7 Introduction to Markov Chains

Suppose that we observe the evolution of a system, described through discrete random variables $X_0, X_1, ..., X_n, ...$, where each $X_t$ takes values from the set of states $\{1, 2, ..., N\}$, the possible states of the system at any time. If there exist $p_{ij}, i, j = 1, 2, ..., N$, such that

$$p_{ij} = P(X_{n+1} = j|X_n = i) = P(X_{n+1} = j|X_n = i, X_{n-1}, ..., X_0),$$

then we say that the collection $\{X_n : n \geq 0\}$ forms a Markov Chain with transition probabilities $p_{ij}$, $i, j = 1, 2, ..., N$, where we must have $\sum_{j=1}^{N} p_{ij} = 1, i = 1, 2, ..., N$, since these are all the possible states for the chain to evolve into after it leaves state $i$, that is these are all the possible states for $X_{n+1}$.

Now a Markov Chain is said to be irreducible if for each pair of states $i$ and $j$ there is a positive probability that starting at state $i$, the chain will eventually reach state $j$. We let $\pi_j$ denote the long-run probability that the chain process is in state $j$. The $\pi_j$'s can be shown to be the unique solutions to the set of linear equations

$$\pi_j = \sum_{i=1}^{N} \pi_i p_{ij}, \quad j = 1, 2, ..., N,$$

where of course $\sum_{j=1}^{N} \pi_j = 1$. Hence we may write the equations as

$$\pi_j = \sum_{i=1}^{N-1} \pi_i p_{ij} + (1 - \pi_1 - \pi_2 - ... - \pi_{N-1}) p_{Nj} \Rightarrow$$

$$\pi_j = \sum_{i=1}^{N-1} \pi_i (p_{ij} - p_{Nj}) \Rightarrow$$

$$\pi_1 (p_{1j} - p_{Nj}) + \pi_2 (p_{2j} - p_{Nj}) + ... + \pi_{j-1} (p_{j-1j} - p_{Nj}) + \pi_j (p_{jj} - p_{Nj} - 1) + \pi_{j+1} (p_{j+1j} - p_{Nj}) + ... + \pi_{N-1} (p_{N-1j} - p_{Nj}) = -p_{Nj},$$

for $j = 1, 2, ..., N - 1$.

If we generate $X_0$ according to the $\{\pi_j\}$, then the $\{\pi_j\}$ are referred to as the stationary probabilities of the Markov Chain, and it can be shown that

$$\pi_j = P(X_n = j), \quad \text{for all } n \text{ and } j.$$

The $\{\pi_j\}$ are often interpreted as the limiting probability that the chain is in state $j$, i.e.,

$$\pi_j = \lim_{n \to +\infty} P(X_n = j), \quad \text{for all } j,$$

provided that the chain is irreducible and aperiodic. In this case the $\{\pi_j\}$ are called often the equilibrium distribution.

A chain is said to be aperiodic if for some $n \geq 0$ and some state $j$ it holds

$$P(X_n = j|X_0 = j) > 0 \text{ and } P(X_{n+1} = j|X_0 = j) > 0.$$

Finally, when $\pi_i p_{ij} = \pi_j p_{ji}$, for all $i \neq j$, the Markov chain is said to be time reversible, since under this condition, if we choose the initial state according to $\{\pi_j\}$, then at any time $n$, the reverse sequence $X_n, X_{n-1}, ..., X_0$, will also be a Markov chain with transition probabilities $p_{ij}$.
7.1 Introduction to Markov Chain Monte Carlo

So now suppose we want to generate values from a discrete random variable with p.m.f. $\pi_j = P(X = j)$, $j = 1, 2, \ldots, N$. If we can somehow create an irreducible aperiodic Markov Chain with limiting distribution \{$\pi_j$\}, then we could run the chain for $n$ steps, where $n$ is large, and approximate a realization from the discrete p.m.f. using $X_n$. Moreover, we can use realizations we obtain after a large step $k$ (which can be thought as approximate generated values from the target distribution) to estimate for example $E(g(X)) = \sum_{i=1}^{N} g(i)\pi_i$, using

$$E(g(X)) \approx \frac{1}{n-k} \sum_{i=k+1}^{N} g(X_i).$$

The problem with the early states of the chain (the first $k$ for instance as above) is that they are heavily influenced by the initial value $X_0 = j$. Hence, we drop them. Other people like to use what is called the burn-in, namely use the first $k$ generated states to estimate a starting state for the chain (e.g., take the average of the first $k$ states and start the chain again with $X_0$ being this average). There is no theoretical justification for the burn-in, it is rather something intuitive. If the chain is indeed an irreducible aperiodic Markov Chain where we start doesn’t matter.

Thus, the combination of Markov Chains and Monte Carlo as we discussed above, gave rise to the widely used approach called Markov Chain Monte Carlo (MCMC), where one constructs a Markov Chain of random variables $X_0, X_1, \ldots$, that has equilibrium a given distribution $F$. Thus, the resulting $X_i$ values are only approximately distributed according to $F$, and furthermore they are dependent. Still, the values of the chain may be used to compute estimates of $E(g(X))$, for any $g(x)$.

Markov Chain Monte Carlo is best suited for high dimensional multivariate distributions where it is difficult to find other methods of generating values, such as rejection samplers with high acceptance probabilities and so forth. Although, we only discussed the discrete case for illustration, we will proceed to discuss both discrete and continuous cases, with the latter being defined in a similar fashion as usual.