A Summary of Time-Homogeneous Finite State Markov Sequences.

Main objects.

State Space $\mathcal{S} = \{a_1, \ldots, a_N\}$;

The Sequence $\mathbb{X} = (X_n,\ n \geq 0),\ X_n \in \mathcal{S}$;

The Tail sigma algebra of $\mathbb{X} : \mathcal{I} = \bigcap_{n\geq 0} \sigma(X_k,\ k \geq n+1)$;

(One Step) Transition Probability Matrix $P = (p_{ij},\ i,j = 1,\ldots,N):\ p_{ij} = \mathbb{P}(X_{n+1} = a_j | X_n = a_i)$;

m Step Transition Probability Matrix $P^{(m)} = (p_{ij}^{(m)},\ i,j = 1,\ldots,N):\ p_{ij}^{(m)} = \mathbb{P}(X_{n+m} = a_j | X_n = a_i);\ P^{(1)} = P$.

Basic results.

(1) Markov Property: $\mathbb{P}(X_n = a_i | X_{n-1}, \ldots, X_0) = \mathbb{P}(X_n = a_i | X_{n-1})$, $n \geq 1$.

(2) Transition probability matrices are stochastic: $p_{ij}^{(m)} \geq 0,\ \sum_{j=1}^N p_{ij}^{(m)} = 1,\ i = 1,\ldots,N,\ m = 1,2,\ldots$.

(3) Chapman-Kolmogorov equation: $P^{(m)} = P^m$.
Proof. Start with $m = 2$:

$$\begin{aligned} p_{ij}^{(2)} &= \mathbb{P}(X_2 = a_j | X_0 = a_j) = \sum_{\ell=1}^N \mathbb{P}(X_2 = a_j, X_1 = a_\ell | X_0 = a_i) \\ &= \sum_{\ell=1}^N \mathbb{P}(X_2 = a_j | X_1 = a_\ell, X_0 = a_i)\mathbb{P}(X_1 = a_\ell, X_0 = a_i) \\ &= \sum_{\ell=1}^N \mathbb{P}(X_2 = a_j | X_1 = a_\ell)\mathbb{P}(X_1 = a_\ell, X_0 = a_i) = \sum_{\ell=1}^N p_{i\ell}p_{\ell j}, \end{aligned}$$

where the third equality is a particular case of

$$\mathbb{P}(A \cap B | C) = \mathbb{P}(A | B \cap C)\mathbb{P}(B|C)$$

and the fourth equality is the Markov property. The general m is similar.

(4) Distribution of X_n: if $\mu_i = \mathbb{P}(X_0 = a_i)$, $i = 1,\ldots,N$, then

$$\mathbb{P}(X_n = a_j) = \sum_{i=1}^N \mu_i p_{ij}^{(n)}.$$

(5) Linear Model Representation: If $\mathbf{X}_n$ is a column vector in $\mathbb{R}^N$ with component number i equal to $I(X_n = a_i)$, then

$$\mathbf{X}_{n+1} = P^\top \mathbf{X}_n + \mathbf{V}_{n+1},$$

where $P^\top$ is the transpose of the matrix P and the random variables $\mathbf{V}_n$, $n \geq 1$, defined by

$$\mathbf{V}_n = \mathbf{X}_n - P^\top \mathbf{X}_{n-1},$$

have zero mean [because, by construction, $\mathbb{E}(\mathbf{X}_n | \mathbf{X}_{n-1}) = P^\top \mathbf{X}_{n-1}$], and are *uncorrelated*: $\mathbb{E}\mathbf{V}_n^\top \mathbf{V}_m$ is the zero matrix for $m \neq n$ [similarly, by conditioning].

(6) Recursive generation of a sample path of $\mathbb{X}$: on step n, generate a random variable U that is uniform on $(0,1)$ and, if $X_n = a_i$, then set $X_{n+1} = a_1$ if $U \leq p_{i1}$, $X_{n+1} = a_2$ if $p_{i1} < U \leq p_{i1} + p_{i2}$, etc.

(7) Exponential (Geometric) Ergodicity: If there exists an ℓ such that $p_{ij}^{(\ell)} > 0$ for all $i,j = 1,\ldots,N$, then there exists a unique probability distribution $\pi = (\pi_1, \ldots, \pi_N)$ on $\mathcal{S}$ and numbers $C > 0$ and $r \in (0,1)$ with the following properties:

- $\pi_i > 0$ for all $i = 1,\ldots,N$
- π_i is the almost-sure limit of $\frac{1}{n}\sum_{k=1}^n \mathbb{I}(X_k = a_i)$ as $n \to +\infty$;
- $1/\pi_i$ is the average return time to a_i if $X_0 = a_i$;
- $\pi_j = \sum_{i=1}^N \pi_i p_{ij}$ (invariance of π: if $\mathbb{P}(X_0 = a_i) = \pi_i$, then $\mathbb{P}(X_n = a_i) = \pi_i$ for all $n > 0$; also, as a *row* vector, π is the *left* eigenvector of P);
- The distribution of X_n converges to π exponentially quickly regardless of the initial distribution of X_0:

$$\max_{i,j} |\pi_j - p_{ij}^{(n)}| \leq Cr^n, \quad \max_j \Big| \sum_{i=1}^N \mu_i p_{ij}^{(n)} - \pi_j \Big| = \max_j \Big| \sum_{i=1}^N \mu_i (p_{ij}^{(n)} - \pi_j) \Big| \leq Cr^n. \tag{1}$$

An outline of the proof. Taking for granted that π exists, let $\delta > 0$ be such that $p_{ij}^{(\ell)} > \delta\pi_j$ for all i. Take $\theta = 1 - \delta$ and define Q by $P^\ell = (1-\theta)\Pi + \theta Q$, where the matrix Π has all rows equal to π. Next, argue by induction that

$$P^{n\ell} = (1-\theta^n)\Pi + \theta^n Q^n$$

so that, for $m \geq 1$,

$$P^{n\ell+m} - \Pi = \theta^n (Q^n P^m - \Pi),$$

from which (1) follows with $r = \theta^{1/\ell}$ and $C = 1/\theta$.

Further Definitions.

(1) Chain is an indication that the state space is countable; Sequence is an indication that the time is discrete.
(2) Irreducible Markov chain $\mathbb{X}$: for every i, j there is an $n \geq 1$ such that $p_{ij}^{(n)} > 0$ (each state is accessible from every other state);
(3) The Period d_i of the state a_i is the greatest common divisor of the numbers n for which $p_{ii}^n > 0$. Aperiodic state has $d_i = 1$.
(4) Total variation distance between two probability distributions μ and ν on $\mathcal{S}$ is s

$$\|\mu - \nu\|_{\mathrm{TV}} = \frac{1}{2}\sum_{i=1}^N |\mu_i - \nu_i|.$$

Further Results.

(1) A state a_i is aperiodic if and only if there exists an $n_0 \geq 1$ such that, for all $n > n_0$, $p_{ii}^{(n)} > 0$.
(2) All states in an irreducible chain have the same period.
(3) If the chain is irreducible and aperiodic, then there exists an m such that $p_{ij}^{(m)} > 0$ for all $i, j = 1, \ldots, N$. [Indeed, $p_{ij}^{(m+q+r)} \geq p_{ik}^{(m)} p_{kk}^{(q)} p_{kj}^{(r)}$.]
(4) A stronger version of (1) is

$$\|\pi - p_{i\bullet}^{(n)}\|_{\mathrm{TV}} \leq Cr^n.$$

The fundamental question: given the chain $\mathbb{X}$, how to find the r from (1).

Further developments: Mixing time and cut-off phenomena; Metropolis-Hastings algorithm and MCMC; Hidden Markov Models (HMM) and the estimation algorithms of Viterbi and Baum-Welch.

The reference: D. A. Levin and Y. Peres. Markov chains and mixing times, Second edition. American Mathematical Society, Providence, RI, 2017. xvi+447 pp. ISBN: 978-1-4704-2962-1.

The bottom line: A *nice* discrete time Markov chain is time-homogeneous, finite state, irreducible and aperiodic; for a time-homogenous finite-state Markov chain $\mathbb{X}$ with transition matrix P, the following three properties are equivalent: (a) all entries of the matrix P^n are non-zero for some $n \geq 1$; (b) as $n \to \infty$, every row of P^n converges to the same (unique) invariant distribution π for $\mathbb{X}$ and $\pi_i > 0$ for all i; (c) $\mathbb{X}$ is irreducible and aperiodic.

Beyond the nice setting.

Starting point: a discrete-time, time-homogenous Markov chain $\mathbb{X} = (X_0, X_1, X_2, \ldots)$ with a *countable* state space $\mathcal{S} = \{a_1, a_2, \ldots\}$ and transition matrix $P = (p_{ij},\ i,j \geq 1)$; $p_{ij} = \mathbb{P}\big(X_{n+1} = a_j | X_n = a_i\big)$, and, for $m \geq 2$, $P^m = (p_{ij}^{(m)},\ i,j \geq 1)$, $p_{ij}^{(m)} = \mathbb{P}\big(X_{n+m} = a_j | X_n = a_i\big)$, with conventions $p_{ii}^{(0)} = 1,\ p_{ij}^{(1)} = p_{ij}$.

Definitions based on the *arithmetic* properties of P.

(1) State a_i is absorbing if $p_{ii} = 1$;
(2) State a_j is accessible from a_i, $i \neq j$ if $p_{ij}^{(m)} > 0$ for some $m \geq 1$ [by convention, every state is accessible from itself];
(3) Communicating states a_i and a_j are mutually accessible: $p_{ij}^{(n_1)} > 0,\ p_{ji}^{(n_2)} > 0$ for some $n_1, n_2 \geq 0$ [by convention, each state communicates with itself, making communication an equivalence relation];
(4) State a_j is essential if it is absorbing or if it communicates with every state that is accessible from it: either $p_{jj} = 1$ or, for every $i \neq j$, if $p_{ji}^{(n)} > 0$ for some n then $p_{ij}^{(m)} > 0$ for some m;
(5) Inessential states are those that are not essential: $\mathbb{X}$ eventually gets out of an inessential state but never comes back to it;
(6) Irreducible class is an equivalence class of (essential) communicating states;
(7) The Period d_i of the state a_i is the greatest common divisor of the numbers n for which $p_{ii}^n > 0$;

(8) IRREDUCIBLE CHAIN has only one irreducible class;
(9) APERIODIC CHAIN has all states with period 1.
(10) An INVARIANT (STATIONARY) MEASURE for P (or for $\mathbb{X}$) is a positive measure μ on $\mathcal{S}$ such that $\sum_i \mu_i p_{ij} = \mu_j$, with $\mu_i = \mu(a_i)$. *Note that* $\mu(\mathcal{S}) = +\infty$ *is allowed.* If $\mu(\mathcal{S}) < +\infty$ and $\pi_i = \mu_i/\mu(\mathcal{S})$, then π is the INVARIANT (STATIONARY) DISTRIBUTION for P.
(11) An invariant measure μ for P is called REVERSIBLE if

$$\mu_i p_{ij} = \mu_j p_{ji} \quad \text{for all } i, j. \tag{2}$$

Note that a measure satisfying (2) *must be invariant: just sum up both sides over i (or over j).*
(12) A function $f : \mathcal{S} \to \mathbb{R}$ is called [SUB](SUPER)-HARMONIC if $f(a_i)\,[\le](\ge) = \sum_j p_{ij} f(a_j)$.

Definitions based on the *asymptotic* properties of P^n as $n \to \infty$.

(1) State a_i is RECURRENT if $\sum_{n\ge1} p_{ii}^{(n)} = +\infty$ and is TRANSIENT otherwise;
(2) State a_i is POSITIVE RECURRENT if

$$\sum_{n=1}^{\infty} n f_i^{(n)} < \infty, \quad f_i^{(1)} = p_{ii}, \quad p_{ii}^{(n)} = \sum_{k=1}^{n} f_i^{(k)} p_{ii}^{(n-k)}, \ n \ge 2.$$

(3) State a_i is NULL RECURRENT if it is recurrent but not positive recurrent;
(4) A chain $\mathbb{X}$ is (positive) recurrent if all the states are (positive) recurrent;
(5) A chain $\mathbb{X}$ is ERGODIC if, as $n \to \infty$, all rows of P^n converge to the (unique) invariant distribution π for P and $\pi_i > 0$ for all i.

Basic facts:

(1) If the state space has N elements, then $\mathbb{X}$ is irreducible if and only if the matrix $\mathbb{I}_{N\times N} + P + P^2 + \cdots + P^N$ has all entries strictly positive;
(2) All states within one irreducible class have the same period, leading to a *cycle, or cyclic, decomposition* of the class;
(3) If a_i is recurrent and a_j is accessible from a_i, then a_j is also recurrent and communicates with a_i;
(4) With $\tau_i = \inf\{n \ge 1 : X_n = a_i\}$, we have

$$f_i^{(k)} = \mathbb{P}(\tau_i = k | X_0 = a_i),$$

so that

$$\mathbb{P}(\tau_i < +\infty | X_0 = a_i) = \sum_{k=1}^{\infty} f_i^{(k)} < \infty, \ \sum_{n\ge1} p_{ii}^{(n)} = \mathbb{P}(\tau_i < +\infty | X_0 = a_i)\left(1 + \sum_{n\ge1} p_{ii}^{(n)}\right),$$

and the state a_i is
- transient if and only if $\mathbb{P}(\tau_i = +\infty | X_0 = a_i) > 0$;
- recurrent if and only $\mathbb{P}(\tau_i < +\infty | X_0 = a_i) = 1$,
- positive recurrent if and only if $\mathbb{E}(\tau_i | X_0 = a_i) < \infty$;

(5) Inessential states are transient, and, if the state space is finite, then all transient states are inessential (that is, in a finite state Markov sequence, a state is (positive) recurrent if and only if it is essential);
(6) If a_j is accessible from a_i, $i \ne j$, and a_j is absorbing, then a_i is inessential;
(7) In terminology of graph theory, a chain on a graph is
- irreducible if and only if the graph is connected;
- aperiodic and irreducible if and only if the graph is non-bipartite;

Further facts.

(1) If a_i is a recurrent state and $\tau_i = \inf\{n \ge 1 : X_n = a_i\}$, then

$$\mu_j^{(i)} = \sum_{n=0}^{\infty} \mathbb{P}(X_n = a_j, \tau_i > n | X_0 = a_i)$$

defines an invariant measure for P; the measure is finite if a_i is positive recurrent.
(2) If the chain is irreducible, then the following three statements are equivalent: (a) at least one state is positive recurrent; (b) all states are positive recurrent; (c) an invariant *distribution* exists.
(3) If π is an invariant *distribution* for P and $\pi(a_i) > 0$, then a_i is recurrent;
(4) If $\mathbb{X}$ is irreducible, aperiodic, and recurrent, then the tail sigma algebra of $\mathbb{X}$ is trivial;

(5) If $\mathbb{X}$ is irreducible and recurrent, and all states have period $d > 1$, then the tail sigma algebra of $\mathbb{X}$ is determined by the *cycle decomposition* of the state space $\mathcal{S}$ and is non-trivial.
(6) A function $f : \mathcal{S} \to \mathbb{R}$ is [sub](super)-harmonic if and only if the sequence $f(X_n), n \geq 0$ is a [sub](super)-martingale with respect to the filtration generated by $\mathbb{X}$. An irreducible $\mathbb{X}$ is recurrent if and only if every non-negative super-harmonic function is constant.

Examples

(1) Simple symmetric random walk on the line with arbitrary starting point is irreducible; each state is null recurrent and has period 2; the tail sigma algebra consists of four sets.
(2) Ehrenfest chain has $\mathcal{S} = \{0, 1, \ldots, N\}$,
$$p_{ij} = \begin{cases} j/N, & j = i-1, \\ (N-j)/N, & j = i+1, \\ 0, & \text{otherwise} \end{cases}$$
and, even though all states have period 2, there is a unique invariant distribution π with $\pi_k = 2^{-N}\binom{N}{k}$. Still, the chain is not ergodic in the sense that the rows of P^n do not converge to π as $n \to \infty$.
(3) A birth and death chain with $\mathcal{S} = \{0, 1, 2 \ldots\}$, $p_{00} = 2/3, p_{01} = 1/3$,
$$p_{ij} = \begin{cases} 2/3, & j = i-1, \\ 1/3, & j = i+1, \\ 0, & \text{otherwise,} \end{cases}$$
is reversible, with invariant distribution $\pi_k = 2^{-k-1},\ k = 0, 1, 2, \ldots.$
(4) $M/G/1$ queue with arrival intensity λ and service time cdf F: if $a_k = \frac{1}{k!}\int_0^{+\infty} e^{-\lambda t}(\lambda t)^k dF(t)$ [probability that k customers arrive while one is served], then the numbers of customers X_n in the buffer at the time n-th customer enters service is a Markov chain with
$$p_{ij} = \begin{cases} a_0 + a_1, & i = j = 0, \\ a_k, & j = i - 1 + k \text{ and } i \geq 1 \ \text{ or } k > 1, \\ 0, & \text{otherwise.} \end{cases}$$
With $\nu = \sum_{k\geq 1} k a_k$ [average number of new customers arriving while one is served], the chain is transient if $\mu > 1$, null recurrent if $\mu = 1$, and positive recurrent, with a unique invariant distribution, if $\mu < 1$.
(5) A particular example of an $M/M/\infty$ queue is $X_{n+1} = \sum_{m=1}^{X_n} \xi_{mn} + Y_n$, where ξ_{mn} are iid Bernoulli with probability of success $p \in (0,1)$ [representing the customers still served at time $n+1$] and Y_n are iid (also independent of ξ) Poisson random variables with mean λ [representing new customers arriving in one time unit], then the chain $(X_n), n \geq 1$, is ergodic, with the unique invariant distribution equal to Poisson with mean $\lambda/(1-p)$.
(6) Gambler's ruin model has $a_i = i,\ i = 0, \ldots, N$, with absorbing states a_0 and a_N and
$$p_{ij} = \begin{cases} p, & j = i+1,\ i < N \text{ [winning one unit]} \\ q = 1-p, & j = i-1,\ i > 0 \text{ [losing one unit]} \\ 0, & \text{otherwise.} \end{cases}$$
Here, all states $a_i,\ i = 1, \ldots, N-1$ are inessential, and there are infinitely many invariant distributions: any probability distribution of the form $\pi_0 = \alpha \in [0,1]$, $\pi_N = 1-\alpha$, $\pi_i = 0$ otherwise. Accordingly, for this problem, the object of interest is not the long-time behavior but the *probability of ruin* r_n, that is, the probability of reaching a_0 before reaching a_N if the starting state is a_n so that n represents gambler's starting capital. Because $r_n = pr_{n+1} + qr_{n-1}$ and $r_0 = 1, r_N = 0$, we have $r_n = 1-(n/N)$ for $p = 1/2$ and $r_n = (\beta^n - \beta^N)/(1-\beta^N),\ \beta = q/p \neq 1$. When the odds are not in gambler's favor [that is, $p < q$], the ruin is essentially certain: for example, with $N = 100$, $p = 0.455$, and $q = 0.545$ we have $\beta \approx 1.1978$ and $r_{80} \approx 0.97$.