Introduction to Markov Chains: Discrete spaces

1. **Stochastic Process:** A stochastic process is an indexed collection of random variables $\{X(t) : t \in T\}$, where T is an index set, with a particular *law* or distribution for any finite *extract* of the above collection. Note that any simple sequence of random variables $\{X_n\}$ is a stochastic process with index set T being the set of non-negative integers, as long as the distribution of any finite set $(X_{i_1}, \ldots, X_{i_k})$ is well-defined.

2. **Discrete time stochastic process:** A sequence $\{X_n\}$ of random variables with values in a set E is called a *discrete-time stochastic process* with *state-space* E. We will mostly (if not exclusively) be dealing with countable state-spaces. So, without loss of generality, we will assume $E = \{1, 2, 3, \ldots\}$.

3. **Markov Chains:** Let $\{X_n\}_{n\geq 0}$ be a discrete-time stochastic process with countable state-space E. If, for all integers $n \geq 0$ and all states $i_0, i_1, \ldots, i_{n-1}, i, j$ in E,

$$P(X_{n+1} = j|X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_{n+1} = i|X_n = j),$$

   whenever both sides are well-defined, this stochastic process is called a *Markov Chain*. It is called a *homogeneous Markov Chain* (abbreviated as HMC) if in addition, the right hand side in the above equation is independent of n.

4. **Transition probabilities** of a Markov Chain:

$$p_{ij}^{(n)} = P\left(X_{n+1} = j|X_n = i\right), \text{ where } i, j \in E.$$

5. **The Transition matrix** of a Markov Chain has its $(i, j)^{th}$ element as p_{ij}. Thus,

$$[P]_{ij} = p_{ij}$$

   The matrix dimension of P is finite or infinite depending upon cardinality of E. We will denote P_{i*} as the i^{th} row and P_{*j} as the j^{th} column of the transition matrix.

6. Since its entries are probabilities, and each row is conditioned on the same state i, each element of the transition matrix is "non-negative" and the sum of the elements of each row is 1:

$$p_{ij} \geq 0; \sum_{k\in E} p_{ik} = 1.$$

We say that the transition matrix is a *non-negative* and *row-stochastic* matrix. In linear algebraic terms, we write this as:

$$P\mathbf{1} = \mathbf{1}.$$

Remark: $\mathbf{1}$ is an eigenvector of P corresponding to the eigenvalue 1. Study of finite Markov Chains has a very close relationship with linear algebra of non-negative matrices. We will explore some of this later.

7. **Initial state and its distribution:** The random variable X_0 denotes the start of the Markov Chain and represents the *initial state*. The corresponding distribution $\nu_0(i_0) = P(X_0 = i_0)$ is called the *initial distribution*.

8. **Distribution of an HMC:** The distribution of an HMC is determined by its initial distribution $\nu(i_0)$ and its transition matrix P. This is clearly because the joint distribution of the Markov Chain until any time point k is given by:

$$P(X_0 = i_0, X_1 = i_1, \ldots, X_k = i_k) = \nu_0(i_0)p_{i_0 i_1} \cdots p(i_{k-1} i_k).$$

9. **Marginal distributions:** The distribution at time n of the chain is the vector with i-th element given by the marginal distribution of X_n:

$$\nu_n(j) = P(X_n = j).$$

Using simple probability rules we arrive at a recursive relationship:

$$\begin{aligned}
\nu_n(j) = P(X_n = j) &= \sum_{i \in E} P(X_{n-1} = i) P(X_n = j | X_{n-1} = i) \\
&= \sum_{i \in E} \nu_{n-1}(i) p_{ij} \\
&= \boldsymbol{\nu}_{n-1}^T P_{*j},
\end{aligned}$$

where $\boldsymbol{\nu}_{n-1}^T = (\nu_{n-1}(1), \nu_{n-1}(2), \ldots)$ and P_{*j} is the j^{th} column of P. In this notation, we can further write:

$$\boldsymbol{\nu}_n = \boldsymbol{\nu}_{n-1}^T P.$$

Iteration of this relationship yields:

$$\boldsymbol{\nu}_n = \boldsymbol{\nu}_0^T P^n.$$

10. **The n-step transition matrix:** The matrix P^n is called the n-step transition matrix because its $(i, j)^{th}$ term is given by:

$$p_{ij}(n) = P(X_n = j | X_0 = i).$$

Note that, by virtue of the homogenous Markov property, the above is equal to $P(X_{m+n} = j | X_m = i)$ for any time-point m. Indeed, using simple conditional probability rules and the Markov property:

$$\begin{aligned}
[P^n]_{ij} = P^{n-1}_{i*} P_{*j} &= \sum_{i_{n-1} \in E} [P^{n-1}]_{i i_{n-1}} p_{i_{n-1} j} \\
&= \sum_{i_{n-1}} (P^{n-2}_{i*} P_{* i_{n-1}}) p_{i_{n-1} j} = \sum_{i_{n-1}} \sum_{i_{n-2}} [P^{n-3}]_{i i_{n-2}} p_{i_{n-2} i_{n-1}} p_{i_{n-1} j} \\
&= \cdots\cdots \\
&= \sum_{i_{n-1}, \ldots, i_1 \in E} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} j} \\
&= \sum_{i_{n-1}, \ldots, i_1 \in E} P(X_n = j, X_{n-1} = i_{n-1} \ldots, X_1 = i_1 | X_0 = i_0) = P(X_n = j | X_0 = i).
\end{aligned}$$

11. **Chapman-Kolmogorov equations:** The Chapman-Kolmogorov equations concern the n-step and m-step transition matrices and is a simple consequence of the associative nature of matrix multiplication:

$$P(X_{n+m} = j | X_0 = i) = \sum_{k \in E} P(X_n = k | X_0 = i) P(X_m = j | X_0 = k).$$

This is a simple consequence of the matrix identity:

$$P^{n+m} = P^n P^m,$$

but here is a direct verification:

$$P(X_{n+m} = j | X_0 = i) = \sum_{k \in E} P(X_{n+m} = j | X_n = k) P(X_n = k | X_0 = i) = \sum_{k \in E} P(X_m = j | X_0 = k) P(X_n = k | X_0 = i).$$

Note how the *time-homogeneity* was used to write $P(X_{n+m} = j | X_n = k)$ as $P(X_m = j | X_0 = k)$.

12. **HW:** Prove that the Markov property extends to:

$$P(X_{n+1} = j_1, \ldots, X_{n+k} = j_k | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_{n+1} = j_1, \ldots, X_{n+k} = j_k | X_n = i)$$

for all $i_0, \ldots, i_{n-1}, i, j_1, \ldots, j_k$.

13. Writing,

$$A = \{X_{n+1} = j_1, \ldots, X_{n+k} = j_k\},\ B = \{X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\}$$

the HW result reads $P(A|X_n = i, B) = P(A|X_n = i)$, which in turn is equivalent to:

$$P(A \cap B|X_n = i) = P(A|X_n = i)P(B|X_n = i).$$

In words: The future at time n and the past at time n are conditionally independent given the present state $X_n = i$. Thus, the Markov property is *independent of the direction of time.*

14. **Example:** 2-state Markov Chain: $E = \{0, 1\}$

$$P(X_{n+1} = 1|X_n = 0) = p$$

$$P(X_{n+1} = 0|X_n = 1) = q$$

and let $P(X_0 = 0) = \nu_0(0)$. The transition matrix is:

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}.$$

Now, we have,

$$\begin{aligned}\nu_{n+1}(0) = P(X_{n+1} = 0) &= (1-p-q)\,P(X_n = 0) + q = (1-p-q)\nu_n(0) + q \\ &= (1-p-q)^{n+1}\nu_0(0) + q\sum_{j=0}^{n}(1-p-q)^j\end{aligned}$$

If $p = q = 0$, $P(X_{n+1} = i) = \nu_0(i)$ for $i \in \{0, 1\}$. If $p + q > 0$, then

$$\sum_{j=0}^{n}(1-p-q)^j = \frac{1-(1-p-q)^{n+1}}{p+q},$$

so that

$$\nu_{n+1}(0) = \frac{q}{p+q} + (1-p-q)^{n+1}\left(\nu_0(0) - \frac{q}{p+q}\right),$$

and consequently that:

$$\nu_{n+1}(1) = \frac{p}{p+q} + (1-p-q)^{n+1}\left(\nu_0(1) - \frac{q}{p+q}\right).$$

Now, if p and q are neither both 0 nor both 1, we have $0 < p + q < 2$ which means $|1 - p - q| < 1$, so we can take $n \to \infty$ above to obtain:

$$\lim_{n\to\infty}\nu_{n+1}(0) = \frac{q}{p+q};\ \lim_{n\to\infty}\nu_{n+1}(1) = \frac{p}{p+q}.$$

15. **Theorem: HMC's driven by white-noise** Let $\{Z_n\}_{n\geq 1}$ be an i.i.d. sequence of random variables with values in an arbitrary space F. Let E be a countable set and $f : E \times F \to E$ be any function. Let X_0 be a random variable with values in E independent of the Z_n's. Then the recurrence equations:

$$X_{n+1} = f(X_n, Z_{n+1})$$

defines a homogeneous Markov Chain.

Proof: We first need to argue that this is a Markov Chain. That is, X_{n+1} is conditionally independent of $X_0, \ldots, X_{n-1}$ given X_n. But from the recursion, it is clear that X_{n+1} depends only upon X_n and Z_{n+1}. Thus, X_{n+1} will be a Markov Chain if we can show that Z_{n+1} is independent of $X_0, \ldots, X_n$. Note the recursion:

$$\begin{aligned}
X_1 &= f(X_0, Z_1) = g_1(X_0, Z_1), \\
X_2 &= f(X_1, Z_2) = f(f(X_0, Z_1), Z_2) = g_2(X_0, Z_1, Z_2), \\
\ldots &= \ldots = \ldots = \ldots, \\
X_n &= f(X_{n-1}, Z_n) = \ldots = g_n(X_0, Z_1, \ldots, Z_n).
\end{aligned}$$

Thus, X_n is completely determined by $X_0, Z_1, \ldots, Z_n$ all of which are independent of Z_{n+1}. Therefore, Z_{n+1} is independent of X_n and, in fact, of any $X_k, k \leq n$. So, X_n's are a Markov Chain.

More formally, we have:

$$\begin{aligned}
P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) &= P(f(i, Z_{n+1}) = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \\
&= P(f(i, Z_{n+1}) = j) = P(X_{n+1} = j | X_n = i),
\end{aligned}$$

since Z_{n+1} is independent of $X_0, \ldots, X_n$.

To show that this $\{X_n\}$ is homogeneous, note that Z_n's are i.i.d. and so have the same distribution. Therefore,

$$P(f(i, Z_{n+1}) = j) = P(f(i, Z_1) = j)$$

which is independent of n. Thus, we have a transition matrix P with $(i,j)^t h$ element:

$$p_{ij} = P(f(i, Z_1) = j).$$

16. **Constructing a Markov Chain from a given transition matrix:** Consider any transition matrix P on E. Then we can *construct* or simulate a Markov Chain corresponding to this transition matrix with any starting distribution ν_0, using the above result, namely

$$X_{n+1} = \sum_{j \in E} jI\left(Z_{n+1} \in \left[\sum_{k=0}^{j-1} p_{X_n k}, \sum_{k=0}^{j} p_{X_n k}\right]\right),$$

where $\{Z_n\}_{n\geq 1}$ are i.i.d. $U(0,1)$ and $X_0 \sim \nu_0$. It is easy to use the above Theorem and check that this HMC indeed has the desired transition matrix. [**Convince yourself!**]

17. **HW: Prove the following Theorem** Let $\{Z_n\}_{n\geq 1}$ be an i.i.d. sequence of random variables with values in an arbitrary space F. Let E be a countable set and $f : E \times F \to E$ be any function. Let X_0 be a random variable with values in E. Suppose that for all $n \geq 0$, Z_{n+1} is conditionally independent of $Z_n, \ldots, Z_1, X_{n-1}, \ldots, X_0$ given X_n. More precisely, for all $k, k_1, \ldots, k_n \in F$, $i_0, i_1, \ldots, i_n \in E$

$$P(Z_{n+1} = k | X_n = i, \ldots, X_0 = i_0, Z_n = k_n, \ldots, Z_1 = k_1) = P(Z_{n+1} = k | X_n = i),$$

where the latter quantity is independent of n.

Then $\{X_n\}_{n\geq 0}$ is an HMC with transition matrix P given by

$$p_{ij} = P(f(i, Z_1) = j | X_0 = i).$$

18. **Examples of Markov Chains** We will explore some examples of Markov Chains, casting them in the framework of the HMC-Noise Theorem.

    (a) **One-dimensional Random Walk** Let X_0 be a random variable with values in the space of integers. Let $\{Z_n\}_{n\geq 1}$ be a sequence of i.i.d. random variables, independent of X_0, taking the values $+1$ or -1 and with the probability distribution:

    $$P(Z_n = +1) = p,$$

    where $p \in (0,1)$. The process X_n is defined by the recursion:

    $$X_{n+1} = X_n + Z_{n+1}.$$

    (b) **Repair shop** During day n, Z_{n+1} machines break down, and all of them enter the repair shop on day $n+1$. Each day one machine among those waiting for service is repaired. Let X_n be the number of machines in the shop on day n. Then,

$$X_{n+1} = (X_n - 1)^+ + Z_{n+1},$$

where $a^+ = \max(a, 0)$. *So, when is this a Markov Chain?* If Z_n is an i.i.d. sequence independent of the initial state X_0, then $\{X_n\}$ is an HMC. We will need to specify the probability distribution of the Z_n's:

$$P(Z_1 = k) = a_k, \; k \geq 0$$

which results in the transition matrix:

$$\begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \cdots \\ a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & a_0 & a_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

(c) **The Ehrenfest Chain** This famous Markov Chain was formulated by Austrian physicists Tatiana and Paul Ehrenfest to describe the diffusion of molecules through a porous membrane and the exchange of heat between two systems at different temperatures. Suppose there are N particles that can be either in compartment A or in compartment B. Suppose that at time $n \geq 0, \; X_n = i$ particles are in compartment A. One then chooses a particle at random, and this particle is moved at time $n+1$ from where it is to the other compartment. Thus,

$$P(X_{n+1} = i - 1 | X_n = i) = \frac{i}{N};$$
$$P(X_{n+1} = i + 1 | X_n = i) = 1 - \frac{i}{N}.$$

Obviously this is an HMC, but *how does this relate to the HMC-Noise framework?* Well, it turns out that is better explained by the extension of that theorem in your HW (bullet 17). For all $n \geq 0$,

$$X_{n+1} = X_n + Z_{n+1},$$

where $Z_n \in \{-1, +1\}$ and

$$P(Z_{n+1} = -1 | X_n = i) = \frac{i}{N}.$$

The non-zero entries of the transition matrix are therefore,

$$p_{i,i+1} = 1 - \frac{i}{N},\ p_i i - 1 = \frac{i}{N}.$$

(d) **Branching Process: The Galton-Watson process** This is a famous model that is used in population genetics and evolutionary studies. Let $\{Z_n\}_{n\geq 1}$ be a *vector sequence*, so $Z_n = (Z_{n1}, Z_{n2}, \dots)$ where $\{Z_{nj}\}$ are i.i.d. and integer valued. The recurrence equation

$$X_{n+1} = \sum_{k=1}^{X_n} Z_{n+1,k},$$

with the convention $X_{n+1} = 0$ if $X_n = 0$. Thus, X_n is the number of individuals of a given population (humans, particles, etc.). The j^{th} individual in the n^{th} generation gives birth to $Z_{n+1,j}$ descendants.

19. **A useful tool: the probability generating function.** Probability generating functions (pgf) are useful transforms (functionals) associated with a random variable as:

$$\psi_X(t) = E[t^X].$$

Other popular transforms are moment-generating functions, Laplace transforms and characteristic functions.

20. **The Galton-Watson recursion.** We return to the Galton-Watson setting. Let g be the pgf of the i.i.d. random variables Z_{nj} and let $\psi_n(t)$ be that of X_n. Then,

$$\psi_{n+1}(t) = E[t^{X_{n+1}}] = E\left[t^{\sum_{k=1}^{X_n} Z_{n+1,k}}\right] = E\left[\prod_{k=1}^{X_n} E[t^{Z_{n+1,k}}|X_n]\right].$$

But note that since X_n depends only upon $Z_1, \dots, Z_n$ which are independent of Z_{n+1}, X_n is independent of $Z_{n+1,k}$. Therefore, using the common generating function g, we have

$$\psi_{n+1}(t) = E\left[\prod_{k=1}^{X_n} E[t^{Z_{n+1,k}}|X_n]\right] = E\left[\prod_{k=1}^{X_n} E[t^{Z_{n+1,k}}]\right] = E\left[g(t)^{X_n}\right] = \psi_n(g(t)).$$

Iterating this recursion all the way yields $\psi_{n+1}(t) = \psi_0(g^{(n+1)}(t))$, where $\psi_0(t)$ is the pgf of X_0 and $g^{n+1}(t)$ is $(n+1)^{th}$ iterate of g. Note that if there is only one ancestor then $\psi_0(t) = t$. In that case, $\psi_{n+1}(t) = g^{(n+1)}(t) = g(g^n(t))$, so:

$$\psi_{n+1}(t) = g(\psi_n(t)).$$

21. **The Galton-Watson extinction probabilities.** Since

$$\psi_n(t) = E[t^{X_n}] = P(X = 0) + E[t^{X_n} I(X_n > 0)],$$

we have $\psi_n(0) = P(X_n = 0)$. Therefore, using the Galton-Watson recursion with only one ancestor, we have

$$P(X_{n+1} = 0) = g(P(X_n = 0)).$$

Define $\mathcal{E}$ as the event that "an extinction occurs" which means that at least one generation is empty. Hence, defining $E_n = [X_n = 0]$,

$$\mathcal{E} = \cup_{n=0}^{\infty} E_n.$$

Note that $E_n \uparrow \mathcal{E}$, and so, by the monotone property of the probability measure,

$$P(\mathcal{E}) = \lim_{n \uparrow \infty} P(E_n).$$

The generating function $g(t)$ is continuous in t and since, $P(E_{n+1}) = g(P(E_n))$, using the continuity of g we have

$$P(\mathcal{E}) = \lim P(E_{n+1}) = \lim g(P(E_n)) = g(\lim P(E_n)) = g(P(\lim E_n)) = g(P(\mathcal{E})).$$

22. **Analyzing $g(t)$:** Consider a non-negative integer-valued random variable X and the corresponding pgf $g(t) = E[t^X]$ defined on $t \in [0, 1]$. Then,

    (a) $g(t)$ is non-decreasing and convex. Moreover, if $P(X = 0) < 1$, then $g(t)$ is strictly increasing, and if $P(X \leq 1) < 1$, it is strictly convex.

    (b) Suppose $P(X \leq 1) < 1$. If $E[X] \leq 1$, the equation $t = g(t)$ has a unique solution $t \in [0, 1]$, namely $t = 1$. If $E[X] > 1$, it has two solutions in $[0, 1]$, $t = 1$ and $t = t_0 \in (0, 1)$.

    Proof: Observe that for $t \in (0, 1)$,

$$g'(t) = \sum_{i=1}^{\infty} i t^{i-1} P(X = i) \geq 0$$
$$g''(t) = \sum_{i=2}^{\infty} i(i-1) t^{i-2} P(X = i) \geq 0.$$

So, $g(t)$ is non-decreasing (by the first equation) and convex (by the second equation). Also, for $g'(t) = 0$ for some $t \in (0, 1)$, it is *necessary* to have $P(X = i) = 0$ for all $i \geq 1$ and therefore

$P(X = 0) = 1$. For $g''(t) = 0$ for some $t \in (0,1)$, one must have $P(X = i) = 0$ for all $i \geq 2$, which means $P(X = 0) + P(X = 1) = 1$.

Now consider the strictly increasing strictly convex case. Note that $g(0) = P(X = 0)$ and $g(1) = 1$. So $t = 1$ is always a solution. Also, if $P(X = 0) = 0$ then the solutions are at the end points only $t = 0$ and $t = 1$. Next we split up the cases into $g'(1) = E[X] \leq 1$ and $g'(1) = E[X] > 1$. **Draw a diagram of $g(t)$ and the (t,t) line.** We see that in the former $t = 1$ can be the only solution, while in the latter there will be an intermediate point $t_0 \in (0,1)$ that will also be a solution. **You can formalize this if you want.**

23. **Returning to Galton-Watson extinction probabilities.** Now let Z denote any of the random variables in Z_{nk} in the Galton-Watson process. Exclude the trivial cases $P(Z = 0) = 1$ or $P(Z \geq 2) = 0$. By applying the above characterization of $g(t)$ we see that if $E[Z] \leq 1$, then the only solution of $t = g(t)$ in $[0,1]$ is 1, and therefore, $P(\mathcal{E}) = g(P(\mathcal{E})) = 1$. *Thus the branching process eventually becomes extinct.* If $E[Z] > 1$ then there is a solution of $t_0 \in (0,1)$ of the fixed point system, besides the solution 1. It now follows from the *strict convexity* of $g(t)$ that any sequence $y_n = P(X_n = 0)$ satisfying $y_0 = 0$ and $y_{n+1} = g(y_n)$ converges to t_0 (**Convince yourself**). Therefore, the solution $t = 1$ is ruled out and we have $P(\mathcal{E}) \in (0,1)$. Hence, *there is a non-null probability of the population not becoming extinct.*

24. **HW** Consider the generating function:

$$g(z) = \frac{1}{4} + \frac{1}{4}z + \frac{1}{2}z^2,$$

for a branching process. Find the extinction probabilities.