Prob(S) = 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 (*)?
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)

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

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

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

$$= \frac{\text{Prob}(S')}{\text{Prob}(S)}$$
So, we 'jump' based on ratio of desired probabilities.

Why does this process converge to desired global probability distribution?

\[ \text{Prob}(S) = e^{(f(S)/T)} / Z \]
Demonstration by example.

(Neetropolis - Hastings)

sampling. (Good strategy for sampling from complex distribution using local information only.)

NMC: Markov Chain Monte Carlo

Consider particle jumps around between 3 states.

We want it in

<table>
<thead>
<tr>
<th>State</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>$\alpha_1$</td>
</tr>
<tr>
<td>#2</td>
<td>$\alpha_2$</td>
</tr>
<tr>
<td>#3</td>
<td>$\alpha_3$</td>
</tr>
</tbody>
</table>

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

$\alpha_1 + \alpha_2 + \alpha_3 = 1$
Sample:
1) Start random node i
   \( \text{current} = i \)
2) Pick random weight j of current
3) If \( \text{Prob}(j) > \text{Prob}(i) \)
   \( \text{current} = j \)
   else with probability \( \frac{\text{Prob}(j)}{\text{Prob}(i)} \)
   \( \text{current} = j \)
4) Back to 2

\[ \text{[compute SA / step]} \]

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 \( (\text{RF}(S') - \text{RF}(S)) > 0 \), move to \( S' \), i.e. set \( S := S' \)
   (i.e., jump to node with better value)
   else with probability \( e^{(\text{RF}(S') - \text{RF}(S)) / T} \) move to \( S' \), i.e., set \( S := S' \)

4) "Go back to 2"
Sample definition makes Cho-Chi
Transition Matrix
(chosen at hour 0)

\[
\begin{pmatrix}
0 & \frac{1}{2} & \frac{1}{2} \\
\frac{\alpha_1}{\alpha_2} & 1 - \frac{1}{2} - \frac{1}{2} \frac{\alpha_1}{\alpha_2} & \frac{1}{2} \\
\frac{\alpha_1}{\alpha_3} & \frac{\alpha_2}{\alpha_3} & 1 - \frac{1}{2} \frac{\alpha_1}{\alpha_3} - \frac{1}{2} \frac{\alpha_2}{\alpha_3}
\end{pmatrix}
\]

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

why? why? why?
What is special about transition matrix $A$?

Claim:

$(\alpha_1, \alpha_2, \alpha_3).A = (\alpha_1, \alpha_2, \alpha_3)$

i.e. distribution (stationary) we want is

fixed point of Markov Chain.
let's verify $\alpha_2$ case.

\[
\begin{pmatrix}
\alpha_1 & \alpha_2 & \alpha_3 \\
0 & 1/2 & 1/2 \\
1/2 & \frac{\alpha_1}{\alpha_2} & \frac{1}{2} - \frac{1}{2} \frac{\alpha_1}{\alpha_2} & 1/2 \\
1/2 & \frac{\alpha_2}{\alpha_3} & \frac{1}{2} & 1/2 - \frac{1}{2} \frac{\alpha_1}{\alpha_2} - \frac{1}{2} \frac{\alpha_2}{\alpha_3}
\end{pmatrix}
\]

\[
\frac{1}{2} \cdot \alpha_1 + \alpha_2 \left( \frac{1}{2} - \frac{1}{2} \frac{\alpha_1}{\alpha_2} \right) + \alpha_3 \cdot \frac{1}{2} \cdot \frac{\alpha_2}{\alpha_3} = \frac{1}{2} \alpha_2 + \frac{1}{2} \alpha_3
\]

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

SA is similar chain with stationary distribution:

\[
\text{Prob}(\sigma) = \frac{e^{f(\sigma)/T}}{Z}
\]

\[
\text{Prob}(S) = e^{(f(S)/T)} / Z
\]