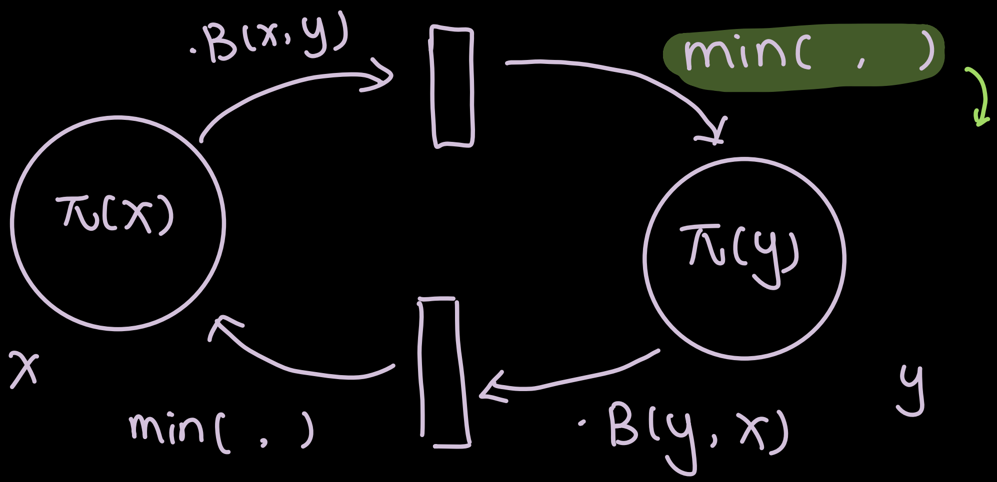
Metropolis Algorithm

target $\pi(x)$ $\left(= \underbrace{\frac{1}{Z}}_{\text{unknown}} e^{-\mathcal{E}(x)} \right)$



base chain : $B(x,y)$: $P(X_{\text{proposal}} = y \mid X_t = x)$

(Simple, can be symmetric : $B(x,y) = B(y,x)$)

$$P(\text{accept } X_{\text{prop}} = y) = \min\left(1, \frac{\boxed{\pi(y)}\, B(y,x)}{\boxed{\pi(x)}\, B(x,y)}\right)$$

Z gets cancelled

$$\pi(x) B(x,y) \cdot P(\text{accept } y)$$

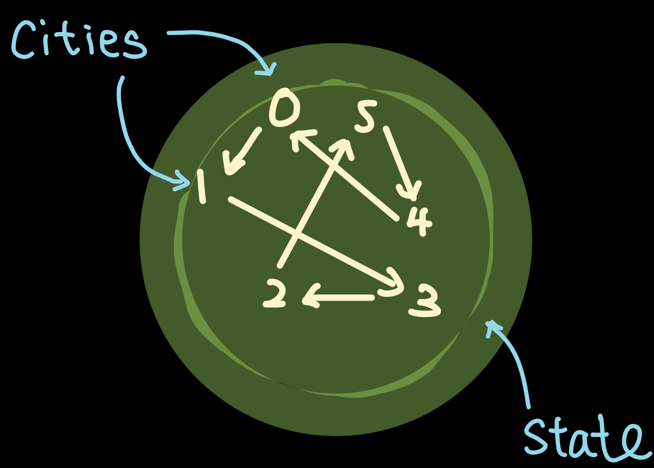
$$= \min\left(\pi(x) B(x,y), \pi(y) B(y,x)\right)$$

For symmetric $B(x,y) = B(y,x)$

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

Traveling Salesman Problem

goal: find the path with shortest distance



each state x is a permutation of $\{1, 2, \dots, m\}$

state space = $\{$all $m!$ permutations$\}$

sample space $\longrightarrow$ n samples

Too big, $m!$ $\downarrow$ min

n samples $\downarrow$ $\sim \exp(-\mathcal{E}(x))$

$$X = (\sigma_1\ \sigma_2 \cdots \sigma_{i-1} \ast \sigma_i \ast \sigma_{i+1} \cdots \sigma_{j-1} \ast \sigma_j \ast \sigma_{j+1} \cdots \sigma_m)$$

B random (i, j)

$$B(x, y) = B(y, x) \ \frac{1}{\binom{m}{2}}$$

$$Y = (\sigma_1\ \sigma_2 \cdots \sigma_{i-1} \ast \sigma_j \ast \sigma_{i+1} \cdots \sigma_{j-1} \ast \sigma_i \ast \sigma_{j+1} \cdots \sigma_m)$$

each is simple. Each step only involves local change

a small / minor update

Rejection Step :

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

$$= \min\left(1, \frac{\exp(-\mathcal{E}(y))}{\exp(-\mathcal{E}(x))}\right)$$

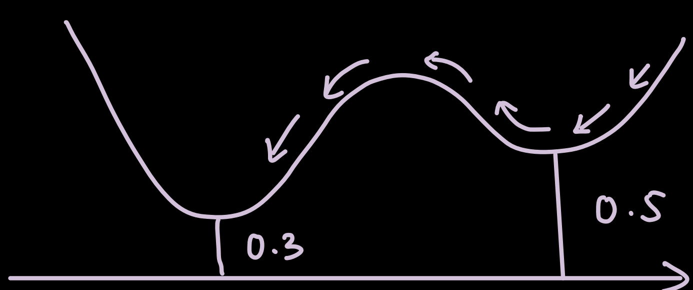
$$= \min\left(1, \exp(\mathcal{E}(x) - \mathcal{E}(y))\right)$$

$$\begin{aligned} \mathcal{E}(x) - \mathcal{E}(y) = & \; d(\sigma_{i-1}, \sigma_i) + d(\sigma_i, \sigma_{i+1}) \\ & + d(\sigma_{j-1}, \sigma_j) + d(\sigma_j, \sigma_{j+1}) \\ & - d(\sigma_i, \sigma_j) - d(\sigma_j, \sigma_{i+1}) \\ & - d(\sigma_{j-1}, \sigma_i) - d(\sigma_i, \sigma_{j+1}) \end{aligned}$$

If $\mathcal{E}(x) \geq \mathcal{E}(y)$, $P(\text{accept}) = 1$ → exploitation

If $\mathcal{E}(x) > \mathcal{E}(y)$, $P(\text{accept}) = e^{\mathcal{E}(x) - \mathcal{E}(y)}$

(e.g. 0.1) → exploitation



$\mathcal{E}(x)$

Simulated Annealing : temperature

$$\pi(x) = \frac{1}{Z_T} \exp\left(-\frac{\mathcal{E}(x)}{T}\right)$$

If $T \to \infty$, $\pi_T(x) \to$ Unif

If $T \to \infty$, $\pi_T(x) \to \begin{cases} 1 & x = \text{global min} \\ 0 & \text{otherwise} \end{cases}$

Example : Start from a high $T \overset{\text{e.g.}}{=} 1000$

Sample $\pi_T(x)$ by metropolis

gradually reduce T

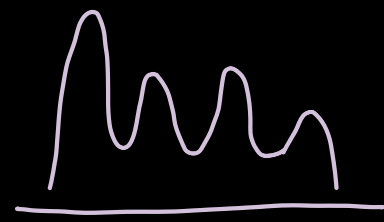
· annealing : putting hot metal in water to cool down

$$P(\text{acceptance}) = \begin{cases} 1 & , \ \mathcal{E}(x) \geq \mathcal{E}(y) \\ \exp\left(\frac{\mathcal{E}(x) - \mathcal{E}(y)}{T}\right) & , \ \mathcal{E}(x) > \mathcal{E}(y) \end{cases}$$

$$\to \begin{cases} 1 & , \text{ if } T \to \infty \\ 0 & , \text{ if } T \to 0 \end{cases}$$

What if we start from a low T?

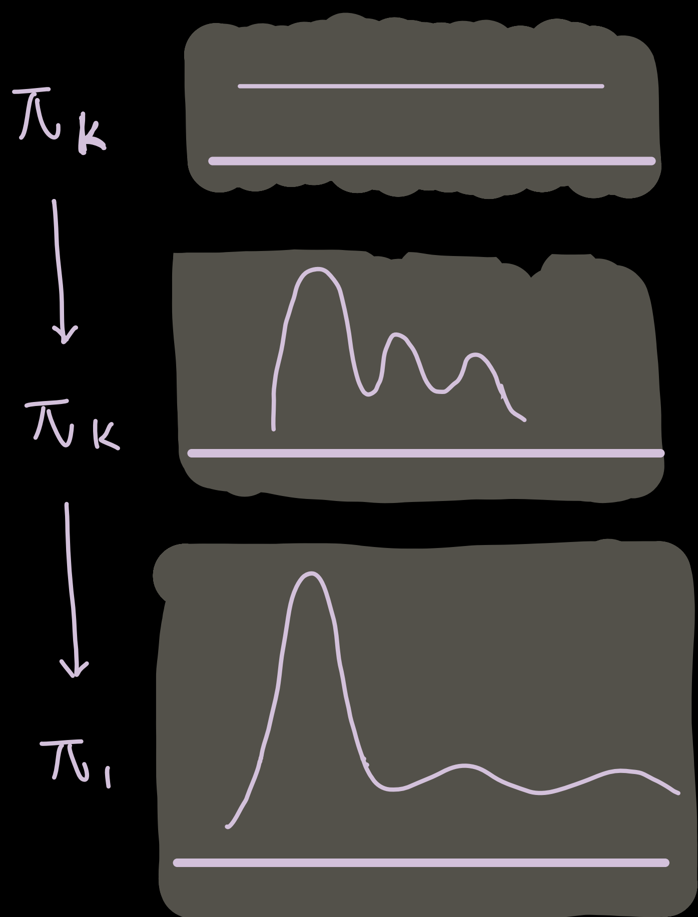
$\pi(x)$:



· Parallel Tempering / replicate exchange

we choose a set of temperatures

$$
\begin{array}{ccccc}
T_1 & T_2 & \cdots & T_k & \cdots & T_K \\
0.1 & 0.2 & & & & 100 \\
\downarrow & \downarrow & & & & \downarrow \\
\pi_1(x_1) & \pi_2(x_2) & & & & \pi_K(X_K) \approx \text{Unif}
\end{array}
$$



multi-modal

Replica : $(X_1 \quad X_2 \quad \dots \quad X_k \quad X_{k+1} \quad \dots \quad X_K)$

$\sim \pi_1(X_1)\ \pi_2(X_2)\ \dots\ \pi_k(X_k)\ \dots\ \pi_K(X_K)$

run K parallel chains by using $B(x,y)$

Replica Exchange move:

$$(X_k, X_{k+1}) \longrightarrow (X_{k+1}, X_k)$$

accept with p:

- A rigorous implementation of simulated annealing
- used in molecule dynamics

MCMC : each step $\rightarrow$ small change

- Gibbs Sampler $\pi(\vec{x})$

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \\ \vdots \\ x_d \end{pmatrix}$$

multivariate

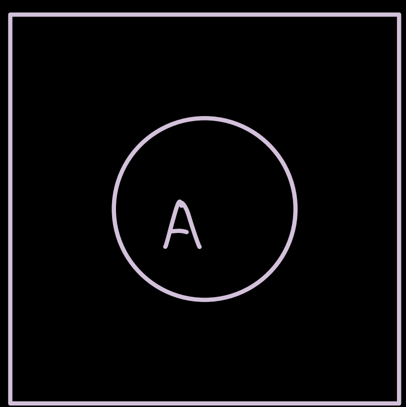
Small change

- randomly pick k
- change x_k while fixing
  $$x_{-k} = (x_1 \dots x_{k-1}\, x_{k+1} \dots x_d)$$
- update $x_k \sim \pi(x_k \mid x_{-k})$

(similar to coordinate descent)

Background on conditioning

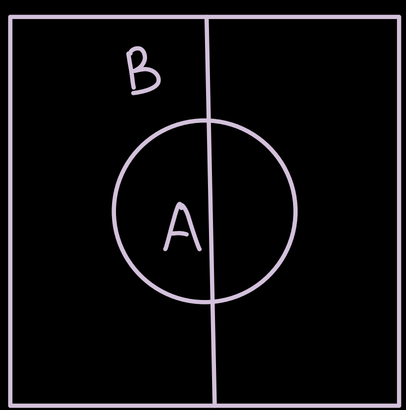
Ω



randomly throw a point into Ω

$$P(\text{falls into } A) = \frac{|A|}{|\Omega|}$$

Ω



If I tell you it falls into B

$$P(A \mid B) = \frac{|A \cap B|}{|B|}$$

think as if we randomly threw into B

$$= \frac{|A \cap B| \,/\, |\Omega|}{|B| \,/\, |\Omega|}$$

$$= \frac{P(A \cap B)}{P(B)}$$

Sample from a population

$X =$ gender
- 1 male
- 0 female

$Y =$ height
- 1 tall (≥ 6 ft.)
- 0 short (< 6 ft.)

Ω population

X \ Y		1 tall	0 short	
1	male	0.3	0.2	0.5
0	female	0.1	0.4	0.5
		0.4	0.6	

$$P_X(0) = P_X(1) = 0.5$$

$$\left.\begin{aligned} P_{Y|X}(1|1) &= P(Y=1|X=1) = \frac{0.3}{0.5} = 0.6 \\ P_{Y|X}(0|1) &= P(Y=0|X=1) = \frac{0.2}{0.5} = 0.4 \end{aligned}\right\}$$

Among males, distr. of Y

$$\left.\begin{aligned} P_{Y|X}(1|0) &= \frac{0.1}{0.5} = 0.2 \\ P_{Y|X}(0|0) &= \frac{0.4}{0.5} = 0.8 \end{aligned}\right\} \text{Among females}$$

$$P(x,y) = P(x)P(y|x)$$

$$P(y) = \sum P(x,y)$$

$$P(x|y) = \frac{P(x,y)}{P(y)}$$

Gibbs Sampler