Section 6 (Stat 171)

- TF: Jiaze Qiu
  - Email: jiazeqiug.harvard.edu
  - Section/OH: Wed 9-12 pm (ET)
- TF: Fernando Vicente
  - Email: fernando_vicente@fas.harvard.edu
  - Section/OH: Mon 7-9 pm and Thu 8-10 pm (ET)

- Acknolegement: This handout is partially based on notes created by Lisa Ruan and Christy Huo.

1 Review

We use $\pi(\dot{})$ to denote the target density.

- Metropolis-Hastings: pick proposal distribution $q(x, \cdot)$, and do the following.
  1. Pick initial state x_0.
  2. From $t = 0$ to $T - 1$:
     - draw $x^* \sim q(x_t, \cdot)$.
     - Let x_{t+1} be x^* with probability $\min\{1, \frac{\pi(x^*)q(x^*,x_t)}{\pi(x_t)q(x_t,x^*)}\}$, and be x_t otherwise.

  Under mild conditions, the empirical distributions of samples $x_{0:T}$ will approach π as $T \to \infty$. *You only need to be able to evaluate the density up to a normalizing constant in order to implement this.*

- Gibbs sampler: for π a distribution $\mathbb{R}^d$, we have d conditional distributions, denoted $\pi(x_j|x_{[-j]})$, i.e., distribution of a single coordinate conditional on all other coordinates. Here $x_{[-j]} = x_1, \dots, x_{j-1}, x_{j+}, \dots, x_d)$.
  1. Pick initial state $x^{(0)}$.
  2. Let $x^{(t+1)} = x^{(t)}$ - this is temporary; we will update all the elements of $x^{(t+1)}$
  3. From $t = 0$ to $T - 1$: update $x_1^{(t+1)} \sim \pi(x_1|x_[-1] = x_[-1]^{(t+1)}$

  Under mild conditions, the empirical distribution of samples of $x^{(0:T)}$ will approach π as $T \to \infty$.

- Total Variance (TV) distance: $d_{TV}(\pi, \pi') = \sup_{A\in\mathcal{S}} |\pi(A) - \pi'(A)|$ for π, π' probability distributions on state space $\mathcal{S}$. We let $\nu(n) = \max_{i\in\mathcal{S}} d_{TV}(P^n(i,\cdot), \pi)$ represents the worst case TV-distance to stationary distribution π over the choice of X_o.
- ϵ-Mixing time: $\inf\{n : \nu(n) < \epsilon\}$
- Spectral conditions for convergence of Markov Chain
  - Finite state space: $|\mathcal{S}| < \infty$
  - $\{X_n : n \geq 1\}$ with transition matrix P.
  - P is ergodic.
  - P is reversible with respect to the stationary distribution π.

  Then the chain has geometric convergence to the stationary distribution with rate driven by eigenvalues of the transition matrix P.

2 Problems

1. Total variation distance Prove that if the state space is final, i.e. $|\mathcal{S}| < \infty$, then the total variation distance can be calculated as:

$$d_{TV}(\pi, \pi') = \frac{1}{2} \sum_{i \in \mathcal{S}} |\pi_i - \pi'_i|$$

where $\pi_i = \pi(\{i\})$ and $\pi'_i = \pi'(\{i\})$.
Solution.
Let $B = \{i \in \mathcal{S} : \pi_i - \pi'_i > 0\}$.
Step 1: Prove that B gives an upper bound.
For any $A \subseteq \mathcal{S}$.

$$\pi(A) - \pi'(A) \leq \pi(B) - \pi'(B)$$

Thus B upper bounds d_{TV}. Moreover, since $B \subset \mathcal{S}$, $B = d_{TV}$.
Step 2 Notice $|\pi(B) - \pi'(B)| = |\pi'(B^c) - \pi(B^c)|$
Proof: $\pi(B) + \pi(B^c) = 1 = \pi(B)' + \pi(B^2)' \implies \pi(B) - \pi(B)' = \pi'(B) - \pi(B)$.
Step 3 Equate results to d_{TV}
Because state space is finite, $d_{TV}(\pi, \pi') = \max_{A \in \mathcal{S}} |\pi(A) - \pi'(A)|$ (i.e., we can use max instead of sup).

$$\begin{aligned}
d_{TV}(\pi, \pi') &= \max_{A \in \mathcal{S}} |\pi(A) - \pi'(A)| \\
&= \pi(B) - \pi'(B) \textit{ from Step 1} \\
&= \frac{1}{2}(\pi(B) - \pi'(B)) + \frac{1}{2}(\pi'(B^c) - \pi(B)) \textit{ from step 2} \\
&= \frac{1}{2}|\pi(B) - \pi'(B)| + \frac{1}{2}|\pi'(B^c) - \pi(B)| \textit{ the terms in the parenthesis were net positive to begin with} \\
&= \frac{1}{2} \sum_{i \in \mathcal{S}} |\pi_i - \pi'_i|
\end{aligned}$$

2. Finite Markov Chain: Ergodic and Regular We want to prove that a **finite** Markov Chain is ergodic if and only if its transition matrix $\mathbf{P}$ is regular.
(a) Prove that $\mathbf{P}$ regular implies ergodic.

Solution.

For some $N > 0$, $\mathbf{P}^N \geq 0$, i.e., it has all positive entries. It follows that the Markov Chain is **irreducible**.
Since $\mathbf{P}^n \geq 0$ for all $n \geq N$, then $P_{ii}^n > 0$ and $P_{ii}^{n+1} > 0$, which implies that $\gcd(i) = 1$ for all $i \in \mathcal{S}$. Thus the Markov Chain is **aperiodic**.
Since the Markov Chain is finite, irreducible, and aperiodic, it follows that it is an ergodic chain.

(b) Prove that if i is an aperiodic state, there exists a $N > 0$ such that $P_{ii}^n > 0$ for all $n \geq N$
Solution.
Let i be an aperiodic state i.e. $\gcd(i) = 1$.
Let $T_i = \{n : P_{ii}^n > 0\}$ be the set of all possible return times.
Step 1: Obtain smallest element of T_i. Call this element k. Assume that there is not a N such that $P_{ii}^n > 0$ for all $n \geq N$. It follows that there must be a smallest $k \in T_i$, $k \geq 2$, such that $|m - n| \leq k$ for all $m, n \in T_i$.
Claim 1 T_i is closed under addition, i.e., $n, m \in T_i$ implies $n + m \in T_i$. Proof:

$$P_{ii}^{m+n} \geq P_{ii}^m P_{ii}^n > 0.$$

Step 2: Generate two numbers that are in T_i that are closer than k.
Since $\gcd(T_i) = 1$, there is $n, m \in T_i$ such that n does not share a common divisor with $m+k$. Since k is the smallest element in T_i, $k < n$, and we can write $n = kq + r$ for $q \in \mathbb{N}$ and $0 < r < k$.
Since T_i is closed under addition, $(q + 1)(m + k), n + (q + 1)m \in T_i$. The difference in these numbers is

$$(q + 1)(m + k) - n + (q + 1)m = q(k + 1) - n = q(k + 1) - qk - r = k - r < k.$$

Thus we have two elements in T_i that are closer than k steps. This is a contradiction, and thus there must be a N such that $P_{ii}^n > 0$ for all $n \geq N$.

(c) Prove that an ergodic chain implies a regular transition matrix.

Solution.

An ergodic chain means the chain is irreducible and aperiodic (we assume the chain is finite).

Since the chain is aperiodic, for each state $i \in \mathcal{S}$, there exists a N_i such that $P_{ii}^n > 0$ for all $n \geq N_i$ by problem (b) above. Set $N = \max_{i \in \mathcal{S}} N_i$ to be the step number after which all states return to themselves with positive probability i.e. the diagonal entries of $\mathbf{P}^n$ are positive for all $n \geq N$.

Since the chain is irreducible, for each $i, h \in \mathcal{S}$ there is $m_{ij} \in \mathbb{N}$ such that $P_{ij}^{m_{ij}} > 0$. Set $M = \max_{i,j \in \mathcal{S}} m_{ij}$.

Set $X = M + N$. For any i, j

$$P_{ij}^X \geq P_{ii}^{N+M-m_{ij}} P_{ij}^{m_{ij}} > 0$$

since $N + M - m_{ij} > N$. Thus $\mathbf{P}^n$ has positive entries for all $n \geq X$, and $\mathbf{P}$ is a regular transition matrix.