$\boxed{\text{29 Nov 2021}}$ Markov Chains and Stationary Distributions

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

<u>Def.</u> An unnormalized distribution on a finite set Ω is a function
$$\mu : \Omega \longrightarrow [0, \infty)$$
such that $\sum_{\omega \in \Omega} \mu(\omega) > 0$.

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

Typical situation:
- Ω is exponentially large
- $\mu(\omega)$ 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 unif random q-coloring of a graph $G=(V,E)$.
(Try to do this efficiently when $q >$ max 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" $(d_1, d_2, \dots, d_n)$ 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 Ω is a prob. distrib. on infinite sequences $(X_0, X_1, X_2, \dots)$ of elements of Ω, that satisfies

$$\forall t \in \mathbb{N} \ \forall (x_0, \dots, x_t)$$

$$\Pr\left(X_t = x_t \mid X_0 = x_0, X_1 = x_1, \dots, X_{t-1} = x_{t-1}\right)$$

$$= \Pr(X_t = x_t \mid X_{t-1} = x_{t-1})$$

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 π_t denotes the marginal distribution of X_t then

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

$$= \sum_{y \in \Omega} P_{xy} \cdot \pi_{t-1}(y)$$

Treating π_t as a vector in $\mathbb{R}^{\Omega}$,

$$\pi_t = P \cdot \pi_{t-1}.$$

Inductively $\pi_t = P^t \pi_0$.

Def. A stationary distribution for P is a probability vector π such that $P\pi = \pi$.

Q1: Given unnormalized distribution μ, how can we design a Markov transition matrix P whose stationary distribution is $\bar{\mu}$?

Q2: For Markov chain P, does a stationary distrib exist? Is it unique? If so, is $\pi_t \to \pi$ guaranteed for every initial π_0? And how fast is the convergence?

For answering Q1, the Metropolis-Hastings procedure.

Def. Markov chain P is reversible w.r.t. distribution π if

$$\forall x,y \quad P_{xy}\pi_y = P_{yx}\pi_x$$

<u>Lem</u> If P is reversible w.r.t. π then π is a stationary distrib for P.

<u>Proof</u>.

$$(P\pi)_x = \sum_y P_{xy}\pi_y$$

$$= \left(\sum_y P_{yx}\right)\pi_x = \pi_x$$

Metropolis-Hastings requires:

1. Unnormalized distributions

   μ — that we want to sample from

   $K : \Omega \to [0,1]$

   "reference distrib that's easy to sample from"

   Assume we have algs. to efficiently compute $\mu(\omega)$, $K(\omega)$ $\forall \omega \in \Omega$.

2. Markov chain K that's reversible w.r.t. K.

   Assume we can efficiently simulate one state transition of K.

Often, $K(w) = 1 \quad \forall w \in \Omega$. Then reversibility of K just means K is a symmetric matrix.

Then the matrix P defined by

$$P_{xy} = K_{xy} \cdot \min\left\{\frac{\mu_x}{\mu_y}, 1\right\} \cdot K_y \quad \text{if } x \neq y$$

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

is a Markov transition matrix and is reversible w.r.t. $\pi = \bar{\mu}$.

P is non-negative because

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

$$\leq \sum_{x \neq y} K_{xy} K_y \leq K_y \leq 1.$$

so $P_{yy} \geq 0$.

<u>Reversibility:</u> $P_{xy}\pi_y = K_{xy} \min\{\pi_x, \pi_y\} K_y$

$P_{yx}\pi_x = K_{yx} \min\{\pi_y, \pi_x\} K_x$

One state transition of M.-H.:

1. Sample state x with prob. K_{xy}
2. Calculate $\min\left\{ \frac{\mu_x}{\mu_y}, 1 \right\} \cdot K_{y} = \rho$.
3. With probability ρ transition to state x. Else remain at y.