$$\boxed{Prob(S) = e^{(f(S)/T)} / Z} \qquad (*) \qquad \frac{e^{f(S)/T}}{Z}$$

So, exponential distribution focuses probability mass on global maxima of $f(S)$, by lowering T.

But, why does SA converge to sampling from $(*)$?

DETAILS ON MARKOV CHAIN UNDERLYING SIMULATED ANNEALING.

Simulated Annealing (SA)

Consider the following "random walker" in hypercube space:

1) **Start at a random node S (the "current node").**

   (How do we generate such a node?)

2) **Select, at random, one of the N neighbors of S, call it S'**

3) **If (f(S') – f(S)) > 0, move to S', i.e. set S := S'**

   (i.e., jump to node with better value)

   else with probability e^(f(S')-f(S))/T move to S', i.e., set S := S'

4) **Go back to 2)**

$$\boxed{\text{Prob}(S) = e^{(f(S)/T)} / Z}$$

$$= \frac{e^{f(S')/T}}{e^{f(S)/T}}$$

$$= \frac{\dfrac{e^{f(S')/T}}{Z}}{\dfrac{e^{f(S)/T}}{Z}}$$

$$= \frac{\text{Prob}(S')}{\text{Prob}(S)} \quad (**)$$

So, we 'jump' based on ratio of desired probabilities.

Why does this process converge to desired global probability distribution?

$$\boxed{Prob(S) = e^{(f(S)/T)} / Z}$$

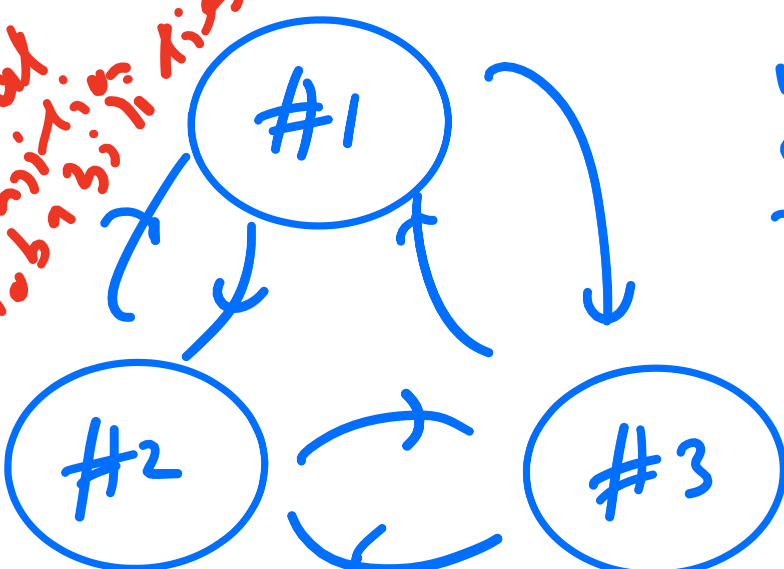
Demonstration by example. :)

(Metropolis–Hastings sampling. Clever strategy for sampling from complex distribution using local information only.)

MCMC: Markov Chain Monte Carlo

Consider particle jumping around between 3 states.

What transition probabilities?

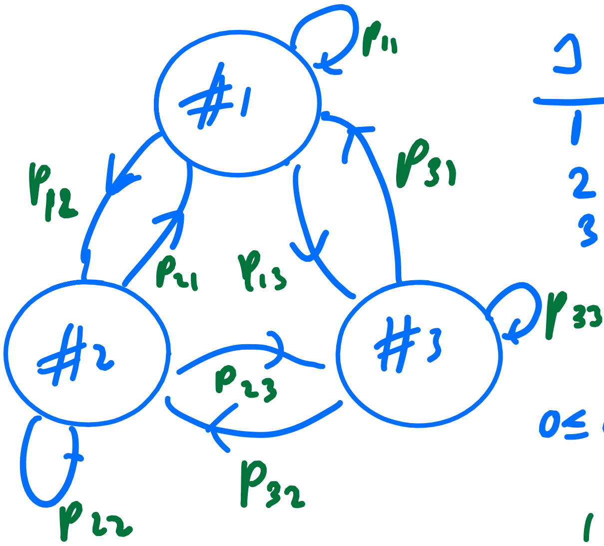


We want it in

State	Prob.
#1	α_1
#2	α_2
#3	α_3

Assume: $0 \le \alpha_1 \le \alpha_2 \le \alpha_3 \le 1$

$$\alpha_1 + \alpha_2 + \alpha_3 = 1$$



j	Prob(j)
1	α_1
2	α_2
3	α_3

desired

$0 \le \alpha_1 \le \alpha_2 \le \alpha_3 \le 1$

Sampler:

1) Start random state i
   current $= i$
2) Pick random neighbor j of current
3) If Prob(j) > Prob(i)
   current $= j$
   else with probability $\frac{\text{Prob}(j)}{\text{Prob}(i)}$ current $= j$
4) Back to 2)

jump to lower prob. state

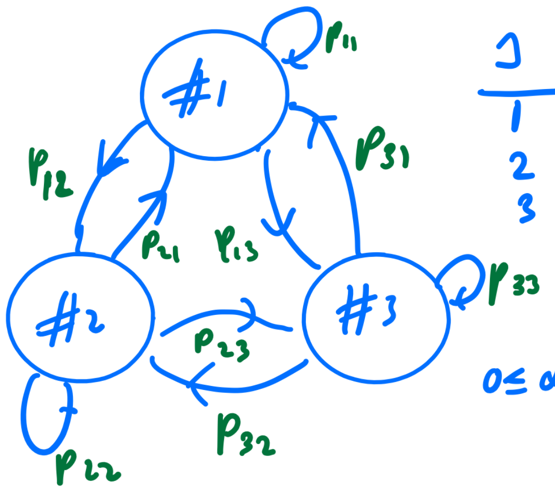
[Compare SA / **]

1) Start at a random node S (the "current node").
   (How do we generate such a node?)

2) Select, at random, one of the N neighbors of S, call it S'

3) If (f(S') – f(S)) > 0, move to S', i.e. set S := S'
   (i.e., jump to node with better value)
   else with probability e^(f(S')-f(S))/T move to S', i.e., set S := S'

4) Go back to 2)



s	Prob(s)
1	α_1
2	α_2
3	α_3

desired

$0 \le \alpha_1 \le \alpha_2 \le \alpha_3 \le 1$

1) Start random state i
   current $= i$
2) Pick random neighbor j of current
3) If Prob(j) > Prob(i)
   current $= j$
   else with probability $\frac{\text{Prob}(j)}{\text{Prob}(i)}$
   current $= j$
4) Back to 2)

Sample defines Markov Chain Transition Matrix —

(do this at home ?)

why? p_{11}

why? p_{12}

$$
\begin{array}{c|c}
 & j \rightarrow \\
i \downarrow &
\begin{array}{c|ccc}
 & \#1 & \#2 & \#3 \\
\hline
\#1 & 0 & \frac{1}{2} & \frac{1}{2} \\
\#2 & \frac{1}{2}\frac{\alpha_1}{\alpha_2} & 1-\frac{1}{2}-\frac{1}{2}\frac{\alpha_1}{\alpha_2} & \frac{1}{2} \\
\#3 & \frac{1}{2}\frac{\alpha_1}{\alpha_3} & \frac{1}{2}\frac{\alpha_2}{\alpha_3} & 1-\frac{1}{2}\frac{\alpha_1}{\alpha_3}-\frac{1}{2}\frac{\alpha_2}{\alpha_3}
\end{array}
\end{array}
$$

why? p_{11}

$$A = \begin{array}{c} \\ \#1 \\ \#2 \\ \#3 \end{array} \begin{array}{c} \xrightarrow{\;j\;} \\ \begin{array}{ccc} \#1 & \#2 & \#3 \end{array} \\ \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2}\frac{\alpha_1}{\alpha_2} & 1-\frac{1}{2}-\frac{1}{2}\frac{\alpha_1}{\alpha_2} & \frac{1}{2} \\ \frac{1}{2}\frac{\alpha_1}{\alpha_3} & \frac{1}{2}\frac{\alpha_2}{\alpha_3} & 1-\frac{1}{2}\frac{\alpha_1}{\alpha_3}-\frac{1}{2}\frac{\alpha_2}{\alpha_3} \end{pmatrix} \end{array}$$

What is special about transition matrix A?

Claim:

$$(\alpha_1 \;\; \alpha_2 \;\; \alpha_3) \cdot A = (\alpha_1 \;\; \alpha_2 \;\; \alpha_3)$$

i.e. distribution (stationary) we want is

<u>fixed</u> point of Markov Chain.

let's verify α_2 case.

$$(\alpha_1 \ \alpha_2 \ \alpha_3) \cdot A = (\alpha_1 \ \alpha_2 \ \alpha_3)$$

$$\begin{pmatrix} 0 & \boxed{\tfrac{1}{2}} & \tfrac{1}{2} \\ \tfrac{1}{2}\dfrac{\alpha_1}{\alpha_2} & \boxed{1 - \tfrac{1}{2} - \tfrac{1}{2}\dfrac{\alpha_1}{\alpha_2}} & \tfrac{1}{2} \\ \tfrac{1}{2}\dfrac{\alpha_1}{\alpha_3} & \boxed{\tfrac{1}{2}\dfrac{\alpha_2}{\alpha_3}} & 1 - \tfrac{1}{2}\dfrac{\alpha_1}{\alpha_3} - \tfrac{1}{2}\dfrac{\alpha_2}{\alpha_3} \end{pmatrix}$$

$$\overbrace{\tfrac{1}{2} \cdot \alpha_1 + \alpha_2 \left(\tfrac{1}{2} - \tfrac{1}{2} \dfrac{\alpha_1}{\cancel{\alpha_2}}\right.}^{x} \left. \vphantom{\dfrac{\alpha_1}{\alpha_2}}\right)$$

$$+ \cancel{\alpha_3} \cdot \tfrac{1}{2} \cdot \frac{\alpha_2}{\cancel{\alpha_3}} \qquad = \tfrac{1}{2}\alpha_2 + \tfrac{1}{2}\alpha_2$$

$$= \alpha_2 \quad \checkmark$$

ok :)

So, cleverly constructed sampler has desired stationary distribution $(\alpha_1, \alpha_2, \alpha_3)$.

SA is similar chain with stationary distribution:

$$\mathrm{Prob}(s) = \frac{e^{f(s)/T}}{Z}$$

$\boxed{\textbf{Prob(S) = e^(f(S)/T) / Z}}$

$\boxed{\textbf{Prob(S) = e^(f(S)/T) / Z}}$