Task: draw random samples from a "complicated" probability distribution.

Def: An unnormalized distribution on a finite set \( \Omega \) is a function
\[
\mu : \Omega \to [0, \infty)
\]
such that \( \sum_{w \in \Omega} \mu(w) > 0 \).

For any such \( \mu \), we can define
\[
Z(\mu) = \sum_{w \in \Omega} \mu(w)
\]
and
\[
\bar{\mu}(w) = \frac{1}{Z(\mu)} \mu(w).
\]
This \( \bar{\mu} \) is a probability distribution, the "normalization" of \( \mu \).

Typical situation:
- \( \Omega \) is exponentially large
- \( \mu(w) \) is easy to compute
- \( Z(\mu) \) is (believed) hard to compute.

When that happens, can we efficiently sample from \( \bar{\mu} \)?
Ex. 1. Sample a unit random \( q \)-coloring of a graph \( G = (V, E) \).
(Try to do this efficiently when \( q > \max \text{degree of } G \).)

\[ \Omega = \{ \text{all functions } V \to [q] \} \]

\[ \mu(x) = \begin{cases} 1 & \text{if } x \text{ is a proper coloring} \\ 0 & \text{if not} \end{cases} \]

Ex. 2. Given a "degree sequence" \( (d_1, d_2, ..., d_n) \), sample a uniformly random graph with vertex set \([n]\) in which \( \forall i \in [n] \), degree of vertex \( i \) is \( d_i \).

Def. A Markov chain with state set \( S \) is a first order Markov chain in infinite sequences \( (X_0, X_1, X_2, ...) \) of elements of \( S \), that satisfies

\[ \forall t \in \mathbb{N} \forall (x_0, ..., x_t) \\
\quad \Pr(X_t = x_t | X_0 = x_0, X_1 = x_1, ..., X_{t-1} = x_{t-1}) \]
If this probability depends on the values $x_t, x_{t-1}$ but not on $t$, we call the Markov chain time-homogeneous.

For time-homogeneous Markov chains, we can define the transition matrix $P$ by

$$p_{xy} = \Pr(X_t = x \mid X_{t-1} = y).$$

With this convention, if $\pi_t$ denotes the marginal distribution of $X_t$ then

$$\pi_t(x) = \sum_{y \in \mathbb{Y}} \Pr(X_{t-1} = y) \cdot \Pr(X_t = x \mid X_{t-1} = y)$$

$$= \sum_{y \in \mathbb{Y}} p_{xy} \cdot \pi_{t-1}(y).$$

Treating $\pi_t$ as a vector in $\mathbb{R}^{|\mathbb{Y}|}$,

$$\pi_t = \rho \cdot \pi_{t-1}.$$
Inductively  \( \pi_t = p^t \pi_0 \).

**Def.** A stationary distribution for \( P \) is a probability vector \( \pi \) such that  \( P \pi = \pi \).

**Q1:** Given unnormalized distribution \( \mu \), how can we design a Markov transition matrix \( P \) whose stationary distribution is \( \overline{\mu} \)?

**Q2:** For Markov chain \( P \), does a stationary distribution exist? Is it unique? If so, is \( \pi_t \overset{\text{st}}{\to} \pi \) guaranteed for every initial \( \pi_0 \)? And how fast is the convergence?

For answering Q1, the Metropolis-Hastings procedure.

**Def.** Markov chain \( P \) is reversible with distribution \( \pi \) if  
\[ \forall x,y : P_{xy} \overline{\pi}_y = P_{yx} \overline{\pi}_x \]
Lemma: If \( P \) is reversible w.r.t. \( \pi \), then \( \pi \) is a stationary distrib for \( P \).

Proof:
\[
\begin{align*}
(p_{\pi})_x &= \sum_y P_{xy} \pi_y \\
&= \left( \sum_y p_{yx} \right) \pi_x = \pi_x
\end{align*}
\]

Metropolis-Hastings requires:

1. Unnormalized distributions \( \mu \) - that we want to sample from

\[ K: \Omega \rightarrow [0,1] \]

"reference distrib that's easy to sample from"

Assume we have alg. to efficiently compute \( \mu(w), K(w) \) \forall w \in \Omega.

2. Markov chain \( K \) that's reversible w.r.t. \( \pi \).

Assume we can efficiently simulate one state transition of \( K \).
Often, \( K(w) = 1 \) \( \forall \omega \in \Omega \). Then reversibility of \( K \) just means \( K \) is a symmetric matrix.

Then the matrix \( P \) defined by

\[
P_{xy} = K_{xy} \cdot \min \left\{ \frac{\pi_x}{\pi_y}, 1 \right\} \cdot \pi_y + \pi_{xy}
\]

\[
P_{yy} = 1 - \sum_{x \neq y} P_{xy}
\]

is a Markov transition matrix and is reversible with \( \pi \mathcal{T} = \pi \).

\( P \) is non-negative because

\[
\sum_{x \neq y} P_{xy} = \sum_{x \neq y} K_{xy} \min \left\{ \frac{\pi_x}{\pi_y}, 1 \right\} \pi_y
\]

\[
\leq \sum_{x \neq y} K_{xy} \pi_y \leq \pi_y \leq 1.
\]

so \( P_{yy} \geq 0 \).

Reversibility:

\[
P_{xy} \pi_y = K_{xy} \min \left\{ \frac{\pi_x}{\pi_y}, 1 \right\} \pi_y
\]

\[
P_{yx} \pi_x = K_{yx} \min \left\{ \frac{\pi_y}{\pi_x}, 1 \right\} \pi_x
\]
The state transition of $M - H$:

1. Sample state $x$ with prob. $K_{xy}$.
2. Calculate $\min\left\{ \frac{k_x}{k_y}, 1 \right\}$. $K_y = p$.
3. With probability $p$ transition to state $x$. Else remain at $y$. 