Markov Chain Monte Carlo (MCMC)

Monte Carlo - metoda výpočtu deterministických veličin náhodnými algoritmy

Příklad: $\Omega \subset [0,1]^d$, zajímá nás $|\Omega|$ ... Lebesgueova míra Ω.

Předpokládejme, že
- $\int_{[0,1]^d} \chi_\Omega(x)dx$ nelze jednoduše a přesně spočítat
- pro bod $x \in [0,1]^d$ lze jednoduše rozhodnout jestli $x \in \Omega$ nebo $x \notin \Omega$.

$\Rightarrow |\Omega| = \int_\Omega 1 dx = \int_{[0,1]^d} \chi_\Omega(x)dx$ nahradíme $\frac{1}{n}\sum_{i=1}^{n} \chi_\Omega(x_i)$,

kde $x_i \in [0,1]^d$ jsou voleny nezávisle a rovnoměrně rozdělení.

... obecně $\mathbb{E}X$ nahradíme $\frac{1}{n}\sum_{j=1}^{n} X_j$

Jaká je průměrná chyba?!

$$\mathbb{E}\left|\mathbb{E}X - \frac{1}{n}\sum_{j=1}^{n} X_j\right|^2 = \mathbb{E}\left|\frac{1}{n}\sum_{j=1}^{n}(\mathbb{E}X - X_j)\right|^2$$

$$= \frac{1}{n^2}\,\mathbb{E}\sum_{j,k=1}^{n}(\mathbb{E}X - X_j)(\mathbb{E}X - X_k) = \frac{1}{n^2}\sum_{j=1}^{n}\mathbb{E}(\mathbb{E}X - X_j)^2 = \frac{n\sigma^2}{n^2}$$

$\Rightarrow$ Průměrná chyba (v L_2-smyslu) je tedy $\frac{\sigma}{\sqrt{n}}$, kde $\sigma^2 = var(X)$

a n je počet opakování.

Příklady metody Monte Carlo:

• Výpočet π: $[0,1]^2$, $\Omega = \{(x_1,x_2)\in[0,1]^2,\ x_1^2+x_2^2\le 1\}$

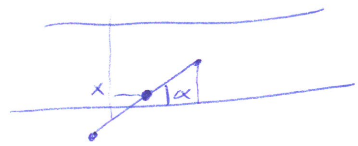
$(x_1,x_2)\in[0,1]^2$ rovnoměrně, $X = \begin{cases} 1 \ \dots\ (x_1,x_2)\in\Omega \\ 0 \ \dots\ (x_1,x_2)\notin\Omega \end{cases}$

$\mathbb{E}X = \pi/4$ ... Volíme (x_1,x_2) náhodně a počítáme podíl těch, které padnou do jednotkového kruhu...

• Podobně funguje experiment s jehlou délky 1 házenou náhodně na rovnoběžné přímky se vzdáleností 1:

- x ... vzdálenost středu jehly od nejbližší přímky ... $[0,1/2]$
- α ... úhel mezi jehlou a přímkami $\alpha\in[0,\pi/2]$

$B = \{(x,\alpha)\in[0,1/2]\times[0,\pi/2] : 0\le x\le \frac{1}{2}\sin\alpha\}$



$\mathbb{P}(B) = \frac{4}{\pi}\int_0^{\pi/2}\frac{1}{2}\sin\alpha\, d\alpha = \frac{2}{\pi}.$

... Buffon 1777, Laplace 1812

• Vlastnosti náhodných polygonů

Sylvester (1864): Pro konvexní $K\subset\mathbb{R}^2$ zvolme nezávisle $x_1,\dots,x_4\in K$.
Jaká je pravděpodobnost, že $\mathrm{conv}(x_1,\dots,x_4)$ je trojúhelník?

... stačí spočítat střední hodnotu plochy $\mathrm{conv}(x_1,x_2,x_3)$ ...

$\Rightarrow$ šestirozměrný integrál

... možno simulovat metodou Monte Carlo

... nebo spočítat analyticky

(známé hodnoty pro kruh, trojúhelník ...)

Pokud je integrál tvaru $I = \int_{[0;1]^d} f(x)\,\psi(x)\,dx$, pak lze brát

$\frac{1}{n}\sum_{i=1}^{n} f(x_i)$, kde ale x_i jsou brány náhodně s hustotou ψ

- $X = f(x)$, $\mathbb{E}X = \int f(x)\psi(x)\,dx = I$.

=> Pro použití metody Monte Carlo je nutné umět generovat náhodné vzorky (x_i) vzhledem k dané distribuci.

... Vyrobíme Markovův řetězec, který bude mít danou distribuci jako stacionární rozdělení a vezmeme X_n pro n velké!

=> Pro S a π na S dané hledáme Markovský řetězec tak, aby

- π byl stacionární stav ... $\pi = \pi P$
- X_n konvergovalo rychle k π (v nějakém smyslu)

  ... tedy aby $\mathbb{P}(X_n = j) \sim \pi_j$ pro n malé...

- Další podmínky ... $X_n \to X_{n+1}$ lze realizovat rychle

Mějme tedy dáno S a π na S.

Metropolisův algoritmus:

- Předpokládejme, že na S máme (nerozložitelný) Markovský řetězec se symetrickou maticí pravděpodobností přechodu $(\Psi_{x,y})_{x,y \in S}$
- Modifikujeme tento řetězec tak, aby měl stac. rozdělení π.
- Přechod z x do y akceptujeme s pravd. $a(x,y)$, odmítneme s $1-a(x,y)$

Tedy $$P_{xy} = \begin{cases} \Psi_{xy}\, a(x,y), & y \neq x \\ 1 - \sum_{y \neq x} \Psi_{xy}\, a(x,y), & y = x \end{cases}$$

- π bude stac. pro P, pokud budou splněny rovnice detailní rovnováhy:

$$\pi(x) P_{xy} = \pi(y) P_{yx} \quad , x \neq y$$

tedy $\pi(x) \cancel{\Psi_{xy}}\, a(x,y) = \pi(y) \cancel{\Psi_{yx}}\, a(y,x)$

Ψ symetrická

$$\pi(x) a(x,y) = \pi(y) a(y,x)$$

a chceme volit co největší ($\Rightarrow$ rychlost konvergence)

$$\Rightarrow \pi(x) a(x,y) = \pi(y) a(y,x) = \min(\pi(x), \pi(y))$$

$$\ldots\ a(x,y) = \frac{\min(\pi(x), \pi(y))}{\pi(x)}, \quad a(y,x) = \frac{\min(\pi(x), \pi(y))}{\pi(y)}$$

$$= \min\left(1, \frac{\pi(y)}{\pi(x)}\right)$$

$$\Rightarrow P_{xy} = \begin{cases} \Psi_{xy} \cdot \min\left(1, \frac{\pi(y)}{\pi(x)}\right), & x \neq y \\ 1 - \sum_{y \neq x}, & x = y \end{cases}$$

Poznámky: • Metropolisův algoritmus závisí jen na podílu $\pi(x)/\pi(y)$

Pokud $h: S \to \mathbb{R}_+$ a $\pi(x) = \frac{h(x)}{Z}$, $Z = \sum'_{x \in S} h(x)$

(tedy π je známé až na normalizační konstantu Z), pak $\pi(x)/\pi(y) = h(x)/h(y)$ a Z není třeba počítat.

• Pokud T není symetrická (ale je nerozložitelná)

$$P(x,y) = T(x,y) \cdot \min\left\{1, \frac{\pi(y)\, T(y,x)}{\pi(x)\, T(x,y)}\right\}, \quad x \neq y$$

$$= 1 - \sum'_{z \neq x} \quad - \text{"} - \qquad\qquad , x = y$$

Příklady: • (V,E) - graf, obarvení V q-barvami

$f: V \to \{1, \dots, q\}$

Množina všech obarvení: $\{1, \dots, q\}^V$

Vlastní obarvení - sousední vrcholy mají různou barvu

... Pro všechny hrany $\{v, w\} \in E$: $f(v) \neq f(w)$

S = množina vlastních obarvení (V,E) s q barvami

- neznáme velikost S

Chceme generovat prvky $\mathcal{S}$ uniformně: $\pi(f) = \frac{1}{\# \mathcal{S}}$.

Glauberova dynamika vlastních q-obarvení je následující řetězec:

- pro konfiguraci f a vrchol v řekneme, že barva $j \in \{1, \dots, q\}$ je "dovolená", pokud je různá od barev všech sousedů: $j \notin \{f(w), w \sim v\}$

Řetězec:
- vybereme $v \in V$ náhodně (rovnoměrně)
- vybereme náhodně (rovnoměrně) jednu z dovolených barev j
- Obarvíme v barvou j.

Tento řetězec má jako stac. rozdělení rovnoměrné rozdělení na $\mathcal{S}$:

- Mezi dvěma konfiguracemi f, g lze projít jen tehdy liší-li se v jednom vrcholu v

$$\mathbb{P}(X_{n+1} = g \mid X_n = f) = \underbrace{\frac{1}{\# V}}_{\text{vybereme } v} \cdot \frac{1}{\# \{j : j \text{ je dovolená ve } v\}}$$

$$= \mathbb{P}(X_{n+1} = f \mid X_n = g) \quad \dots \quad P(f,g) = P(g,f)$$

... rovnice detailní rovnováhy pro rovnoměrné rozdělení!

• další příklad: "hardcore configuration"

Do vrcholů grafu (V,E) umísťujeme částice tak, aby žádné dvě nesousedily

$$\mathbb{S} = \{ f : \{0,1\} \to V : f(v)\cdot f(w) = 0 \text{ pro } v \sim w \}$$

Řetězec pro výběr náhodné hardcore configuration:

- Vybereme náhodně rovnoměrně vrchol $v \in V$ bez ohledu na jeho obsazení
- je-li nějaký soused obsazen ($\exists w \in V : v \sim w \ \& \ f(w) = 1$) necháme v neobsazen ($f(v) = 0$)
- jinak jej s $P = 1/2$ obsadíme a s $P = 1/2$ neobsadíme.

7.4 Ising model

The *Ising model* is a fundamental model of statistical physics. Let us assume that we have a d-dimensional grid with $N \times N \times \cdots \times N$ points. Later on, we stick to $d = 2$ with N^2 vertices as the most transparent (and most understood) system.[11] The vertices are neighbours if (and inly if) they differ only in one coordinate by the value 1.

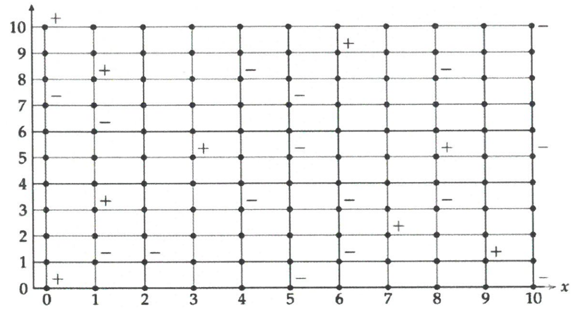


Figure 1: 11×11 grid in $\mathbb{R}^2$ with 11^2 vertices, some of them with an assigned spin.

There are 2^{N^2} ways, how one can assign a spin $\{-1, +1\}$ to each of these vertices. We define a probability dostribution on all these 2^{N^2} possible ways, how to distribute the spin across the grid. This will make some of the assignments more likely than the others. To describe the details, we need some notation. We denote the set of vertices of the grid by V and we denote by $v \sim w$ the fact, that v and w are neighbours. If $\sigma : V \to \{-1, +1\}$ is an assigments of spins to the vertices, we denote by $H(\sigma)$ its *energy*

$$H(\sigma) = -\sum_{\substack{v,w \in V \\ v \sim w}} \sigma(v)\sigma(w).$$

The idea behing this formula is clear: if two neighbouring vertices have the same spin (i.e., $\sigma(v) = \sigma(w) = +1$ or $\sigma(v) = \sigma(w) = -1$), then the energy is decreased by 1, if the spins differ, then $\sigma(v) \cdot \sigma(w) = -1$ and the energy is increased. So, $H(\sigma)$ is the number of pairs of neighbouring vertices with different spin minus the number of pairs of neighbouring vertices with the same spin.

The probability distribution on all mappings $\sigma : V \to \{-1, +1\}$ is then proportional to $e^{-\beta H(\sigma)}$, where $\beta > 0$ is a real parameter. This means that

$$\mu(\sigma) = \frac{1}{Z(\beta)} e^{-\beta H(\sigma)}$$

and

$$Z(\beta) := \sum_{\sigma} e^{-\beta H(\sigma)}$$

is the normalization factor, which ensures that $\sum_\sigma \mu(\sigma) = 1$.

Some of the properties of this model are easy to observe, namely

[11] In general, one can consider an arbitrary graph $G = (V, E)$ instead.

- The probability space of all σ's is huge, it has 2^{N^2} elements.
- For a given σ, $H(\sigma)$ is quite easy to compute.
- It is rather difficult to compute $Z(\beta)$ as it would require to calculate the (huge) sum over all σ's.
- However, for two different σ_1 and σ_2, the ratio $\mu(\sigma_1)/\mu(\sigma_2)$ is again easy to compute, as the normalization factor $Z(\beta)$ cancels out.

We are mainly interested in two (somehow connected) questions

- How does a typical configuration of σ chosen randomly with respect to μ look like? Is it more random or is it more structured?
- How do we generate a random sample with respect to μ?

The question about the typical configuration of σ is rather tricky. From the 2^{N^2} configurations, vast majority is rather chaotic, but their energy $H(\sigma)$ is large (we expect them to have many neighbours with different spin) and, therefore, $\mu(\sigma)$ is very small. The "structured" configurations, with large areas in the grid of vertices with the same spin, are much less in number, but their energy $H(\sigma)$ is very small making $\mu(\sigma)$ large.

It turns out, that the parameter β plays a crucial role. In $d = 2$, there is a *phase trnasition*: For small values of β (or large values of the temperature $T = 1/\beta$) the values of $H(\sigma)$ (and of $\mu(\sigma)$) do not differ much and the random configuration is rahter chaotic. For large values of β (and small values of T) structured configurations are more likely. Somewhere inbetween is the *critical temperature* ($\beta = \ln(1 + \sqrt{2})/2$). The analytical description of this solution was given by Lars Onsager in 1944. The analytical solution for $d \geq 3$ is unknown so far.

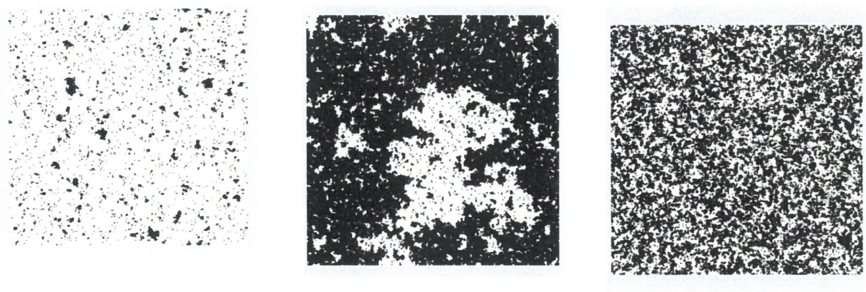


Figure 2: Low, critical, and high temperature on 250×250 grid, from the book of David Levin and Yuval Peres.

To answer the second question, namely to generate random spin configurations with respect to μ, one can use the Markov Chain Monte Carlo method. We construct a Markov Chain on the state space $\mathbb{S} = \{-1, +1\}^{N^2}$, which has stationary distribution μ, we start randomly and iterate for long enough until the actual state is (almost) random with respect to μ. For that sake, we use the *Glauber dynamics*, which starts with arbitrary σ, and then repeatedly picks a vertex w randomly and uniformly from all vertices and then re-defines the spin of w randomly.