Note: Problems marked with a star (⋆) are required for graduate students to receive 4 credit hours; undergraduate students who solve these problems will receive extra credit points.

Submission: Write your name, netid, and U for undergrad/G for grad in the upper right-hand corner of the first page of your written solutions. Typewritten solutions will receive 5 extra credit points.

Problems to be handed in

1 In class, we saw how to generate a Bernoulli(p) random variable from a random variable U uniformly distributed on the unit interval $[0,1]$: if $0 \le U < p$, output 1; else, if $p \le U < 1$, output 0.

(i) Modify the above procedure for the case when, instead of $U \sim \text{Uniform}(0,1)$, we can generate a random variable V with cdf given by

$$F_V(a) = \begin{cases} 0, & a < 0 \\ a^2, & 0 \le a < 1 \\ 1, & a \ge 1 \end{cases} .$$

Hint: Can you find a function $f : \mathbb{R} \to \mathbb{R}$, such that $f(V) \sim \text{Uniform}(0,1)$?

(ii) Now suppose that we wish to generate a random variable X taking values in the finite set $\{1, \dots, n\}$ with probabilities $p_i \triangleq \mathbf{P}[X = i]$. Construct a procedure that takes $U \sim \text{Uniform}(0,1)$ as the input and generates a sample of X as the output.

(iii) Finally, modify your procedure from part (ii) to work with input V from part (i).

2 Consider a Markov chain $X = (X_t)_{t \in \mathbb{Z}_+}$ with finite state space $\mathsf{X} = \{1, \dots, n\}$, and let $M = [M(x,y)]_{x,y \in \mathsf{X}}$ be the matrix of its one-step transition probabilities. This constitutes a declarative description of X. Construct an imperative description of X in the form $X_{t+1} = f(X_t, U_t)$, where the U_t's are i.i.d. Uniform$(0,1)$ random variables.

3 Consider the following imperative description of a discrete-time stochastic signal with a finite state space X: We have a coin with bias (i.e., probability of coming up HEADS) p. At each time t, we toss the coin. Let $Z_t \in \{0,1\}$ denote the outcome of the toss (0 for TAILS, 1 for HEADS). Then

$$X_{t+1} = \begin{cases} f_0(X_t, U_t), & \text{if } Z_t = 0 \\ f_1(X_t, U_t), & \text{if } Z_t = 1 \end{cases}$$

where f_0 and f_1 are two given update rules, and where the initial state X_0, the i.i.d. random variables U_t, and the coin tosses Z_t are all mutually independent.

(i) Prove that the resulting signal $X = (X_t)_{t \in \mathbb{Z}_+}$ is a Markov chain.

(ii) For each pair of states $x, y \in \mathsf{X}$, and for $i \in \{0,1\}$, le $M_i(x,y) \triangleq \mathbf{P}[f_i(x, U_0) = y]$. Express the one-step transition probabilities of X in terms of p, M_0, and M_1.

4 Prove that any discrete-time Markov chain has the following property:

$$\begin{aligned}\mathbf{P}[X_{t_{n+1}} = y_1, X_{t_{n+2}} = y_2, \ldots, X_{t_{n+m}} = y_m | X_{t_1} = x_1, X_{t_2} = x_2, \ldots, X_{t_n} = x_n] \\ = \mathbf{P}[X_{t_{n+1}} = y_1, X_{t_{n+2}} = y_2, \ldots, X_{t_{n+m}} = y_m | X_{t_n} = x_n]\end{aligned}$$

for all $t_1 < t_2 < \ldots < t_n < t_{n+1} < \ldots < t_{n+m}$ and for all states $x_0, \ldots, x_t, y_1, \ldots, y_m$. This is why we say that, in a Markov chain, the future is conditionally independent of the past given the present.

Hint: Pass to a suitable imperative description of the form $X_{t+1} = f(X_t, U_t)$ and iterate.

5 ($\star$) The goal of this problem is to develop some basic tools for quantifying short-run and long-run behavior of Markov chains. Fix a finite state space $\mathsf{X} = \{1, \ldots, n\}$.

(i) For any two probability distributions $p = (p(x))_{x \in \mathsf{X}}$ and $q = (q(x))_{x \in \mathsf{X}}$ on X, define

$$d(p,q) \triangleq \frac{1}{2} \sum_{x \in \mathsf{X}} |p(x) - q(x)|.$$

Prove that d has the following properties:

(a) positivity and nondegeneracy — $d(p,q) \ge 0$, and $d(p,q) = 0$ if and only if $p = q$

(b) symmetry — $d(p,q) = d(q,p)$

(c) triangle inequality — for any three distributions p, q, and r,

$$d(p,q) \le d(p,r) + d(r,q).$$

Because d has these properties, it can be thought of as a *distance* between distributions. In fact, it has a name — it is called the *total variation distance*.

(ii) Prove that, for any Markov matrix M on X and for any two probability distributions p, q on X,

$$d(pM, qM) \le d(p,q).$$

If we think of p and q as two possible distributions for a starting state X_0 of a Markov chain with transition probability matrix M, then this inequality shows that the corresponding next-state distributions will be closer to one another.

(iii) Consider the "Google" matrix G defined in the lecture on the PageRank algorithm. Prove that it has the stronger property that

$$d(pG, qG) \le (1 - \alpha) d(p,q)$$

for any two distributions p, q on $\mathsf{X} = \{1, \ldots, n\}$.

(iv) Use the result of part (iii) to prove that G has a unique equilibrium distribution $r = rG$.

Hint: Fix an arbitrary initial distribution p_0 and consider the sequence of distributions $p_1 = p_0 G$, $p_2 = p_0 G^2, \ldots, p_t = p_0 G^t, \ldots$. Prove that this sequence tends to a limit, i.e., that there exists a probability distribution p^*, such that $d(p_t, p^*) \xrightarrow{t\to\infty} 0$. Next, prove that the limit p^* is an invariant distribution of G: $p^* = p^* G$. Finally, prove that p^* does not depend on the initial distribution p_0. Then $r = p^*$.

(v) We can now use these results to determine the number of iterations of PageRank required to approximate the solution to $r = rG$ to any desired accuracy. Let $\alpha = 0.15$, and determine the number of iterations N to guarantee that $d(r^{(N)}, r) \leq 0.05$ for any starting $r^{(0)}$.