

Moderní metody počítačové fyziky II (NEVF161)

Štěpán Roučka

MFF UK,
2020

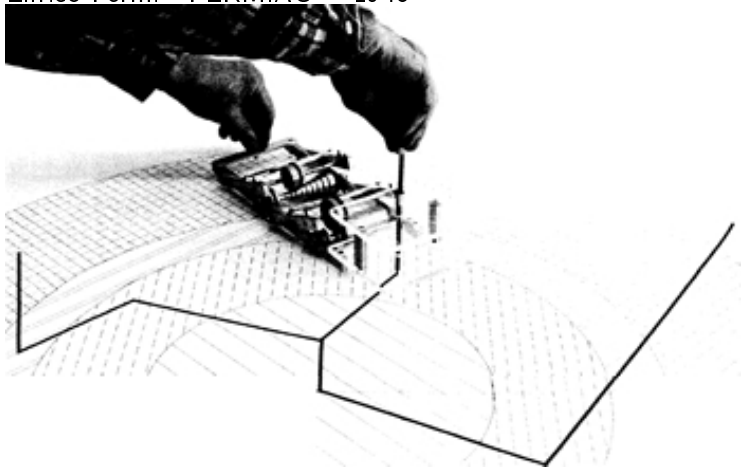


Základy metody Monte Carlo



Počátky metody Monte Carlo

Enrico Fermi - FERMIAC \sim 1948



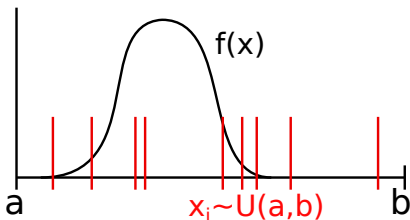
Monte Carlo (MC) - příklady

- ▶ Princip MC – vyhodnocování parametrů za využití náhodného vzorkování.

Značení

- ▶ $x_i \sim X$ náhodně vybraná proměnná x_i z rozdělení X
- ▶ $U(a, b)$ rovnoměrné rozdělení na intervalu $[a, b]$
- ▶ $\mathcal{N}(\mu, \sigma^2)$ Normální rozdělení

Příklad: střední hodnota funkce



Skutečná hodnota

$$E_{U(a,b)}[f(x)] = \frac{1}{b-a} \int_a^b f(x) dx$$

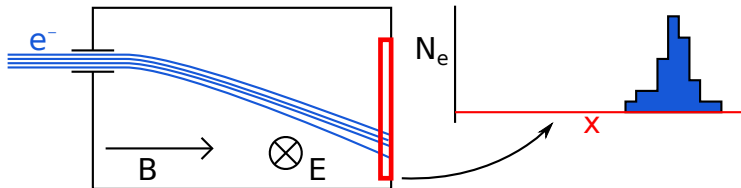
Odhad z n vzorků

$$\bar{f}_n = \frac{1}{n} \sum_{i=1}^n f(x_i)$$



MC - jiný příklad

- ▶ Simulace trochoidálního elektronového spektrometru



- ▶ Každý vzorek je užitečný pro určení stř. hodnoty (při vhodné počáteční hodnotě)
- ▶ Příklad důležitostního vzorkování



Odhad chyby MC

Zkoumaná veličina

$$E_p[f(x)] = \int f(x)p(x)dx$$

MC aproximace (konzistentní, nevychýlená...)

$$\bar{f}_n = \frac{1}{n} \sum_{i=1}^n f(x_i) \quad x_i \sim p$$

Rozptyl MC aproximace

$$\text{Var}[\bar{f}_n] = \text{Var} \left[\frac{1}{n} \sum_{i=1}^n f(x_i) \right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[f(x_i)] = \frac{1}{n} \text{Var}[f(x)]$$



Důležitostní vzorkování - Snížení chyby MC

Je vhodné snížit varianci vzorkované veličiny $f(x)$, neboť

$$\text{Var}[\bar{f}_n] = \frac{1}{n} \text{Var}[f(x)]$$

Lze dosáhnout volbou jiné výchozí distribuce než p :

$$\mathbb{E}_p[f(x)] = \int f(x)p(x) \frac{h(x)}{h(x)} dx = \mathbb{E}_h \left[f(x) \frac{p(x)}{h(x)} \right]$$

Pokud $p = U$ (rovnoměrné rozdělení), tak h by mělo splňovat

- ▶ $h \propto |f(x)|$
- ▶ $f(x) \neq 0 \Rightarrow h > 0$
- ▶ Asymptoticky nesmí $h \ll |f(x)|$

Jak generovat $x \sim f$?



MCMC - Monte Carlo Markov Chains

Doposud: Ordinary Monte Carlo (OMC) – přímo generované *nezávislé* x_i

Nyní - Markovovské řetězce - X_1, X_2, X_3, \dots

- ▶ Nemají paměť: $\Pr(X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_1 = x_1) = \Pr(X_{n+1} = x_{n+1} | X_n = x_n)$
- ▶ Určené počátečním rozdělením $\Pr(X_1 = x_1)$ a maticí přechodu $\Pr(X_{n+1} = x_{n+1} | X_n = x_n)$
- ▶ Značíme $\Pr(X_{n+1} = x' | X_n = x) = P_n(x \rightarrow x')$

statické

- ▶ $P_n(x \rightarrow x')$ nezávisí na n
- ▶ X_n vzorkuje rovnovážné rozdělení

dynamické

- ▶ $P_n(x \rightarrow x')$ může záviset na n
- ▶ X_n závisí na n - studium dynamiky



Statické MCMC



Statické MCMC

Často je obtížné vytvořit generátor n.p. ze zkoumaného rozdělení Π
Je jednodušší vytvořit matici přechodu, která zachovává Π
Rozdělení X_n potom konverguje k Π

Podmínky:

- ▶ Stacionarita $\Pi(x)P(x \rightarrow x') = \Pi(x')P(x' \rightarrow x)$ (detailní rovnováha, mikroreverzibilita...)
- ▶ Ergodicita - aperiodický, rekurentní, ireducibilní (propojený)



Metropolisův Hastingsův algoritmus I

Jak vytvořit matici přechodu, která zachovává Π ?

Podmínku stacionarity lze zapsat jako

$$\frac{P(x \rightarrow x')}{P(x' \rightarrow x)} = \frac{\Pi(x')}{\Pi(x)}$$

Definujeme g a A jako

$$P(x \rightarrow x') = \underbrace{g(x \rightarrow x')}_{\text{proposal}} \underbrace{A(x \rightarrow x')}_{\text{acceptance}}$$

Podmínka stacionarity potom je

$$\frac{A(x \rightarrow x')}{A(x' \rightarrow x)} = \frac{\Pi(x') g(x \rightarrow x')}{\Pi(x) g(x' \rightarrow x)}$$

$r(x, x')$ -Hastingsův poměr

A lze ji splnit nezávisle na g , pokud zvolíme

$$A(x \rightarrow x') = \min\{1; r(x, x')\}$$



Metropolisův Hastingsův algoritmus II

Jeden krok MH algoritmu potom je

- ▶ Generujeme x' z $g(x \rightarrow x')$
- ▶ Přejdeme do stavu x' s pravděpodobností $A(x \rightarrow x')$

Výhoda MH algoritmu: nepotřebujeme znát normované $\Pi(x)$, stačí poměr $\Pi(x')/\Pi(x)$

Příklad: Boltzmannovo rozdělení $\Pi(x) = \frac{1}{Z} \exp\left(-\frac{E(x)}{kT}\right)$

$Z = \sum_x \exp\left(-\frac{E(x)}{kT}\right)$ je obtížně zjistitelné pro velký systém, ale

$$\frac{\Pi(x')}{\Pi(x)} = \exp\left(-\frac{E(x') - E(x)}{kT}\right)$$

je snadné



Metropolisův Hastingsův algoritmus III

Speciální volba g : $g(x \rightarrow x') = g(x' \rightarrow x)$ se nazývá Metropolisův algoritmus.

- ▶ Typicky $x' = x + e$; $e \sim \mathcal{N}(0, \sigma)$
- ▶ Vhodné volit σ tak, aby $\mathcal{N}(x, \sigma) \approx \Pi$
- ▶ Lze diagnostikovat podle poměru zamítnutých pokusů:
 - $\bar{A} \rightarrow 1$: σ příliš malé - pomalý pohyb ve stavovém prostoru
 - $\bar{A} \rightarrow 0$: σ příliš velké - jedním krokem přeskočíme celé Π
- ▶ Ideálně $A \approx 0.23$ pro vícerozměrné gaussovské rozdělení Π

Boltzmannovo rozdělení:

$$A(x \rightarrow x') = \min \left\{ 1; \exp \left(-\frac{\Delta E}{kT} \right) \right\}$$



Příklad: Isingův model I

“Jednoduchý,, model feromagnetika

- ▶ Spiny $\sigma \in \{-1; 1\}$ definované na mříži s N prvky
- ▶ Paralelní konfigurace sousedních spinů a zároveň paralelní s vnějším polem H je výhodná:

$$\mathcal{H} = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - H \sum_i \sigma_i$$

$\langle i, j \rangle$ označuje páry nejbližších susedů, $J > 0$ interakční konstanta

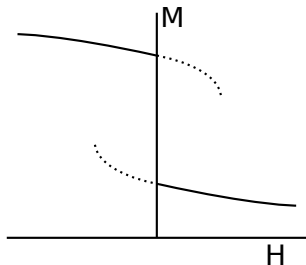
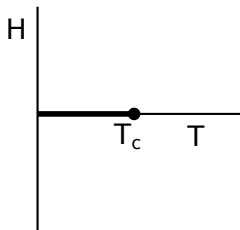


Příklad: Isingův model II

“Jednoduchý,, model feromagnetika

$$\mathcal{H} = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - H \sum_i \sigma_i$$

- ▶ Obsahuje fázový přechod
- ▶ Metastabilní stavy
- ▶ Magnetizace $M = \sum \sigma_i$



Příklad: Isingův model III

Simulace MH algoritmem s Gibbsovým vzorkováním - v každém kroku manipulujeme jen s jedním prvkem N -rozměrného stavového vektoru - překlápění spinu.

Testovací model pro

- ▶ Fázové přechody
- ▶ Metastabilní stavy
- ▶ Modely kondenzace a vypařování tekutin



Odhad chyby a konvergence MCMC

Rozptyl korelovaných dat:

$$\text{Var}_{\Pi}[\bar{f}_n(x)] = \sigma^2/n$$

$$\sigma^2 = \text{Var}_{\Pi}[f(X_1)] + 2 \sum_{i=1}^{\infty} \text{cov}_{\Pi}[f(X_1), f(X_i)]$$

Batch means - průměry podmnožin o délce $b \dots$

Otázka konvergence - těžká

- ▶ Možnost pseudokonvergence
- ▶ Batch means musí být nekorelované
- ▶ Lze zkoušet různé startovací polohy x_1
- ▶ Jiné vzorkování
- ▶ Spolehlivé je počkat...



Příklad - Google PageRank

Fiktivní Google bot náhodně prochází celý web. V každém kroku

- ▶ s 85% pravděpodobností přejde na náhodný odkaz
- ▶ s 15% pravděp. přejde na úplně náhodnou nesouvisející stránku.

Markovovský řetězec - rovnovážné rozdělení je PageRank
(pravděpodobnost náštěvy stránky)

Lze řešit maticově - Hledaný PageRank x je vlastním vektorem matice přechodu A ,

$$x = Ax$$

http://www.mathworks.com/company/newsletters/news_notes/clevescorner/oct02_cleve.html



Simulované žihání



Simulované žíhání

Myšlenka využití MCMC pro minimalizaci složitých funkcí

- ▶ Kirkpatrick et al. 1983 a Černý 1985
- ▶ Inspirace žíháním - zahřátím a následným zchlazením materiálu dojde k relaxaci - minimalizace volné energie.
- ▶ Klasický "Boltzmannovský" algoritmus pro hledání základního stavu fyzikálního systému - použijeme MH algoritmus s vysokou teplotou a postupně "ochlazujeme"
- ▶ Přejít k dynamickému MCMC
- ▶ Systém by měl skončit v základním stavu při $T = 0$.

http://link.springer.com/article/10.1007%2F978-3-642-00940-8_12



Simulované žíhání - algoritmus

Optimalizovaná veličina E , Pravděpodobnost přechodu $g_T(s \rightarrow s')$, pravděpodobnost přijetí $A_T(\Delta E)$, plán (schedule) $T(k)$, stav systému s

- ▶ $s = s_0$
- ▶ pro $k = 0 : k_{\max}$
 - ▶ $T = T(k)$
 - ▶ $s_{\text{proposal}} \sim g(s \rightarrow s')$
 - ▶ pokud $A_T(E(s_{\text{proposal}}) - E(s)) > u \sim U(0, 1)$
 - ▶ $s = s_{\text{proposal}}$



Simulované žíhání - parametry

Klasický algoritmus - Boltzmannovské žíhání, napodobuje fyziku

- ▶ $A_T \approx \exp(-\Delta E/T)$
- ▶ $g_T = (2\pi T)^{-D/2} \exp(-\Delta x^2/2T)$

Je zaručena konvergence pro

$$T(k) = \frac{T_0}{\ln k}$$

Občas se používá

$$T(k) = T_0 \exp(-\epsilon k)$$

simulované kalení (? quenching), výsledek nedefinovaný - špatné



Rychlé simulované žihání - parametry

Cauchyho rozdělení skoků s velkými chvosty - rychlejší vzorkování celého prostoru parametrů

- ▶ $A_T \approx \exp(-\Delta E/T)$
- ▶ $g_T = \frac{T}{(\Delta x^2 + T^2)^{(D+1)/2}}$

Je zaručena konvergence pro

$$T(k) = \frac{T_0}{k}$$

Existuje velmi rychlé žihání (Ingber, JMCM, 1989)

$$T(k) = T_0 \exp(-\epsilon k)$$

a řada dalších variant



Simulované žíhání - aplikace

- ▶ Fyzikální modely, hledání základního stavu systému - statistická fyzika, materiály
- ▶ Obecná optimalizace, problém obchodního cestujícího - návrh propojení procesorů atd...



Kinetické MCMC



Kinetické/dynamické MC

- ▶ Rozdělení X_n závisí na n
- ▶ Odpovídá fyzikálnímu systému, časová osa: $n \rightarrow t$

Modelový problém - chemická kinetika. Například modelování srážek v plazmatu

- ▶ Markovovský proces se spojitým časem
- ▶ Rozdělení času mezi reakcemi $\text{EXP}(w)$, $f(\tau) = \frac{1}{w} \exp(-w\tau)$
- ▶ Lze simulovat jako diskrétní markovovský proces s pravděpodobností reakce $1 - \exp(-w\Delta t)$



MC chemická kinetika

- ▶ Diskrétní stavový vektor - složky jsou počty částic
- ▶ Různé postupy výpočtu:
 - Proměnný časový krok $\tau \sim \text{EXP}(w)$
 - ▶ First reaction method, Next reaction method, Direct method
 - Fixní krátký časový krok, každá reakce probíhá s $\text{Pr} \approx 1 - \exp(-w\Delta t)$
 - Fixní dlouhý časový krok τ (τ -leaping), každá reakce probíhá k -krát v časovém kroku $\text{Pr}(k) = \frac{(w\tau)^k}{k!} \exp(-w\tau)$; $x' = x + ks$



MC zpracování dat



Testování hypotéz

Frekventistické, Př:

- ▶ Nulová hypotéza, H_0 : data jsou z určitého rozdělení např $x_1 \dots x_n \sim \mathcal{N}(\mu, \sigma^2)$
- ▶ Chyby typu I (zamítnutí H_0) a II (přijetí H_0)
- ▶ Zvolíme hladinu významnosti (signifikance) α pravděp chyby I druhu
- ▶ Test:
 - ▶ Vypočítáme statistiku $\hat{\theta}$ z našich dat
 - ▶ Generujeme náhodná data podle hypotézy H_0 a vypočteme jejich rozdělení statistiky θ
 - ▶ Pro rozdělení θ určíme kvantily (α , $1 - \alpha$, nebo $\alpha/2$ a $1 - \alpha/2$ pro levo-/pravo-/obou- stranný test)
 - ▶ Pokud $\hat{\theta}$ je extrémnější než určené kvantily, zamítáme H_0 s hladinou významnosti α .



Testování hypotéz II

- ▶ Pozor na sekvenční testování - nemůžeme nabírat data, dokud nedostaneme signifikanci
- ▶ Multiple comparison problem, alpha spending, Bonferroni correction...



Bootstrap

Př: z naměřených dat $x_1 \dots x_n$ určíme $\hat{\theta}$. Jaký je rozptyl a intervaly spolehlivosti pro θ ?

- ▶ Generujeme bootstrapový výběr $x_1^* \dots x_n^*$ náhodným výběrem s opakováním z $x_1 \dots x_n$.
- ▶ Z bootstrapového výběru určíme $\hat{\theta}^*$.
- ▶ Opakováním procedury určíme rozdělení $\hat{\theta}^*$.
- ▶ Z rozdělení určíme hledané parametry.



Příklady na cvičení

1. Generujte náhodnou proměnnou $x_i \sim 0.3\mathcal{N}(0.1, 0.1^2) + 0.7\mathcal{N}(0.5, 0.07^2)$ s pomocí Metropolisova algoritmu
 - ▶ Porovnejte různé volby kroku
2. Najděte globální minimum funkce $x^2 - \cos(20(\cos^8(4x) - 1))$
 - ▶ Hladovým (greedy) algoritmem
 - ▶ Simulovaným žíháním
3. Aplikujte simulované žíhání na složitější problém
 - ▶ Hledání základní konfigurace Van der Waalsovsky vázaného clusteru, lze použít i složitější potenciály
 - ▶ Řešení problému obchodního cestujícího
 - ▶ Další aplikace podle vlastní volby... Isinguv model feromagnetika: M-H křivky; Libovolný (PIC, fluid, hybrid) model plazmatu; Fyzikální aplikace strojového učení; Stromový algoritmus pro výpočet síly...

