Metody Monte Carlo

Stojíme před následujícím problémem

- potřebujeme spočítat $\int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_{50}}^{b_{50}} f(x_1, \dots, x_{50})\, dx_1 \dots dx_{50}$
- nemůžeme si dovolit klasickou kvadraturu, protože bychom potřebovali 3-4 body „v každé dimenzi" $\to$ odpovídá 3^{50} vyhodnocení f $\leadsto$ příliš časově náročné

Jsme tedy omezeni počtem bodů $\to$ dejme tomu, že chceme aproximovat

$$\left.\begin{array}{ll} \displaystyle\int_{\Omega} f(\vec{x})\, d\vec{x} & \begin{array}{l} f : \Omega \longrightarrow \mathbb{R} \\ \Omega \subseteq \mathbb{R}^d,\ d \in \mathbb{N} \text{ vysoké} \end{array} \end{array}\right\} \quad (I)$$

a máme praktické omezení $\to$ smíme f vyhodnotit nejvýše v N bodech pro nějaké pevně dané $N \in \mathbb{N}$ (např. dáno výkonností počítače). Otázky:

Jak aproximovat (I) na základě $f(\vec{x}_1), \dots, f(\vec{x}_N)$?

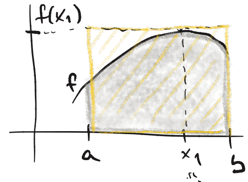
Jak volit $\vec{x}_1, \dots, \vec{x}_N \in \Omega$?

Odvození v $\mathbb{R}$ pro dané $x_1, \dots, x_n$

$\Omega \equiv [a,b]$ & máme <u>dané</u> body $x_1, \dots, x_N \in [a,b]$

• **N=1** → jak aproximujeme $\int_a^b f(x)\,dx$ na základě $x_1 \in [a,b]$ & $f(x_1) \in \mathbb{R}$?

↝ obdélníkové pravidlo:



● → hodnota $\int_a^b f(x)dx$

● → aproximace $\int_a^b f(x)dx$

daný bod $x_1 \in [a,b]$

$$\Rightarrow \int_a^b f(x)\,dx \approx (b-a)\cdot f(x_1)$$

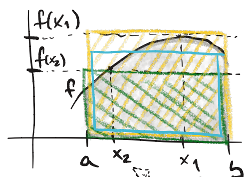
• **N>1** → <u>nechceme</u> zobecnit „N=1" jako u kvadratur:

• problém pro zobecnění do vyšších dimenzí

• nemáme kontrolu nad body $x_1, \dots, x_N$

→ místo toho „$N>1$" ≡ „N-krát $N=1$"

↝ obdélníkové pravidlo:



● → hodnota $\int_a^b f(x)dx$

● → aproximace $\int_a^b f(x)dx$ from x_1

● → aproximace $\int_a^b f(x)dx$ from x_2

● → „Monte-Carlo" aproximace – průměr ● & ●

dané body $\in [a,b]$

⇒ když máme N=2 aproximace

$$\int_a^b f(x)dx \approx (b-a)f(x_1) \quad \& \quad \int_a^b f(x)\,dx \approx (b-a)f(x_2) \rightarrow \text{co s nimi?}$$

„Monte-Carlo": vezmeme jejich průměr → $\int_a^b f(x)dx \boxed{\approx (b-a)\frac{f(x_1)+f(x_2)}{2}}$

$$\Rightarrow \int_a^b f(x)\,dx \approx \frac{b-a}{N}\sum_{i=1}^{N} f(x_i)$$ .... „Monte-Carlo" pro $\Omega = [a,b]$

Zobecnění pro $\mathcal{R} \subseteq \mathbb{R}^d$:

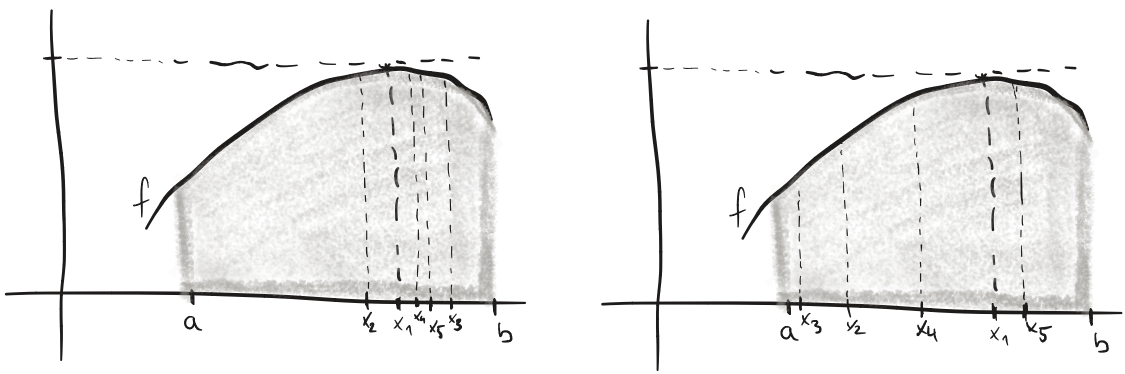
$\mathrm{vol}(\mathcal{R}) \equiv$ „Lebesgueova míra $\mathcal{R}$"
→ přirozené zobecnění délky, obsahu, objemu do $\mathbb{R}^d$

$$\int_{\mathcal{R}} f(\vec{x})\, d\vec{x} \overset{(*)}{\approx} \frac{\boxed{\mathrm{vol}(\mathcal{R})}}{N} \sum_{i=1}^{N} f(\vec{x}_i) \quad \text{pro } \vec{x}_1, \dots, \vec{x}_N \in \mathcal{R} \subseteq \mathbb{R}^d$$

⇒ Pokud máme $\vec{x}_1, \dots, \vec{x}_N \in \mathcal{R} \subseteq \mathbb{R}^d$ dané, definovali jsme jak na jejich základě aproximovat $\int_{\mathcal{R}} f(\vec{x})\, d\vec{x}$.

Jak volit $\vec{x}_1, \dots, \vec{x}_N$ aby aproximace $(*)$ fungovala dobře?

Která z těchto variant nám dá lepší aproximaci?



⇝ zjevně varianta vpravo, protože se body neakumulují v žádné konkrétní podoblasti $\mathcal{R} = [a, b]$

V našem odhadu $\frac{1}{N} \sum_{i=1}^{N} f(x_i)$ jsou si všechny body „rovnocenné", dává smysl „rovnoceně" pokrývali $\mathcal{R}$. Víme, že brát $\{x_i\}_1^N \subset \mathcal{R}$ „rovnoměrně rozdělené" nechceme

→ <u>vezmeme $\{x_i\}_1^N \subset \mathcal{R}$ jako náhodné body z rovnoměrného rozdělení</u>

Intuitivně: Pokud N není „příliš malé", pak $x_1, \dots, x_N$ pokrývají Ω „rovnoceně". Ale nemusíme pokrývat „rovnoceně" každou dimenzi $\rightarrow$ pracujeme automaticky v $\Omega \subseteq \mathbb{R}^d$.

Kombinací myšlenek výše dostáváme Monte-Carlo aproximaci (I). Další krok je vše formálně zapsat a ukázat, že to opravdu funguje

≡ chyba $\rightarrow 0$ pro $N \rightarrow +\infty$

Opáčko z teorie pravděpodobnosti

Náhodná veličina $\vec{X}$ na Ω odpovídá rozdělení

- s distribuční funkcí W $\Rightarrow$ $\mathbb{P}(\vec{X} \in \Omega_0) = \int_{\Omega_0} dW(\vec{x}) \quad \forall \Omega_0 \subseteq \Omega$

  ≡ 1. způsob jak zadat rozdělení

  $\mathbb{P}(\vec{X} \in \Omega_0)$ ≡ pravděpodobnost, že realizace náh. vel. X leží v $\Omega_0 \subseteq \Omega$

  $\int_{\Omega_0} dW(\vec{x})$: integrál vzhledem k pravděpodobnostní míře dané distr. funkcí $W(x)$ $\equiv \int_{\Omega_0} 1 \, dW(\vec{x}) \equiv \int_{\Omega} \chi_{\Omega_0}(\vec{x}) \, dW(\vec{x})$

- s hustotou $w(\vec{x})$ $\Rightarrow$ $\mathbb{P}(\vec{X} \in \Omega_0) = \int_{\Omega_0} w(\vec{x}) \, d\vec{x} \quad \forall \Omega_0 \subseteq \Omega$

  ≡ 2. způsob jak zadat rozdělení

  $\mathbb{P}(\vec{X} \in \Omega_0)$ ≡ pravděpodobnost, že realizace náh. vel. X leží v $\Omega_0 \subseteq \Omega$

  $\int_{\Omega_0} w(\vec{x}) \, d\vec{x}$: integrál vzhledem k pravděpodobnostní míře dané distr. funkcí $W(x)$ $\equiv \int_{\Omega} \chi_{\Omega_0}(\vec{x}) \cdot w(\vec{x}) \, d\vec{x}$

Náhodná veličina $\vec{X}$ na Ω má

- střední hodnotu $\mathbb{E}_W(\vec{X}) = \int_{\Omega} \vec{x} \, dW(\vec{x}) \equiv \int_{\Omega} \vec{x} \cdot w(\vec{x}) \, d\vec{x}$

  ≡ expected value / mean

- rozptyl $\mathrm{var}_W(\vec{X}) = \mathbb{E}_W\left((\vec{X} - \mathbb{E}_W(\vec{X}))^2 \right) \equiv \int_{\Omega} |\vec{x} - \mathbb{E}_W(\vec{x})|^2 \, w(\vec{x}) \, d\vec{x}$

  ≡ variance

Spojité funkce náhodných veličin jsou náhodné veličiny. (ale typicky s jiným rozdělením)

$\vec{X} \equiv \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \in \mathbb{R}^N$ a $x_i \in \mathbb{R}$ jsou nezávislé náhodné veličiny z rozdělení s hustotami $w_i(x)$ $\forall i$

pak je náhodná veličina $\vec{X}$ z rozdělení s hustotou $w(\vec{x}) \equiv \prod_{i=1}^{N} w_i(x_i)$

Teorie Monte-Carlo

Pro dané $\Omega \subseteq \mathbb{R}^d$ omezenou & $f: \Omega \to \mathbb{R}$ integrovatelnou definujeme odhad integrálu $\int_\Omega f(\vec{x})\,d\vec{x}$ klasickou metodou Monte-Carlo s N body jako

$$I_{MC}(N) := \frac{vol(\Omega)}{N} \sum_{i=1}^{N} f(\vec{X}_i) \qquad \text{pro libovolné } N \in \mathbb{N},$$

kde $\{\vec{X}_i\}_{i=1}^N$ jsou nezávislé a odpovídají náhodné veličině na Ω s rovnoměrným rozdělením.

Chyba klasické metody Monte-Carlo odpovídá $\left| \int_\Omega f(\vec{x})\,d\vec{x} - I_{MC}(N) \right|$

Krok 1: $\vec{X}_1, \dots, \vec{X}_N$ jsou nezávislé náhodné veličiny ze stejného rozdělení $\Rightarrow$

tzv. iid (independent, identically distributed)

$\Rightarrow I_{MC}(N)$ je také náhodná veličina & střední hodnota $I_{MC}(N)$ odpovídá

$\mathbb{E}(\alpha_1 Y_1 + \alpha_2 Y_2) = \alpha_1 \mathbb{E}(Y_1) + \alpha_2 \mathbb{E}(Y_2)$

$$\mathbb{E}(I_{MC}(N)) = \mathbb{E}\left(\frac{vol(\Omega)}{N} \sum_{i=1}^{N} f(\vec{X}_i) \right) = \frac{vol(\Omega)}{N} \sum_{i=1}^{N} \mathbb{E}(f(\vec{X}_i)) =$$

$$= \frac{vol(\Omega)}{N} \sum_{i=1}^{N} \int_\Omega f(\vec{x}_i) \cdot w(\vec{x}_i)\,d\vec{x} = vol(\Omega) \cdot \int_\Omega f(\vec{x})\,w(\vec{x})\,d\vec{x} = \int_\Omega f(\vec{x})\,d\vec{x}$$

$w(\vec{x})$ je hustota rovnoměrného rozdělení

$\vec{x}_i$ jsou iid

hustota rovnoměrného rozdělení na Ω odpovídá $1/vol(\Omega)$

$\Rightarrow$ „$I_{MC}(N) \approx \int_\Omega f(\vec{x})\,d\vec{x}$" ve smyslu pravděpodobnosti/střední hodnoty.

Krok 2: Chebyshevovo lemma: $\mathbb{P}(|Z - \mathbb{E}(Z)| \geq \delta) \leq \dfrac{var(Z)}{\delta^2}$

Důkaz: $$\mathbb{P}(|Z - \mathbb{E}(Z)| \geq \delta) = \int_\Omega \chi_{\{|z-\mathbb{E}(z)| \geq \delta\}}(z)\,dW(z) = \int_\Omega \chi^2_{\{|z-\mathbb{E}(z)| \geq \delta\}}(z)\,dW(z)$$

$$\leq \int_\Omega \frac{|z - \mathbb{E}(z)|^2}{\delta^2}\,dW(z) = \frac{var(Z)}{\delta^2}.$$

Krok 3: Díky Chebyshevovu lemma dostáváme

$$\mathbb{P}\left(\left|I_{MC}(N)-\underbrace{\int_{\Omega} f(\vec{x})d\vec{x}}_{\equiv \mathbb{E}(I_{MC}(N))}\right| \geq \delta\right) \leq \frac{var(I_{MC}(N))}{\delta^2}$$

$\Updownarrow$ $\mathbb{P}(|\alpha|\geq\delta)=p \Leftrightarrow \mathbb{P}(|\alpha|<\delta)=1-p$

$$\mathbb{P}\left(\left|I_{MC}(N)-\int_{\Omega} f(\vec{x})d\vec{x}\right| < \delta\right) \geq 1-\frac{var(I_{MC}(N))}{\delta^2}$$

$\Updownarrow$ $\delta \equiv \sqrt{\frac{var(I_{MC}(N))}{\varepsilon}}$

$$\mathbb{P}\left(\left|I_{MC}(N)-\int_{\Omega} f(\vec{x})d\vec{x}\right| < \sqrt{\frac{var(I_{MC}(N))}{\varepsilon}}\right) \geq 1-\varepsilon$$

$\equiv$ pravděpodobnost, že chyba odhadu I_{MC} bude menší než $\sqrt{\frac{var(I_{MC}(N))}{\varepsilon}}$ je alespoň $1-\varepsilon$ $\rightsquigarrow$ tj. pro $\varepsilon \ll 1$ je s velkou pravděpodobností chyba přinejhorším $\sqrt{\frac{var(I_{MC}(N))}{\varepsilon}}$.

Krok 4: opáčko pravděpodobnosti $\to$ $var(I_{MC}(N)) \equiv$

$$\equiv var\left(\frac{vol(\Omega)}{N}\sum_{i=1}^{N} f(X_i)\right) = \frac{vol(\Omega)^2}{N^2}\cdot var\left(\sum_{i=1}^{N} f(X_i)\right) =$$

X_i jsou iid $\Rightarrow$ $\Rightarrow f(X_i)$ jsou iid & a mají stejný rozptyl

$$= \frac{vol(\Omega)^2}{N^2}\sum_{i=1}^{N} var(f(X_i)) = \frac{vol(\Omega)^2}{N} var(f(X_1))$$ a tedy dostáváme

$$\mathbb{P}\left(\left|I_{MC}(N)-\int_{\Omega} f(\vec{x})d\vec{x}\right| < \sqrt{\frac{1}{N}}\cdot\frac{vol(\Omega)\cdot\sqrt{var(f(X))}}{\sqrt{\varepsilon}}\right) \geq 1-\varepsilon$$

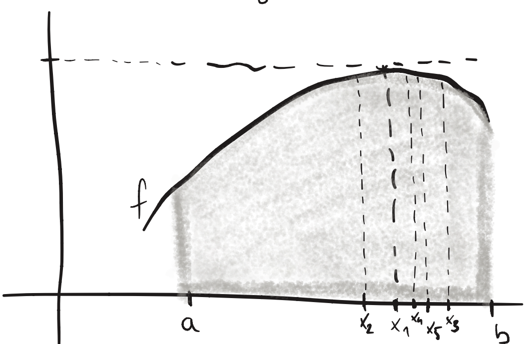
$\equiv$ chyba odhadu $I_{MC}(N)$ klesá alespoň jako $(\varepsilon \cdot N)^{-1/2}$ s pravděpodobností $1-\varepsilon$.

Importance sampling MC

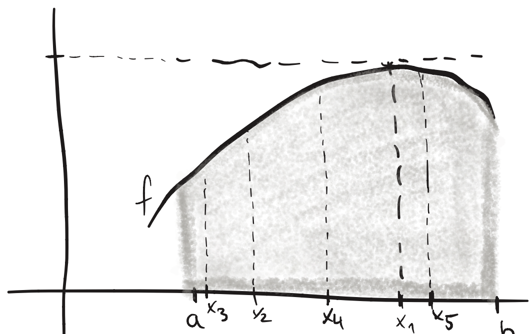
- co když je $\Omega \subseteq \mathbb{R}^d$ neomezená? Pak neexistuje rovnoměrné rozdělení
- co když umím samplovat body z Ω odpovídající jen skrze nějaké dané rozdělení $\mathcal{W}$ s distrib. funkcí $W(\vec{x})$ a hustotou $w(\vec{x})$?

$\Rightarrow$ potřebujeme upravit $I_{MC}(N)$ pro $\{X_i\}_{i=1}^N$ odpovídající danému rozdělení $\mathcal{W}$ s distrib. funkcí $W(\vec{x})$ a hustotou $w(\vec{x})$.

Nebo-li, potřebujeme zvládnout odhadnout $\int_a^b f(x)\,dx$ v „obou případech níže"



$X_i \sim$ nerovnoměrné rozdělení



$X_i \sim$ rovnoměrné rozdělení

Řešení: náš odhad $\int_a^b f(x)\,dx$ bude vážený průměr – každá funkční hodnota $f_i \equiv f(x_i)$ bude škálovaná „pravděpodobností výskytu x_i"

$w(\vec{x}) \neq 0$ skoro všude, protože je to hustota

definice $W(\vec{x})$

Zapsáno formálně: $$\int_\Omega f(\vec{x})\,d\vec{x} = \int_\Omega \frac{f(\vec{x})}{w(\vec{x})} \cdot w(\vec{x})\,d\vec{x} = \int_\Omega \frac{f(\vec{x})}{w(\vec{x})}\,dW(\vec{x})$$

a tedy $$\int_\Omega f(\vec{x})\,d\vec{x} = \mathbb{E}_W\left(\frac{f}{w}\right)$$

„X_i má velkou pravděpodobnost v $\mathcal{W}$ $\Leftrightarrow$ $w(X_i)$ je vysoká" a tedy hodnoty $f_i = f(X_i)$ jsou škálovány nepřímo úměrně jejich šanci na samplování.

střední hodnota funkce f/w vzhledem k míře dW $\rightarrow$ při výpočtu této střední hodnoty „už nejsou všechny jevy stejně pravděpodobné" jako u klasické definice $\rightarrow$ naopak zohledňují jejich pravděpodobnost, právě skrze pravděpodob. míru dW

Pokud máme $\{X_i\}_1^N$ náhodné veličiny, iid, odpovídající rozdělení $W \begin{pmatrix} \text{hust.} & w(\vec{x}) \\ \text{distr.} & W(\vec{x}) \end{pmatrix}$

$$I_{MCIS}(N) := \frac{1}{N} \sum_{i=1}^{N} \frac{f(X_i)}{w(X_i)}$$

Hodnoty $f_i = f(X_i)$ jsou škálovány „pravděpodob.", že se při samplování z W objeví $\Rightarrow$
$\Rightarrow$ „vzácné" hodnoty ovlivní odhad více než ty „běžné"

pak platí:

- $I_{MCIS}(N)$ je náhodná veličina
- $\mathbb{E}_W(I_{MCIS}(N)) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}_W\left(\frac{f(X_i)}{w(X_i)}\right) = \mathbb{E}_W\left(\frac{f}{w}\right) = \int_\Omega f(\vec{x})\, d\vec{x}$
- $\mathrm{var}_W(I_{MCIS}(N)) = \frac{1}{N^2} \sum_{i=1}^{N} \mathrm{var}_W\left(\frac{f(X_i)}{w(X_i)}\right) = \frac{1}{N} \mathrm{var}_W\left(\frac{f}{w}\right)$

a tedy za použití Chebyshevova lemma dostáváme

$$\mathbb{P}_W\left(\left| I_{MCIS}(N) - \int_\Omega f(\vec{x})\, d\vec{x} \right| < \sqrt{\frac{1}{N}} \cdot \frac{\mathrm{var}_W(f(X)/w(X))}{\sqrt{\varepsilon}} \right) \geq 1 - \varepsilon$$

všimněme si, že se nám nutně mění i <u>význam</u> toho odhadu ve smyslu „jak měříme, že je něco ne/pravděpodobné".
To by nemuselo být tak překvapivé $\rightarrow$ Ω může být neomezená a my samplujeme v I_{MCIS} z celé Ω. Tudíž v ěto nelze „klasicky, rovnoměrně" a tedy ani nedává a-priori smysl měřit pravděpodobnost výsledku ve stylu „klasicky, rovnoměrně", ale naopak použít pravděpodobnost indukovanou tím procesem generování $\{X_i\}_{i=1}^N$.

$\rightarrow$ konvergence opět řádu $(\varepsilon \cdot N)^{-1/2}$ s pravděpodobností alespoň $1-\varepsilon$

$\rightarrow$ k rozmyšlení: lze získat I_{MC} jako specifický případ I_{MCIS}?

Další adaptace MC metod

- ## adaptivní MC/MCIS

Dejme tomu, že chceme celkem použít N bodů pro aproximaci $\int_{\mathcal{R}} f(x)\,dx$.
Nejdřív nasampluju $N/2$ bodů z daného rozdělení W (klidně rovnoměrné) a spočtu si odhad $I_{MC/IS}(N/2) \approx \int_{\mathcal{R}} f(x)\,dx$.
Zbývá mi ještě dalších $N/2$ bodů → nasampluju je tam z oblasti $\tilde{\mathcal{R}} \subseteq \mathcal{R}$, kde na základě hodnot $\{f(x_i)/w(x_i)\}_{i=1}^{N/2}$ je největší rozptyl $var_W(f(x_i)/w(x_i))$

- ## stratifikované MC/MCIS

Dejme tomu, že chceme celkem použít N bodů pro aproximaci $\int_{\mathcal{R}} f(x)\,dx$.
Rozdělím si $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2 \cup \dots \cup \mathcal{R}_m$ a použiju MC/IS s $N/2m$ body na každou z oblastí $\mathcal{R}_i$ → stále mi zbývá $N/2$ bodů. Ty využiju k tomu abych „resamploval" $\mathcal{R}_i$ s nejvyšší hodnotou $var_W(f(x_i)/w(x_i))$.

→ souhrnná idea: přidat body tam, kde je vysoký rozptyl a tím ho zmenšit → zlepšujeme odhad

- ## Quazi Monte-Carlo

Idea je založená na pozorování, že pro klasické algoritmy generující náhodná čísla z rovnoměrného rozdělení pozorujeme jejich „shlukování" → demo v pythonu.
Jak jsme tyto čísla dostali? Klasicky číslici po číslici jako zbytky po dělení
Lze i lépe! Lze generovat „také náhodně vypadající", ale bez shlukování
→ tzv. quazi random numbers.

Dobrou volbou lze dosáhnout konvergence $\mathcal{O}\left(\frac{\log^d(N)}{N}\right)$ místo $\mathcal{O}(N^{-1/2})$