

Bayesiánská analýza

Vybrané konvergenční diagnostiky

Konvergenční diagnostiky - přehled

- 1 Konvergenční diagnostiky - úvod
- 2 Gelman a Rubin (1992)
- 3 Brooks a Gelman (1998)
- 4 Geweke (1992)
- 5 Yu a Mykland (1994)

- 1 Konvergenční diagnostiky - úvod
- 2 Gelman a Rubin (1992)
- 3 Brooks a Gelman (1998)
- 4 Geweke (1992)
- 5 Yu a Mykland (1994)

Konvergence MCMC

- Cowles a Carlin (1996) + původní články .
- Problém, kdy ukončit MCMC algoritmus.
- Otázka - kdy můžeme předpokládat reprezentativnost vzorků z hlediska stacionarity rozdělení MC řetězce.
- Obecnější označení konvergence než je obvyklé pro iterativní postupy.
- Důvod – výsledkem algoritmu není číslo či rozdělení, ale jen výběr z rozdělení.
- Markovovský charakter algoritmu = vzorky jsou navzájem korelovány
⇒ zpomalení algoritmu z hlediska efektivity výběrů z celého stacionárního rozdělení ⇒ negativní ovlivnění Monte Carlo odhadu variance odhadu charakteristik modelu na základě výstupu.

Konvergence MCMC - řešení problému I

- Dvě oblasti řešení.
- 1. teoretická:
 - analýza Markovského jádra přechodu řetězce \Rightarrow predeterminace počtu iterací zajišťujících konvergenci v rámci definované tolerance od skutečného rozdělení;
 - sofistikovaná matematika a pracné výpočty pro každý model (modelově specifické);
 - mnohdy získané meze dosti volné (o několik řádu přesahují počet iterací vyhovujících v praktických aplikacích).

Konvergence MCMC - řešení problému II

- 2. aplikace diagnostických nástrojů:
 - aplikace na výstup algoritmu;
 - první pokusy – porovnání empirických rozdělení výstupů na základě po sobě jdoucích iterací (konvergence na základě „zanedbatelnosti“ rozdílu dvou po sobě jdoucích iterací);
 - \Rightarrow generátory využívající velký počet paralelních, nezávislých řetězců k získání odhadu jednoduchých momentů, kvantilů a hustot;
 - hojně užívaná diagnostika „thick felt-tip pen test“ (Gelfand a Smith (1960)) – konvergence v případě, kdy odhady hustot vzdálené od sebe dostatečně daleko (dostatečný počet běhuů), aby byly nezávislé (např. 5 iterací) se lišily graficky méně než tloušťka onoho fixu;
 - problém – nevyužití předkonvergenčních vzorků (masivní paralelní přístup) + identifikace předčasné konvergence pro „pomalu se mísící“ vzorkovače.

Konvergence MCMC - řešení problému III

- Měření vzdálenosti výběrového rozdělení ve dvou iteracích, ne rozdíl příslušného rozdělení od skutečného (teoretického).
- Stacionární rozdělení vždy neznámé (v praxi) \Rightarrow problém jakékoliv konvergenční diagnostiky.
- \Rightarrow dle řady teoretiků všechny tradiční diagnostiky zcela špatné.
- Stejný závěr i v jiných oblastech využívajících MCMC metody (fyzika, operační výzkum).
- Mnoho statistiků tyto diagnostiky využívá – „slabá“ diagnostika vždy lepší než žádná.

Konvergenční diagnostiky - kategorizace I

- 1. Diagnostika konvergence kvantit pro jednorozměrný případ (jedna proměnná) nebo pro celou sdruženou hustotu.
- 2. Výsledky kvantitativní nebo kvalitativní (např. grafické).
- 3. Diagnostika analyzuje problém zkreslení (rozdíl odhadu kvantit, které nás zajímají na základě iterací od skutečných hodnot cílového rozdělení) a rozptylu výsledných odhadů charakteristik (tzn. kvalita odhadů) stacionárního rozdělení (téměř všechny diagnostiky usilují o odhalení zkreslení pocházející z výběrů, které nereprezentuje příslušné rozdělení).
- 4. Metoda využívá jediný řetězec nebo několik paralelních řetězců.

Konvergenční diagnostiky - kategorizace II

- 5. Diagnostiky říkající uživateli kolik (autokorelovaných) výběrů je třeba generovat k odhadům, jejichž variabilita je dostatečně malá (kvůli věrohodnosti jejich přesnosti).
- 6. Diagnostiky modelově specifické, jen pro Gibbsův vzorkovač nebo pro jakýkoliv MCMC algoritmus.
- 7. Snadnost použití – existence obecného kódu; obecný kód je třeba napsat a aplikovat na MCMC výstup jakéhokoliv problému; potřeba problémově-specifického kódu; potřeba analytického odvození a problémově-specifického kódu.

- 1 Konvergenční diagnostiky - úvod
- 2 Gelman a Rubin (1992)**
- 3 Brooks a Gelman (1998)
- 4 Geweke (1992)
- 5 Yu a Mykland (1994)

Charakteristika

- Dvoukrokový algoritmus.
- Původně vytvořený pro Gibbsův vzorkovač.
- Fakticky uplatnitelný pro výstup jakéhokoliv MCMC algoritmu.
- Pro jednotlivé parametry (funkce parametrů).

První krok

- Před algoritmem generování vzorků.
- Zisk rozptýlených počátečních hodnot (z cílového rozdělení).
 - hledají se módy (K modů) cílového rozdělení \rightarrow pro každý z nich matice druhých derivací \rightarrow oblasti nejvyšších posteriorních hustot jsou aproximovány mixem K vícerozměrných normálních rozdělení (se střední hodnotou μ_k a varianční maticí Σ_k odpovídající matici druhých derivací v každém modu);
 - vzorky z rozptýleného rozdělení: výběr z mixu normálních rozdělení a každý vzorek dělen skalární náhodnou veličinou (obvykle volba z χ^2_η , dělená η) \rightarrow nové rozdělení mixem vícerozměrných t -rozdělení; konzervativní výběr je $\eta = 1$ (Cauchyho rozdělení).
- 10 počátečních hodnot (unimodálního rozdělení), více hodnot (multimodální případ).

Druhý krok

- Pro každou skalární kvantitu na základě vygenerovaných vzorků.
- Vzorky délky $2n$; n iterací využíváme:
 - k odhadu cílového rozdělení požadované skalární kvantity (jakožto konzervativního t -rozdělení);
 - k odhadu *scale* parametru zahrnujícího rozptyl mezi vzorky a v rámci vzorku;
 - ke zjištění konvergence na základě odhadu faktoru popisujícího jak by se mohl *scale* parametr zmenšit, pokud by délka vzorku šla k nekonečnu.

Odhad faktoru

$$\sqrt{\hat{R}} = \sqrt{\left(\frac{n-1}{n} + \frac{m+1}{m+n} \frac{B}{W} \right) \frac{df}{df-2}}$$

- B – rozptyl mezi středními hodnotami m paralelních řetězců;
- W – průměr m variancí (rozptylů) v rámci řetězců;
- df – počet stupňů volnosti aproximativní t -hustoty;
- na počátku je B mnohem větší než W pro pomalu se mísící vzorkovače (řetězce startují z rozptýlených pozici vzhledem k cílové hustotě).

Test konvergence

- Návrh dodatečných iterací paralelních řetězců dokud „faktor zmenšení“ pro všechny kvantily nebude blízko 1.
- Analýza na základě posledních n iterací všech řetězců.
- Bodové odhady a .975 kvantily „faktoru zmenšení“.
- Empirické kvantily vzorkovaných kvantit z celých vzorků.
- Odhadované kvantily Studentova t -rozdělení.

Problémy a diskuse

- Hledání počátečních rozptýlených startovních hodnot cílového rozdělení.
- Normální aproximace pro diagnostiku konvergence (diskutabilní).
- „Obhájci“ dlouhých řetězců – neefektivní generovat řadu řetězců a vyhození jejich významné části \times srovnání běhu jednoho řetězce o délce 10000 (pravděpodobnější konvergence) a 10 nezávislých řetězců po 1000.
- Rozšíření pro konvergenci sdružené hustoty – aplikace procedury na -2 krát logaritmus posteriorní hustoty.

- 1 Konvergenční diagnostiky - úvod
- 2 Gelman a Rubin (1992)
- 3 Brooks a Gelman (1998)**
- 4 Geweke (1992)
- 5 Yu a Mykland (1994)

Charakteristika

- Zobecnění a korekce Gelmana a Rubina (1992).
- Rozšíření diagnostiky Gelmana a Rubina (1992) pro vícerozměrný případ.

Scale Reduction Factor (SRF) I

- Skalární náhodná statistika ψ , mající střední hodnotu μ a rozptyl σ^2 v rámci cílového rozdělení.
- Nevychýlený estimátor $\hat{\mu}$ pro μ .
- ψ_{jt} označuje t -tou z n iterací statistiky ψ v řetězci j ; vezmeme $\hat{\mu} = \bar{\psi}_{..}$ a počítáme mezisekvenční rozptyl B/n a rozptyl uvnitř sekvence W :

$$B/n = \frac{1}{m-1} \sum_{j=1}^m (\bar{\psi}_{j.} - \bar{\psi}_{..})^2$$

$$W = \frac{1}{m(n-1)} \sum_{j=1}^m \sum_{t=1}^n (\psi_{jt} - \bar{\psi}_{j.})^2$$

SRF II

- Odhad σ^2 jako vážený průměr B a W .

$$\hat{\sigma}_+^2 = \frac{n-1}{n}W + \frac{B}{n}$$

- nevychýlený odhad skutečné variance σ^2 pokud výchozí hodnoty řad jsou výběry z cílového rozdělení;
- nahodnocuje σ^2 pokud je počáteční rozdělení adekvátně přerозptýlené (pokud přerозptýlenost není, potom $\hat{\sigma}_+^2$ může být příliš nízké \Rightarrow chybná diagnostika konvergence).
- při výběrové variabilitě odhadu $\hat{\mu}$ je posteriorní odhad variance $\hat{V} = \hat{\sigma}_+^2 + B/(mn)$.

SRF a PSRF

- Porovnání analýzy napříč řetězců a v rámci řetězců – poměr rozptylů:

$$R = \frac{\hat{V}}{\sigma^2}$$

- Tzv. scale reduction factor (SRF) – označení spíše pro $\sqrt{\hat{R}}$ \Rightarrow obvykle práce s variance reduction factor (bez odmocniny).
- Jmenovatel v R není znám, třeba odhadnout z dat; získáváme nahodnocený R díky podhodnocení σ^2 prostřednictvím W .

$$\hat{R} = \frac{\hat{V}}{W} = \frac{m+1}{m} \frac{\hat{\sigma}_+^2}{W} - \frac{n-1}{mn}$$

- Tzv. potential scale reduction factor (PSRF) = konvergenční diagnostika-

Intepretace PSRF

- Velké \hat{R} znamená, že odhad rozptylu $\hat{\sigma}^2$ lze snížit více simulacemi nebo další simulace zvýší W , neboť simulace neprošly zcela cílové rozdělení.
- PSRF blízky jedné naznačuje, že každý z m řetězců n simulací se přiblížil cílovému rozdělení.

Korigovaný SRF

- Doladění PSRF pro výběrový rozptyl.
- Odhad stupňů volnosti d pro Studentovu t -aproximaci cílového rozdělení.
- Odhad metodou momentů: $d \approx 2\hat{V}/\widehat{\text{var}}(\hat{V})$.
- Gelman a Rubin (1992) – nekorektně použitý faktor $d/d - 2$ - reprezentuje poměr rozptylů t_d a normálního rozdělení (problémy nekonečnosti či negativních hodnot pro příliš pomalou konvergenci, kdy $d < 2$).
- Korektní faktor $(d + 3)/(d + 1)$; vychází z vyhodnocení Fisherovy informace pro t_d rozdělení, koriguje \hat{V} o stupně volnosti, na kterých je založen.
- Zanedbatelný význam korekce (při konvergenci je d obrovská).

$$\hat{R}_c = \frac{d + 3}{d + 1} \hat{R} = \frac{d + 3}{d + 1} \frac{\hat{V}}{W}.$$

Iterující grafický přístup – motivace

- Gelman a Rubin (1992) – pro konvergenci nutné \widehat{R}_c blízko jedné pro všechny parametry.
- Tento postup ignoruje celou informaci dostupnou v simulacích; pro konvergenci nutné podmínky:
 - 1 Mix (kompozice) rozptylu sekvencí V by se měl stabilizovat jako funkce n (před konvergencí předpokládáme $\hat{\sigma}_+^2$ klesá s n a roste tehdy pokud řady prozkoumávají novou oblast parametrického prostoru \Rightarrow původní řady nebyly dostatečně přerozptýlené pro konkrétní skalární statistiku, která je monitorována)
 - 2 Rozptyl v rámci řady W by se měl stabilizovat jako funkce n (před konvergencí očekáváme, že W je menší než V).
 - 3 \widehat{R}_c by se měl blížit jedné.
- Pozorování samotného \widehat{R}_c předpokládá jen třetí podmínku.

Iterující grafický přístup

- Rozdělení m sekvencí na části délky b .
- Výpočet $V(k)$, $W(k)$ a \hat{R}_c založený na druhé polovině pozorování řad délky $2kb$ pro $k = 1, \dots, n/b$ (pro dostatečně velké n).
- Vyhazujeme první polovinu každého vzorku (kompromis) – více vyhozených vzorků = diagnostika založená na příliš malé části vzorku; méně vyhozených vzorků = vzorek pozorování použitý v každé iteraci by se měnil velmi pozvolnu (\hat{R} by zůstával vysoký příliš dlouho) \Rightarrow diagnostika rozpozná konvergenci později, než je žádoucí.
- Vykreslení konvergence vzhledem ke k nebo $2kb$.
- Volba b : malé b zvyšuje výpočetní náročnost, velké b dává malou dodatečnou informaci ve smyslu pozorování konvergence \Rightarrow ze zkušenosti $b \approx n/20$.

Iterující grafický přístup – další diagnostiky

- Vykreslení nejen $\widehat{R}_c(k)$ vzhledem ke k .
- Další užitečné diagnostiky: vykreslení dvou scale faktorů $V^{1/2}(k)$ a $W^{1/2}(k)$ jako funkci k do jednoho obrázku (mocnina $\frac{1}{2}$ z důvodu přímé interpretovatelnosti faktorů jako „scale“ faktorů).
- Aproximativní konvergence – obě křivky se stabilizují (podmínky 1 a 2), a to na stejné hodnotě (podmínka 3)

Zobecnění pro jednorozměrný případ

- Omezení původní diagnostiky = předpoklad normality marginálního rozdělení každé skalární kvantitativy ψ (normalitu předpokládámě explicitně využitím korekčního faktoru $(d + 3)/(d + 1)$ a implicitně porovnáním mísících se řetězců pozorování středních hodnot a rozptylů).
- MCMC metody – využití pro vysoce nenormální a multimodální hustoty \Rightarrow definice rodiny dalších PSRF s vlastností, že $\hat{R} \rightarrow 1$ v rámci konvergence + nevyžadován předpoklad normality.
- Alternativní interpretace diagnostiky \hat{R} – poměr (čtverců) intervalových délek (oproti poměru rozptylů).
- Pro dostatečně dlouhé sekvence - hledáme $100(1 - \alpha)\%$ interval pro nějaký parametr $\psi \rightarrow$ konstrukce \hat{R} na základě posledních n z $2n$ iterací.

Intervalový PSRF

- Postup konstrukce:
 1. z každého řetězce vezmeme empirický $100(1 - \alpha)\%$ interval (tzn. $100\frac{\alpha}{2}\%$ a $100(1 - \alpha/2)\%$ bod n nasimulovaných výběrů) \Rightarrow vytvoříme m intervalových odhadů v rámci jednotlivých řetězců $\rightarrow IL_i$
 2. z celé množiny mn pozorování (ze všech řetězců) spočítáme empirický $100(1 - \alpha)\%$ interval \Rightarrow odhad délky intervalu úplné řady TIL .
 3. spočítáme \hat{R} definované jako

$$\hat{R}_{interval} = \frac{TIL}{\frac{1}{m} \sum_{i=1}^m IL_i}$$

- Jednoduchá metoda, $\hat{R}_{interval}$ je PSRF (založený na délce empirických intervalů, jakožto nositel informace o konvergenci), podobně jako na normalitě založená diagnostika se blíží jedné v případě konvergence řetězce.

Pravděpodobnost pokrytí

- Postup konstrukce:
 - 1. z každého řetězce vezmeme krajní body $100(1 - \alpha)\%$ interval pro odhad intervalu v rámci řetězce
 - 2. pro každý z intervalů odhadneme empirickou pravděpodobnost pokrytí intervalu na základě kombinace všech m řetězců, tj. počítáme podíl pozorování ze všech m sekvencí spadajících do příslušného intervalu
 - 3. spočítáme průměrnou empirickou pravděpodobnost pokrytí m intervalů; v konvergenci bude rovna nominální pravděpodobnosti pokrytí $1 - \alpha$.

Neintervalové přístupy

- Předpokládáme-li existenci vyšších momentů, které nás zajímají → nahrazení čitatele a jmenovatele ve vztahu pro původní PSRF empirickými odhady centrovaných momentů s -tého řádu ze všech řetězců dohromady respektive průměrný s -tý moment počítaný pro každou sekvenci:

$$\hat{R}_s = \frac{\frac{1}{mn-1} \sum_{j=1}^m \sum_{t=1}^n |\psi_{jt} - \bar{\psi}_{..}|^s}{\frac{1}{m(n-1)} \sum_{j=1}^m \sum_{t=1}^n |\psi_{jt} - \bar{\psi}_j|^s}$$

Grafické zobrazení

- Odpovídající grafické znázornění: souběžné vykreslení čitatele a jmenovatele $\hat{R}_{interval}$ nebo \hat{R}_s pro sekvence délky $2kb$, $k = 1, \dots, n/b$.
- Měřítka momentu druhého řádu \hat{R}_2 podobné původně definovanému R .
- Zanedbání korekčního faktoru (nepředpokládáme normalitu + zanedbatelný vliv na rozhodnutí o dosažení konvergence).
- Pokud $\hat{R}_c < 1$ měli bychom mít dost pozorování (tudíž výběrová variabilita zanedbatelná).
- V případě zanedbatelného vlivu: řetězce nekonvergovaly a $\hat{R} = \hat{V}/W$ je tak jak tak vysoké.
- V případě nutnosti lze korekční faktor vhodně přepočítat (myšlenka aplikace semi-parametrického bootstrapu).

Vícerozměrné rozšíření

- Řada možností pro rozšíření.
- Přímá analogie jednorozměrného přístupu: ψ je vektor parametrů, vzorek $\psi_{jt}^{(i)}$ označuje i -tý prvek vektoru parametrů v řetězci j v čase t
 \Rightarrow odhad posteriorní variančně-kovarianční matice

$$\hat{V} = \frac{n-1}{n}W + \left(1 + \frac{1}{m}\right)B/n$$

$$W = \frac{1}{m(n-1)} \sum_{j=1}^m \sum_{t=1}^n (\psi_{jt} - \bar{\psi}_{j\cdot})(\psi_{jt} - \bar{\psi}_{j\cdot})'$$

$$B/n = \frac{1}{m-1} \sum_{j=1}^m (\bar{\psi}_{j\cdot} - \bar{\psi}_{\cdot\cdot})(\bar{\psi}_{j\cdot} - \bar{\psi}_{\cdot\cdot})'$$

- B je p -rozměrný odhad mezisekvenční kovarianční matice, W je p -rozměrný odhad kovarianční matice v rámci sekvence
- sledujeme \hat{V} a W pro determinaci konvergence v rámci vhodně zvoleného měřítka vzdálenosti.

Skalární měřítko vzdálenosti

- Statistika maximálního kořene: maximum SRF nějaké lineární projekce ψ daná jako

$$\widehat{R}^P = \max_{\alpha} \frac{\alpha' \widehat{V} \alpha}{\alpha' W \alpha}$$

- tzv. multivariate PSRF (MPSRF) – lze spočítat níže uvedeným způsobem.
- Pro nesingulární, pozitivně definitní a symetrické matice M a N platí:

$$\max_{\alpha} \frac{\alpha' M \alpha}{\alpha' N \alpha} = \lambda$$

- kde λ je největší vlastní číslo pozitivně definitní matice $N^{-1}M$

$$\widehat{R}^P = \frac{n-1}{n} + \left(\frac{m+1}{m} \right) \lambda_1,$$

- kde λ_1 je největší vlastní číslo symetrické, pozitivně definitní matice $W^{-1}B/n$.

MPSRF - vyhodnocení I

- Za předpokladu stejných průměrů mezi sekvencemi $\lambda_1 \rightarrow 0$ a tudíž \hat{R}^P se blíží jedné pro dostatečně velké n .
- Nelze spočítat v případě, že W je singularitní; pokud W a B jsou singularitní, potom problém (dva nebo více parametrů mohou být silně korelovány).
- Pokud jen W singularitní – problém ve vzorkovači (jedna nebo více proměnných se neposunuje v rámci daného počtu iterací).
- V případě nespočítatelnosti \hat{R}^P (singularita W) – cenná informace na základě determinantů matic B a W .

MPSRF - vyhodnocení II

- Důvod pro výběr \hat{R}^p z jiných zobecnění SRF: lze využít jako aproximaci horní meze maximální jednorozměrné \hat{R} statistiky přes všech p proměnných.
- Pokud pro $k = 1, \dots, p$ je $\hat{R}(k)$ jednorozměrná PSRF aplikovaná na k -tý prvek vektoru θ a \hat{R}^{max} označuje maximální hodnotu z $\hat{R}(k)$ pro $k = 1, \dots, p$, potom

$$\hat{R}^{max} \leq \hat{R}^p$$

- Sledování MPSRF vede k pozdější diagnostice konvergence než naznačují obrázky jednotlivých PSRF (tento rozdíl je měřítkem chybějícího propojení mezi skalárními funkcionaly, které sledujeme).

- 1 Konvergenční diagnostiky - úvod
- 2 Gelman a Rubin (1992)
- 3 Brooks a Gelman (1998)
- 4 Geweke (1992)**
- 5 Yu a Mykland (1994)

Charakteristika

- Metody spektrální analýzy.
- Původně pro konvergenci Gibbsova vzorkovače.

Princip

- Zajímá nás střední hodnota nějaké funkce g simulovaných parametrů θ .
- V rámci každé iterace získáme $g(\theta^{(j)})$.
- Výsledná řada - časová řada.
- Gewekova metoda - předpoklad, že podstata MCMC procesu a funkce g implikuje existenci spektrální hustoty $S_g(\omega)$ pro tuto časovou řadu, která nemá nespojitosti v nulové frekvenci.

Konvergenční diagnostika I

- Za předchozího předpokladu - estimátor $E[g(\theta)]$ založený na n iteracích

$$\bar{g}_n = \frac{\sum_{i=1}^n g(\theta^{(i)})}{n}$$

- Jeho asymptotická variance je $S_g(0)/n$.
- Odmocnina tohoto rozptylu pro odhad standardní chyby střední hodnoty ("numeric standard error" – NSE).

Konvergenční diagnostika II

- Gewekova diagnostika (CD) po n iteracích:
 - rozdíl mezi střední hodnotou (průměrem) \bar{g}_n^A - na základě prvních n_A iterací a \bar{g}_n^B - na základě posledních n_B ;
 - dělení asymptotickou standardními chybami (vypočítané na základě spektrální hustoty) obou částí.
- Pokud n_A/n a n_B/n jsou pevná čísla a $n_A + n_B < n \Rightarrow$ dle centrální limitní věty rozdělení této diagnostiky se blíží standardizovanému normálnímu rozdělení (pro n jdoucí k nekonečnu).

Praktická aplikace

- $n_A = .1n$ a $n_B = .5n$.
- Kovergence pro $|CD| < 2$ (95% interval spolehlivosti).
- Lze využít pro zjištění kolik výchozích iterací vyhodit, poté musí běžet dostatečný počet iterací pro získání požadované přesnosti (NSE).
- Rozšíření pro konvergenci sdružené hustoty – aplikace procedury pro $g(\theta)$ rovnému -2 krát logaritmus (jádra) posteriorní hustoty.

Problémy a hodnocení

- Gewekova metoda se snaží řešit problém zkreslení i variance (v odhadu požadovaných charakteristik).
- Citlivost na specifikaci spektrálního okna (odhad spektrální hustoty).
- Kvantitativní diagnostika \times Geweke nspecifikuje jak proceduru aplikovat (ponecháno na zkušenostech a subjektivním hodnocení výzkumníka).

- 1 Konvergenční diagnostiky - úvod
- 2 Gelman a Rubin (1992)
- 3 Brooks a Gelman (1998)
- 4 Geweke (1992)
- 5 Yu a Mykland (1994)**

Charakteristika

- Grafická procedura založená na vykreslení CUMSUM trajektorie (kumulativní součty).
- Aplikovatelnosti na jednorozměrnou sumarizující statistiku (jakou je jeden parametr) z jednoho řetězce.
- Aplikovatelnost pro jakýkoliv MCMC algoritmus.

Diagnostika - postup

- Nejprve třeba použít metodu pro determinaci počtu vyhozených iterací n_0 - např. postupné vykreslení souhrnné statistiky iterací za iterací.
- Konstrukce CUSUM trajektorie pro iterace $n_0 + 1$ až n (poslední iterace).
- $T(X)$ - vybraná souhrnná statistika \Rightarrow odhad střední hodnoty na zbytku iterací:

$$\hat{\mu} = \frac{1}{n - n_0} \sum_{j=n_0+1}^n T(X^{(j)})$$

- Pozorovaný kumulativní (částečný) součet:

$$\hat{S}_t = \sum_{j=n_0+1}^t [T(X^{(j)}) - \hat{\mu}], \quad t = n_0 + 1, \dots, n$$

Diagnostika - CUSUM graf

- Obrázek vykreslení $\{\hat{S}_t\}$ vzhledem k $t = n_0 + 1, \dots, n$ a spojení jednotlivých bodů.
- Graf vždy začíná a končí v nule.
- Čím pomaleji se mísící vzorky \rightarrow graf tím hladší a více se vzdalující od nuly.
- Opačná analogie.
- Srovnání s benchmarkem = CUSUM trajektorie pro IID proměnné z normálního rozdělení se střední hodnotou a rozptylem odpovídajícím výběrové střední hodnotě a rozptylu MCMC vzorků.

Shrnutí a hodnocení

- Odstranění potřeby jiné informace než obsažené v řetězci.
- Podobně jako jiné metody - selhání pokud jedny oblasti výběrového prostoru jsou pomaleji se mísící než jiné.
- Není to zcela samostatná metoda (potřeba metody pro zjištění počtu počátečních vzorků k vyhození).
- Užitečná pro identifikaci vzorkovačů, které jsou natolik pomalu se mísící, že je potřeba nalézt alternativní algoritmus nebo parametrizaci pro projití celého parametrického prostoru v rozumném počtu iterací.
- Nepřímo pokrývá otázku variance a zkreslení odhadu (odhaduje závislost mezi iteracemi).