

# 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“  $\rightarrow$  odpovídá  $3^{50}$  výhodnocení  $f$   $\rightarrow$  příliš časově náročné

Jsme tedy omezeni počtem bodů  $\rightarrow$  dejme tomu, že chceme approximovat

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

a máme praktické omezení  $\rightarrow$  summe  $f$  vylučovat nejvýše v  $N$  bodech pro nějaké povídání  $N \in \mathbb{N}$  (např. dleto výkonnosti počítače). Otázky:

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

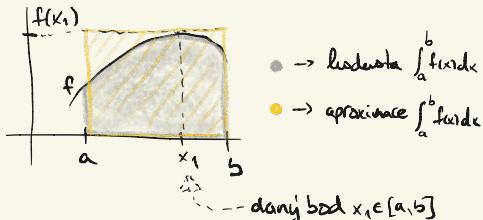
Jak volit  $\vec{x}_1, \dots, \vec{x}_N \in \mathcal{H}$ ?

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

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

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

$\Rightarrow$  obdélníkové pravidlo:

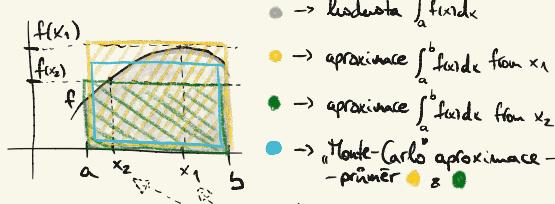


$$\Rightarrow \int_a^b f(x) dx \approx (b-a) \cdot f(x_1)$$

- $N > 1$   $\rightarrow$  nechceme zohlednit " $N=1$ " jde o kvadraturu:

$\rightarrow$  místo toho  $"N > 1" \equiv "N\text{-krát } N=1"$

$\Rightarrow$  obdélníkové pravidlo:



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

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

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

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

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

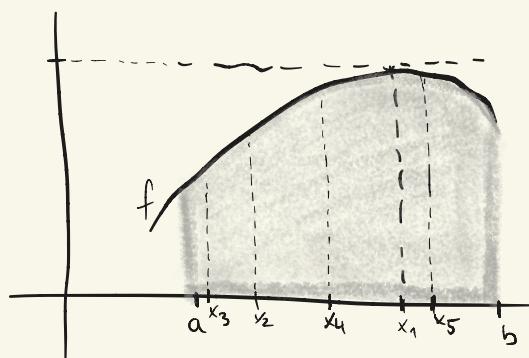
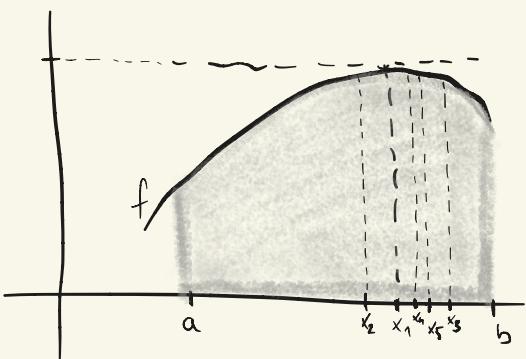
$\text{vol}(\mathcal{H}) = \text{"Lebesgueova měra } \mathcal{H}$ "  
→ přirozené zobecnění délky, obsahu, objemu do  $\mathbb{R}^d$

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

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

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

Která z těchto variant nám dá lepsí approximaci?



vn) zjednodušená varianta vpravo, protože se body neakumulují

✓ žádou konkrétní podoblasti  $\mathcal{H} = [a, b]$

✓ našem odhadu  $\frac{1}{N} \sum_{i=1}^N f(x_i)$  jsou si všechny body "rovnocenné", dává smysl "rovnocenné" pokryvat  $\mathcal{H}$ . Vhodné, že brát  $\{x_i\}_1^N \subset \mathcal{H}$  "rovnoměrně rozdělené" mechateme  
→ vezmeme  $\{x_i\}_1^N \subset \mathcal{H}$  jako "vhodné" body z rovnoměrného rozdělení

## Intuitivné:

Pakend  $N$  neu "prális male", pak  $x_1, \dots, x_N$  pokryvají  $\mathcal{H}$  "rovnocenně". Ale nemusíme pokryvat "rovnocenně" každou dimenzi  $\rightarrow$  pracujeme automaticky v  $\mathcal{H} \subseteq \mathbb{R}^d$ .

Kombinaci myšlenek výše dostáváme Monte-Carlo approximaci (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  $W$   $\Rightarrow P(\vec{X} \in \mathcal{H}_0) = \int dW(\vec{x}) \quad \forall \mathcal{H}_0 \subseteq \mathcal{H}$ 
  - $\equiv$  1. způsob jak zadat rozdělení
  - $\equiv$  pravděpodobnost, že realizace náh. vel.  $X$  leží v  $\mathcal{H}_0 \subseteq \mathcal{H}$
  - integral vzhledem ke pravděpodobnosti vlivé dané distr. funkci  $W_{\text{un}}$
  - $\equiv \int_{\mathcal{H}_0} 1 \, d\vec{x} \equiv \int_{\mathcal{H}_0} x_{\mathcal{H}_0}(\vec{x}) \, dW(\vec{x})$

- s husťotou  $w(\vec{x})$   $\Rightarrow P(\vec{X} \in \mathcal{H}_0) = \int w(\vec{x}) \, d\vec{x} \quad \forall \mathcal{H}_0 \subseteq \mathcal{H}$ 
  - $\equiv$  2. způsob jak zadat rozdělení
  - $\equiv$  pravděpodobnost, že realizace náh. vel.  $X$  leží v  $\mathcal{H}_0 \subseteq \mathcal{H}$
  - integral vzhledem ke pravděpodobnosti vlivé dané distr. funkci  $W_{\text{un}}$
  - $\equiv \int_{\mathcal{H}_0} x_{\mathcal{H}_0}(\vec{x}) \cdot w(\vec{x}) \, d\vec{x}$

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

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

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

$\vec{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \in \mathbb{R}^N$  a  $x_i \in \mathbb{R}$  jsou nezávislé náhodné veličiny z rozdělení s hustotami  $w_i(x_i)$ .  
 pak je náhodná veličina  $\vec{X}$  z rozdělení s hustotou  $w(\vec{x}) = \prod_{i=1}^N w_i(x_i)$

# Teorie Monte-Carlo

Pro dané  $\mathcal{M} \subseteq \mathbb{R}^d$  omezenou a  $f: \mathcal{M} \rightarrow \mathbb{R}$  integratelnou definujeme odhad integračního  $\int_{\mathcal{M}} f(x) dx$  klasickou metodou Monte-Carlo s  $N$  body jako

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

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

Aby byla klasické metody Monte-Carlo odpovídala  $\left| \int_{\mathcal{M}} f(x) dx - I_{MC}(N) \right|$

Krok 1:  $\vec{X}_1, \dots, \vec{X}_N$  jsou nezávislé náhodné veličiny ze stejného rozdělení  $\Rightarrow$   
ter. iid (independent, identically distributed)

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

$$\mathbb{E}(I_{MC}(N)) = \mathbb{E}\left(\frac{\text{vol}(\mathcal{M})}{N} \sum_{i=1}^N f(\vec{x}_i)\right) = \frac{\text{vol}(\mathcal{M})}{N} \sum_{i=1}^N \mathbb{E}(f(\vec{x}_i)) =$$

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

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

$\vec{x}_i$  jsou iid

hustota rovnoměrného rozdělení  
na  $\mathcal{M}$  odpovídá  $1/\text{vol}(\mathcal{M})$

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

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

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

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

Krok 3: Díky Chebyshevova lemma dostáváme

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

$\hat{=} E(I_{MC}(N))$

$\Updownarrow P(|x| > \delta) = p \Leftrightarrow P(|x| < \delta) = 1-p$

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

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

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

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

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

$$\begin{aligned} &= \text{var}\left(\frac{\text{vol}(\Omega)}{N} \sum_{i=1}^N f(X_i)\right) = \frac{\text{vol}(\Omega)^2}{N^2} \cdot \text{var}\left(\sum_{i=1}^N f(X_i)\right) = \\ &= \frac{\text{vol}(\Omega)^2}{N^2} \sum_{i=1}^N \text{var}(f(X_i)) = \frac{\text{vol}(\Omega)^2}{N} \text{var}(f(X_i)) \quad \text{a tedy dostáváme} \end{aligned}$$

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

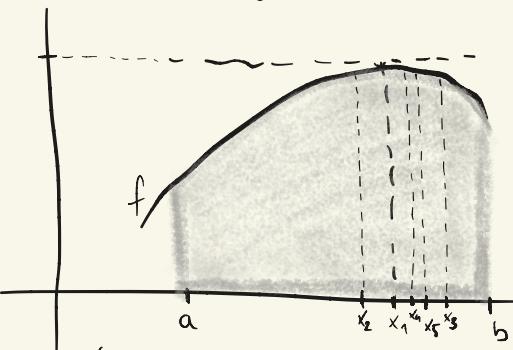
$$P\left(\left|I_{MC}(N) - \int_{\Omega} f(x) dx\right| < \sqrt{\frac{1}{N} \cdot \frac{\text{vol}(\Omega) \cdot \text{var}(f(X))}{\varepsilon}}\right) \geq 1 - \varepsilon$$

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

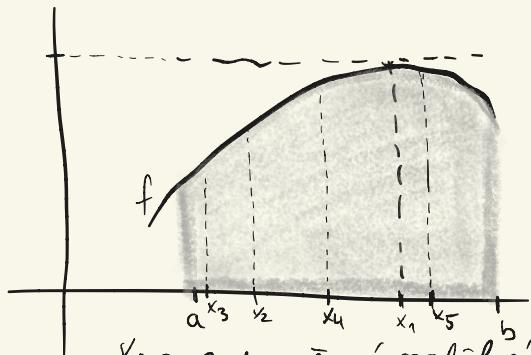
# Importance sampling TIC

- co když je  $\Omega \subseteq \mathbb{R}^d$  neomezená? Pak existuje rovnoměrné rozdělení
  - co když umím sampling body z  $\Omega$  odpovídající jen sítce nějaké dané rozdělení  $W$  s distrib. funkcií  $W(x)$  a hustotou  $w(x)$ ?
- $\Rightarrow$  potřebujeme upravit  $I_{MC}(N)$  pro  $\{x_i\}_{i=1}^N$  odpovídající danému rozdělení  $W$  s distrib. funkcií  $W(x)$  a hustotou  $w(x)$ .

Někdy, potřebujeme zvládnout odhadnout  $\int_a^b f(x) dx$  v „obou případech níže“.



$x_i \sim$  nerovnoměrné rozdělení



$x_i \sim$  rovnoměrné rozdělení

Důsledek: našed odhad  $\int_a^b f(x) dx$  bude važený průměr - každá funkční hodnota  $f_i = f(x_i)$  bude školovaná „pravděpodobností výskytu  $x_i$ “ ( $w(x_i) > 0$  stejně všechny, protože je to hustota definice  $w(x)$ )

Zapsáno formálně:  $\int_{\Omega} f(x) dx \stackrel{?}{=} \int_{\Omega} \frac{f(x)}{w(x)} \cdot w(x) dx = \int_{\Omega} \frac{f(x)}{w(x)} dW(x)$

a tedy  $\int_{\Omega} f(x) dx = \mathbb{E}_{W(\frac{f}{w})}$

$x_i$  má velkou pravděpodobnost v  $\Omega \Leftrightarrow w(x_i)$  je vysoká  
a tedy hodnoty  $f_i = f(x_i)$  jsou školované vejmimo i měřitelné jejich řadci na samplingu.

střední hodnota funkce  $f/w$  vizitkem k mře dW  $\rightarrow$  pravděpodobnost střední hodnoty (už nejsou všechny jen stejné pravděpodobné) jako v klasické definici  $\rightarrow$  napak zohledňují jejich pravděpodobnost, právě stejně pravděpodobnost dW

Pokud máme  $\{X_i\}_{i=1}^N$  náhodné veličiny, iid, odpovídající rozdělení  $W$  (last.  $w(x)$  distr.  $W(x)$ )

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

Hodnoty  $f_i = f(X_i)$  jsou školovanými „pravděpodob.“,  
 že se při samplování z  $W$  objeví  $\Rightarrow$   
 ⇒ „vzdálené“ hodnoty ovlivní odhad více než  
 by „běžné“

pak platí:

- $I_{MCIS}(N)$  je náhodná veličina

- $E_W(I_{MCIS}(N)) = \frac{1}{N} \sum_{i=1}^N E_W\left(\frac{f(X_i)}{w(X_i)}\right) = E_W\left(\frac{f}{w}\right) = \int f(x) dx$

- $\text{Var}_W(I_{MCIS}(N)) = \frac{1}{N^2} \sum_{i=1}^N \text{Var}_W\left(\frac{f(X_i)}{w(X_i)}\right) = \frac{1}{N} \text{Var}_W\left(\frac{f}{w}\right)$

a tedy za použití Chebyshevova lemma dostáváme

$$P_W\left(\left|I_{MCIS}(N) - \int f(x) dx\right| < \sqrt{\frac{1}{N}} \cdot \frac{\text{Var}_W(f/w)}{\sqrt{\epsilon}}\right) \geq 1 - \epsilon$$

→ všimněte si, že se uvažuje měření i význam toho odhadu ve smyslu  
 „jak měříme, že je něco ne/pravděpodobné“.

To by nemuselo být tak překvapivé → může být neomezená  $\infty$   
 samplujeme v  $I_{MCIS}$  z celé  $\Omega$ . Tudíž věto nelze „klasicky, rovnomerne“  
 a tedy ani nedává a-priori smysl měřit pravděpodobnost výsledku ve smyslu  
 „klasicky, rovnomerne“, ale naproti tomu můžete mít pravděpodobnost indukovanou  
 tím procesem generováním  $\{X_i\}_{i=1}^N$ .

→ konvergence opět rádu  $(\epsilon \cdot N)^{-1/2}$  s pravděpodobností alespoň  $1 - \epsilon$

→ k rozmyslení: lze získat  $I_{MC}$  jako specifický případ  $I_{MCIS}$ ?

# Další adaptace MC metod

## • adaptivní MC/MCIS

Definujme tomu, že chceme celkem použít  $N$  bodů pro approximaci  $\int_{\Omega} f(x) dx$ .

Nejdřív nasampljuji  $N/2$  bodů z daného rozdělení  $W$  (blíže rovnoměrné) a spočtu si odhad  $J_{MCIS}(N/2) \approx \int_{\Omega} f(x) dx$ .

Zbývá mi ještě dalších  $N/2$  bodů  $\rightarrow$  nasampljuji je tam z oblasti  $\tilde{\Omega} \subseteq \Omega$ , 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

Definujme tomu, že chceme celkem použít  $N$  bodů pro approximaci  $\int_{\Omega} f(x) dx$ .

Rozdělim si  $\Omega = \Omega_1 \cup \Omega_2 \cup \dots \cup \Omega_m$  a použiju MC/IS s  $N/2m$  body na každou z oblasti  $\Omega_i \rightarrow$  stále mi zbývá  $N/2$  bodů. Ty využiju k tomu abych „resamplingoval“  $\Omega_i$  s nejvyšší hodnotou  $\text{var}_W(f(X_i)/w(X_i))$ .

→ soulímná idea: přidat body tam, kde je vysoký 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? Klasické číslíci po číslici jako zbytky po dělení.  
Lze i lepe! Lze generovat „také náhodně vypadající“, ale bez shlukání  
 $\rightarrow$  tzv. quasi random numbers.

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