Slutskyevoj lemi A.7 o množenju onda vrijedi

$$\exp(iuA_n)\mathbb{E}[\exp(\mathrm{i}vB_n)|\mathcal{F}_n] \xrightarrow[n\to\infty]{\mathcal{D}} \exp(iuA - v^2\sigma_B^2/2).$$

Zato što postoji neprekidna ograničena funkcija koja je identiteta na skupu vrijednosti koje varijable u gornjoj konvergenciji mogu poprimiti, pomoću slabe konvergencije mjera možemo uzeti očekivanje s obje strane. Koristeći jednakost (A5) onda dobivamo

$$\mathbb{E}[exp(\mathrm{i}(uA_n + vB_n))] \xrightarrow[n\to\infty]{} \exp(-u^2\sigma_A^2/2 - v^2\sigma_B^2/2).$$

To je upravo karakteristična funkcija bivarijatne normalne razdiobe, i onda po teoremu neprekidnosti za slučajne vektore (Teorem 13.18. u [20]) slijedi tvrdnja leme. □

Iduća lema nam koristi za račune koji uključuju konvergenciju po distribuciji. Navodimo ju bez dokaza.

Lema A.7 (Slutskyev teorem [7]). *Neka su X_n, X, Y_n (realne) slučajne varijable takve da vrijedi $X_n \xrightarrow[n\to\infty]{\mathcal{D}} X$ i $Y_n \xrightarrow[n\to\infty]{\mathcal{P}} c$ gdje je c konstanta. Tada vrijedi $X_n + Y_n \xrightarrow[n\to\infty]{\mathcal{D}} X + c$, $X_nY_n \xrightarrow[n\to\infty]{\mathcal{D}} cX$ te uz uvjet $c \neq 0$ vrijedi i $X_n/Y_n \xrightarrow[n\to\infty]{\mathcal{D}} X/c$.*

Bibliografija

[1] P. Billingsley, *Probability and Measure*, Wiley Series in Probability and Mathematical Statistics, 1995.

[2] O. Cappé, *Ten years of HMMs*, (2001), http://neuro.bstu.by/ai/To-dom/My_research/Paper-0-again/For-courses/HMM/hmmbib%5B1%5D.pdf, posjećena 2020-01-07.

[3] O. Cappé, E. Moulines i T. Rydén, *Inference in Hidden Markov Models*, Springer Series in Statistics, 2005.

[4] R. Douc i E. Moulines, *Limit Theorems for Weighted Samples with Applications to Sequential Monte Carlo Methods*, The Annals of Statistics **36** (2008), 2344–2376, ISSN 0090-5364, http://doi.org/10.2307/25464714.

[5] A. Doucet i A.M. Johansen, *A tutorial on particle filtering and smoothing: fifteen years later*, in Handbook of Nonlinear Filtering, Eds. D. Crisan and B. Rozovsky, Cambridge University Press, 2009.

[6] R. F. Engle, *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation*, Econometrica **50** (1982), br. 4, 987–1007, ISSN 00129682, 14680262, http://www.jstor.org/stable/1912773.

[7] G.R. Grimmett i D.R. Stirzaker, *Probability and random processes*, sv. 80, Oxford university press, 2001.

[8] J. D. Hamilton, *A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle*, Econometrica **57** (1989), br. 2, 357–384, ISSN 00129682, 14680262, http://www.jstor.org/stable/1912559.

[9] E. Jacquier, N. G. Polson i P. Rossi, *Bayesian Analysis of Stochastic Volatility Models*, Journal of Business and Economic Statistics **12** (1994), br. 4, 371–89, https://EconPapers.repec.org/RePEc:bes:jnlbes:v:12:y:1994:i:4:p:371-89.

[10] R. E. Kalman i R. S. Bucy, *New Results in Linear Filtering and Prediction Theory*, Journal of Basic Engineering **83** (1961), br. 1, 95–108, ISSN 0021-9223, `https://doi.org/10.1115/1.3658902`.

[11] D. Koller i N. Friedman, *Probabilistic Graphical Models - Principles and Techniques*, The MIT Press, 2009.

[12] A. Kong, J. S. Liu i W. H. Wong, *Sequential Imputations and Bayesian Missing Data Problems*, Journal of the American Statistical Association **89** (1994), br. 425, 278–288, ISSN 01621459, `http://www.jstor.org/stable/2291224`.

[13] T. Koski, *Hidden Markov Models for Bioinformatics*, Springer, 2002.

[14] R. Langrock, B. J. Swihart, B. S. Caffo, N. M. Punjabi i C. M. Crainiceanu, *Combining hidden Markov models for comparing the dynamics of multiple sleep electroencephalograms*, Statistics in Medicine **32** (2013), br. 19, 3342–3356, `https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.5747`.

[15] J. S. Liu i R. Chen, *Blind Deconvolution via Sequential Imputations*, Journal of the American Statistical Association **90** (1995), br. 430, 567–576, ISSN 01621459, `http://www.jstor.org/stable/2291068`.

[16] S. Meyn i R. L. Tweedie, *Markov Chains and Stochastic Stability*, Cambridge University Press, 2009.

[17] L. R. Rabiner i B. H. Juang, *Fundamentals of speech recognition*, Prentice Hall signal processing series, 1993.

[18] C. P. Robert i G. Casella, *Monte Carlo Statistical Methods*, Springer Texts in Statistics, 2004.

[19] D. B. Rubin, *The Calculation of Posterior Distributions by Data Augmentation: Comment: A Noniterative Sampling/Importance Resampling Alternative to the Data Augmentation Algorithm for Creating a Few Imputations When Fractions of Missing Information Are Modest: The SIR Algorithm*, Journal of the American Statistical Association **82** (1987), br. 398, 543–546, ISSN 01621459, `http://www.jstor.org/stable/2289460`.

[20] N. Sarapa, *Teorija vjerojatnosti*, Školska knjiga, 2002.

Sažetak

U ovome radu prezentiramo problem izglađivanja na općenitim skrivenim Markovljevim modelima. Problem izglađivanja i srodni problemi sa sekvencijalnom strukturom distribucija su se rješavali Kalmanovim i proširenim Kalmanovi filtrom koji imaju snažne pretpostavke o normalnosti i linearnosti modela. Sekvencijalni Monte Carlo algoritmi su nastali iz potrebe da se takvi problemi riješe na općenitijoj klasi modela, i uz slabije pretpostavke. Prezentiramo i motiviramo nastanak algoritma sekvencijalnog uzorkovanja po važnosti s reuzorkovanjem koji služi kao prototip svih sekvencijalnih Monte Carlo metoda.

Prvo poglavlje motivira proučavanje skrivenih Markovljevih modela s primjerima i pokazuje kako oni pokrivaju vrlo široku klasu modela. Dajemo teoretsku podlogu Markovljevim lancima na općenitom prostoru stanja i formalno definiramo skrivene Markovljeve procese te problem izglađivanja.

Zatim u drugom poglavlju obrađujemo algoritam uzorkovanja po važnosti i njegova proširenja, sve s ciljem kasnije primjene na problemima sa sekvencijalnom strukturom poput izglađivanja. Obrađujemo samo-normalizirajuću varijantu i dodatak reuzorkovanja. Poopćujemo definiciju asimptotske normalnosti na dvostruke nizove i uvodimo alate koji služe za asimptotsku analizu uvjetno nezavisnih dvostrukih nizova, sve s ciljem dokazivanja da korak reuzorkovanja čuva asimptotsku normalnost.

Kao proširenje tehnike uzorkovanja po važnosti na skrivene Markovljeve modele, u trećem poglavlju opisujemo sekvencijalno uzorkovanje po važnosti i dajemo nekoliko načina na koje možemo implementirati instrumentalnu distribuciju u tom algoritmu. Pokazujemo prisutnost problema degeneracije u algoritmu te da ga dodavanje reuzorkovanja uspješno rješava. Na kraju, dajemo dovoljne uvjete i dokazujemo asimptotsku normalnost sekvencijalnog algoritma sa reuzorkovanjem, pritom se oslanjajući na dokaz čuvanja normalnosti koraka reuzorkovanja.

Summary

In this thesis we present the problem of smoothing on general state space hidden Markov models. Smoothing and other related problems with sequential structure have been solved by using the Kalman filter and the extended Kalman filter, which have strong assumptions on normality and linearity of models. Sequential Monte Carlo methods were developed with the goal of not needing such strong model assumptions. We present and motivate the development of an algorithm called sequential importance sampling with resampling, which serves as a prototype for all sequential Monte Carlo methods.

The first chapter motivates exploration of hidden Markov models with examples, and tries to argue that it is a very versatile class of models. We also give a theoretical background on general state space Markov chains and formally define hidden Markov processes and the smoothing problem.

Then in the second chapter, we present importance sampling and some of it's extensions, with the goal of applying it on problems with a sequential structure, as in the case of smoothing in hidden Markov models. The extensions we present are self-normalizing importance sampling and importance sampling with resampling. For proving asymptotic normality of estimators that those algorithms give, we show generalizations of asymptotic normality for triangular arrays and tools which we use for the analysis of conditionally independent triangular arrays.

In the third chapter we present sequential importance sampling and describe several possible ways of implementing instrumental kernels. That algorithm is a generalization of importance sampling which can be used for smoothing in hidden Markov models. We describe how weight degeneracy occurs in sequential importance sampling and present reasons why resampling is a necessary addition to sequential importance sampling. At the end, we present a proof of asymptotic normality of sequential sampling with resampling.

Životopis

Rodio sam se u Koprivnici, 7. travnja 1995. Svake godine od 7. razreda osnovne škole sudjelujem na Državnim natjecanjima iz matematike. Najuspješniji natjecateljski rezultati u srednjoj školi bili su brončana medalja na Međunarodnoj matematičkoj olimpijadi i prolaz na Hrvatsku informatičku olimpijadu.

Kroz studij na Prirodoslovnom-matematičkom fakultetu Sveučilišta u Zagrebu držao sam demonstrature iz Programiranja *I* i *II*, Algebarskih struktura i Matematičke logike. U *V*. gimnaziji sam držao redovnu grupu dodatne matematike kroz četiri godine i član sam udruge *Mladi nadareni matematičari - Marin Getaldić* kroz koju sam sudjelovao u pripremi srednjoškolskih učenika za natjecanja iz matematike.

Obavio sam nekoliko programerskih praksi u Zagrebu i inozemstvu, i istraživačku praksu na Australskom nacionalnom sveučilištu.