

Metody Monte Carlo

Stojíme před následujícím problémem

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

Jsme tedy omezení počtem bodů $\rightarrow$ dejme tomu, že chceme aproximovat

$$\left. \begin{array}{l} \int_{\mathcal{H}} f(\vec{x}) d\vec{x} \\ f: \mathcal{H} \rightarrow \mathbb{R} \\ \mathcal{H} \subseteq \mathbb{R}^d, d \in \mathbb{N} \text{ vysoké} \end{array} \right\} (I)$$

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

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

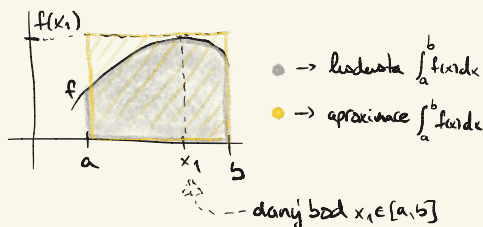
Jak volit $\vec{x}_1, \dots, \vec{x}_N \in \mathcal{H}$?

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

$\mathcal{I} \equiv [a, b]$ a máme dané body $x_1, \dots, x_N \in [a, b]$

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

ms obdélníkové pravidlo:



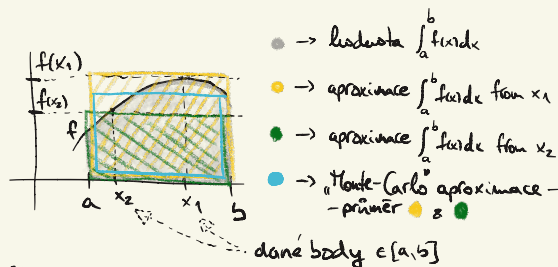
$$\Rightarrow \int_a^b f(x) dx \approx (b-a) \cdot f(x_1)$$

• $N > 1$ $\rightarrow$ necháme zobecnit „ $N=1$ “ jako u kvadratur:

$\rightarrow$ místo toho „ $N > 1$ “ $\equiv$ „ N -krát $N=1$ “

- problém pro zobecnění do vyšších dimenzí
- nemáme kontrolní bod

ms obdélníkové pravidlo:



$\Rightarrow$ když máme $N=2$ aproximace

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

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

$$\Rightarrow \int_a^b f(x) dx \approx \frac{b-a}{N} \sum_{i=1}^N f(x_i) \quad \text{„Monte-Carlo“ pro } \mathcal{I} = [a, b]$$

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

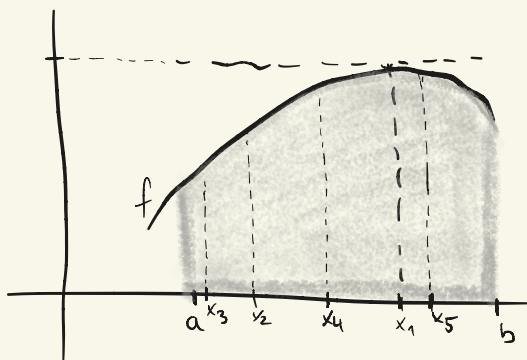
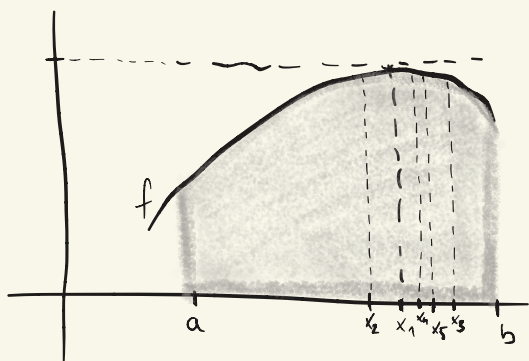
$\text{vol}(\mathcal{H}) \equiv$ "Lebesgueova míra $\mathcal{H}$ "
 $\rightarrow$ přirozené zobecnění délky, obsahu, objemu do $\mathbb{R}^d$

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

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

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

Která z těchto variant nám dá lepší aproximaci?



$\Rightarrow$ zřejmě varianta vpravo, protože se body neakumulují
v žádné konkrétní podoblasti $\mathcal{H} = [a, b]$

V našem odhadu $\frac{1}{N} \sum_{i=1}^N f(x_i)$ jsou si všedny body "rovnoměrné",
dává smysl "rovnoměrné" pokrytí $\mathcal{H}$. Víme, že brát $\{x_i\}_1^N \subset \mathcal{H}$
"rovnoměrně rozdělené" nechceme

$\rightarrow$ vezmeme $\{x_i\}_1^N \subset \mathcal{H}$ jako náhodné body z rovnoměrného rozdělení

Intuitivně: Pokud N není „příliš malá“, pak $x_1, \dots, x_N$ pokrývají $\mathcal{H}$ „rovnoměrně“. Ale nemusíme pokrývat „rovnoměrně“ každou dimenzi $\rightarrow$ pracujeme automaticky v $\mathcal{H} \subseteq \mathbb{R}^d$.

Kombinací myšlenek výše dostáváme Monte-Carlo aproximaci (I).
Další krok je vše formálně zapsat a ukázat, že to opravdu funguje
 $\equiv$ důkaz $\rightarrow 0$ pro $N \rightarrow +\infty$

Opáčko z teorie pravděpodobnosti

Náhodná veličina $\vec{X}$ na $\mathcal{H}$ odpovídá rozdělení

- s distribuční funkcí W $\Rightarrow P(\vec{X} \in \mathcal{H}_0) = \int_{\mathcal{H}_0} dW(\vec{x}) \quad \mathcal{H}_0 \subseteq \mathcal{H}$
 $\equiv$ 1. způsob jak zadat rozdělení
 $\equiv$ pravděpodobnost, že realizace náh. vel. X leží v $\mathcal{H}_0 \subseteq \mathcal{H}$
 $\mathcal{H}_0 \uparrow$
 integrál vezmeme k pravděpodobnostem uvnitř dané distr. funkce W
 $\equiv \int_{\mathcal{H}_0} 1 dW(\vec{x}) = \int_{\mathcal{H}_0} \chi_{\mathcal{H}_0}(\vec{x}) dW(\vec{x})$

- s hustotou $w(\vec{x})$ $\Rightarrow P(\vec{X} \in \mathcal{H}_0) = \int_{\mathcal{H}_0} w(\vec{x}) d\vec{x} \quad \mathcal{H}_0 \subseteq \mathcal{H}$
 $\equiv$ 2. způsob jak zadat rozdělení
 $\equiv$ pravděpodobnost, že realizace náh. vel. X leží v $\mathcal{H}_0 \subseteq \mathcal{H}$
 $\mathcal{H}_0 \uparrow$
 integrál vezmeme k pravděpodobnostem uvnitř dané distr. funkce W
 $\equiv \int_{\mathcal{H}_0} \chi_{\mathcal{H}_0}(\vec{x}) \cdot w(\vec{x}) d\vec{x}$

Náhodná veličina $\vec{X}$ na $\mathcal{H}$ má

- střední hodnotu $E(\vec{X}) = \int_{\mathcal{H}} \vec{x} dW(\vec{x}) \equiv \int_{\mathcal{H}} \vec{x} \cdot w(\vec{x}) d\vec{x}$
 $\equiv$ expected value / mean
- rozptyl $\text{var}(\vec{X}) = E((\vec{X} - E(\vec{X}))^2) \equiv \int_{\mathcal{H}} |\vec{x} - E(\vec{x})|^2 d\vec{x}$
 $\equiv$ variance

Spojitě 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 $w_i(x_i)$ či
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 $\mathcal{H} \subseteq \mathbb{R}^d$ omezenou a $f: \mathcal{H} \rightarrow \mathbb{R}$ integrovatelnou definujeme odhad integrálu $\int_{\mathcal{H}} f(\vec{x}) d\vec{x}$ klasickou metodou Monte-Carlo s N body jako

$$I_{MC}(N) := \frac{\text{vol}(\mathcal{H})}{N} \sum_{i=1}^N f(\vec{x}_i) \quad \text{pro libovolné } N \in \mathbb{N},$$

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

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

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 a střední hodnota $I_{MC}(N)$ odpovídá

$$\begin{aligned} \mathbb{E}(I_{MC}(N)) &= \mathbb{E}\left(\frac{\text{vol}(\mathcal{H})}{N} \sum_{i=1}^N f(\vec{x}_i)\right) \stackrel{\mathbb{E}(aY_1 + \dots + aY_N) = a \cdot \mathbb{E}(Y_1) + \dots + a \cdot \mathbb{E}(Y_N)}{=} \frac{\text{vol}(\mathcal{H})}{N} \sum_{i=1}^N \mathbb{E}(f(\vec{x}_i)) = \\ &\stackrel{w(\vec{x}) \text{ je hustota rovnoměrného rozdělení}}{=} \frac{\text{vol}(\mathcal{H})}{N} \sum_{i=1}^N \int_{\mathcal{H}} f(\vec{x}_i) \cdot w(\vec{x}_i) d\vec{x}_i \stackrel{\vec{x}_i \text{ jsou iid}}{=} \text{vol}(\mathcal{H}) \cdot \int_{\mathcal{H}} f(\vec{x}) w(\vec{x}) d\vec{x} \stackrel{\text{hustota rovnoměrného rozdělení na } \mathcal{H} \text{ odpovídá } 1/\text{vol}(\mathcal{H})}{=} \int_{\mathcal{H}} f(\vec{x}) d\vec{x} \end{aligned}$$

$\Rightarrow I_{MC}(N) \approx \int_{\mathcal{H}} f(\vec{x}) d\vec{x}$ ve smyslu pravděpodobnosti/střední hodnoty.

Krok 2: Chebyshevova lemma: $\mathbb{P}(|Z - \mathbb{E}(Z)| \geq \delta) \leq \frac{\text{Var}(Z)}{\delta^2}$

Důkaz: $\mathbb{P}(|Z - \mathbb{E}(Z)| \geq \delta) = \int_{\mathcal{H}} \chi_{\{|Z - \mathbb{E}(Z)| \geq \delta\}}(Z) dW(Z) = \int_{\mathcal{H}} \chi_{\{|Z - \mathbb{E}(Z)| \geq \delta\}}^2(Z) dW(Z)$

$$\leq \int_{\mathcal{H}} \frac{|Z - \mathbb{E}(Z)|^2}{\delta^2} dW(Z) = \frac{\text{Var}(Z)}{\delta^2}.$$

Krok 3: Díky Chebyshevovu lemmu dostáváme

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

$\Downarrow \quad \mathbb{P}(|x| > \delta) = p \Leftrightarrow \mathbb{P}(|x| < \delta) = 1-p$

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

$\Downarrow \quad \delta \equiv \sqrt{\frac{\text{var}(I_{MC}(N))}{\varepsilon}}$

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

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

Krok 4: opětovně pravděpodobnosti $\rightarrow \text{var}(I_{MC}(N)) \equiv$

$$\begin{aligned} &= \text{var}\left(\frac{\text{vol}(\pi)}{N} \sum_{i=1}^N f(x_i)\right) = \frac{\text{vol}(\pi)^2}{N^2} \cdot \text{var}\left(\sum_{i=1}^N f(x_i)\right) \stackrel{\substack{X_i \text{ jsou iid} \Rightarrow \\ \Rightarrow f(X_i) \text{ jsou iid} \\ \& \text{ a mají stejný rozptyl}}}{=} \\ &= \frac{\text{vol}(\pi)^2}{N^2} \sum_{i=1}^N \text{var}(f(x_i)) = \frac{\text{vol}(\pi)^2}{N} \text{var}(f(x_1)) \text{ a tedy dostáváme} \end{aligned}$$

$$\mathbb{P}\left(\left|I_{MC}(N) - \int_{\pi} f(x) dx\right| < \sqrt{\frac{1}{N}} \cdot \frac{\text{vol}(\pi) \cdot \sqrt{\text{var}(f(x))}}{\sqrt{\varepsilon}}\right) \geq 1 - \varepsilon$$

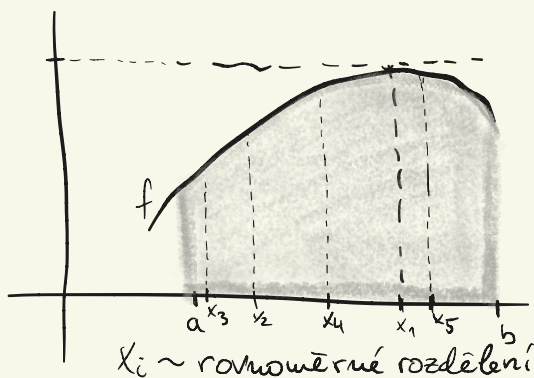
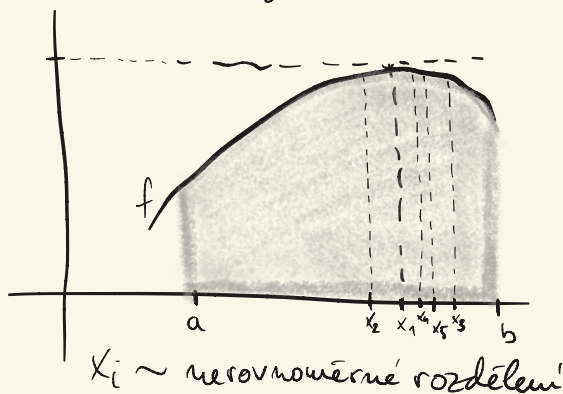
$\equiv$ chyba odhadu $I_{MC}(N)$ klesá alespoň jako $(\varepsilon \cdot N)^{-1/2}$ s pravděpodobností $1 - \varepsilon$.

Importance sampling TC

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

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



Řešení: náš odhad $\int_a^b f(x) dx$ bude vážený průměr - každá funkční hodnota $f_i \equiv f(x_i)$ bude škálována „pravděpodobností výskytu x_i “

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

Zapsáno formálně: $\int_{\mathcal{H}} f(\vec{x}) d\vec{x} \stackrel{!}{=} \int_{\mathcal{H}} \frac{f(\vec{x})}{w(\vec{x})} \cdot w(\vec{x}) d\vec{x} \stackrel{\text{definice } W(\vec{x})}{=} \int_{\mathcal{H}} \frac{f(\vec{x})}{w(\vec{x})} dW(\vec{x})$

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

x_i má velkou pravděpodobnost v $W \Leftrightarrow w(x_i)$ 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 v klasické definici $\rightarrow$ naopak zohledňují jejich pravděpodobnost, právě skrze pravděpodob. míru dW

Pokud máme $\{x_i\}_1^N$ náhodné veličiny, iid, odpovídající rozdělení w (hust. $w(x)$
distr. $W(x)$)

$$I_{MCIS}(N) := \frac{1}{N} \sum_{i=1}^N \frac{f(x_i)}{w(x_i)} \quad \left. \begin{array}{l} \text{Hodnoty } f_i = f(x_i) \text{ jsou skálovány „pravděpodob.“} \\ \text{že se při samplování z } w \text{ objeví} \Rightarrow \\ \Rightarrow \text{„vázané“ hodnoty ovlivní odhad více než} \\ \text{ty „volné“} \end{array} \right\}$$

pak platí:

- $I_{MCIS}(N)$ je náhodná veličina

- $E_W(I_{MCIS}(N)) = \frac{1}{N} \sum_{i=1}^N E_W\left(\frac{f(x_i)}{w(x_i)}\right) = E_W\left(\frac{f}{w}\right) = \int_{\mathcal{X}} f(x) dx$

- $Var_W(I_{MCIS}(N)) = \frac{1}{N^2} \sum_{i=1}^N Var_W\left(\frac{f(x_i)}{w(x_i)}\right) = \frac{1}{N} Var_W\left(\frac{f}{w}\right)$

a tedy za použití Chebysheva lemma dostáváme

$$\mathbb{P}_W\left(\left|I_{MCIS}(N) - \int_{\mathcal{X}} f(x) dx\right| < \sqrt{\frac{1}{N}} \cdot \frac{Var_W(f(x)/w(x))}{\sqrt{\varepsilon}}\right) \geq 1 - \varepsilon$$

↳ Všimněme si, že se nám nutně mění i význam toho odhadu ve smyslu „jak měříme, že je něco ne/pravděpodobné“.

To by nemuselo být tak překvapivé $\rightarrow$ $\mathcal{X}$ může být neomezená a my samploujeme v I_{MCIS} z celé $\mathcal{X}$. Tudiž už to nelze „klasicky, rovnoměrně“ a kdy 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 $\int_{\mathcal{H}} f(x) dx$.
Nejdříve nasamplujeme $N/2$ bodů z daného rozdělení w (klidně rovnoměrné) a spočítáme odhad $I_{MCIS}(N/2) \approx \int_{\mathcal{H}} f(x) dx$.
Zbývá nám ještě dalších $N/2$ bodů $\rightarrow$ nasamplujeme je tam z oblasti $\tilde{\mathcal{H}} \subseteq \mathcal{H}$, kde na základě hodnot $\{f(x_i)/w(x_i)\}_{i=1}^{N/2}$ je největší rozptyl $\text{var}_w(f(x_i)/w(x_i))$.

• stratifikované MC/MCIS

Dejme tomu, že chceme celkem použít N bodů pro aproximaci $\int_{\mathcal{H}} f(x) dx$.
Rozdělíme si $\mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2 \cup \dots \cup \mathcal{H}_m$ a použijeme MC/IS s $N/2m$ body na každou z oblastí $\mathcal{H}_i \rightarrow$ stále nám zbývá $N/2$ bodů. Ty využijeme k tomu abychom „resamplovali“ $\mathcal{H}_i$ s největší hodnotou $\text{var}_w(f(x_i)/w(x_i))$.

$\rightarrow$ souhrnná idea: přidat body tam, kde je vyšší rozptyl a tím ho zmenšit $\rightarrow$ zlepšujeme odhad

• Quasi Monte-Carlo

Idea je založená na pozorování, že pro klasické algoritmy generující náhodná čísla z rovnoměrného rozdělení pozorujeme jejich „shlukování“ $\rightarrow$ demo v pythonu.

Jak jsme tyto čísla dostali? Klasicky číslíci po číslíci jako zbytky po dělení.
Lze i lépe! Lze generovat „také náhodně vypadající“, ale bez shlukování.
 $\rightarrow$ tzv. quasi random numbers.

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