PERSONAL - Markov Chains

Introduction

- **Markov chain**: a stochastic process describing a sequence of possible events in which the probability of each event depends **only on the state attained in the previous event**
  - **discrete-time Markov chain**: the chain moves state at *discrete* time steps (used in this course)
  - **defining property**: a *stochastic process* $(X_t)_{t\in\mathbb{N}_0}$ on a *countable* (finite or countably infinite) state space E is called a *Markov chain* if for every $n \in \mathbb{N}$ and for every $i, j, i_0, ..., i_{n-1} \in E$ such that $P(X_0 = i_0, ..., X_n = i) > 0$, it holds that $P(X_{n+1} = j \mid X_0 = i_0, ..., X_n = i) = P(X_{n+1} = j \mid X_n = i)$
    - **equivalent (homogeneous)**: $P(X_0 = i_0, ..., X_n = i_n) = P(X_0 = i_0, ..., X_{n-1} = i_{n-1})\Pi(i_{n-1}, i_n)$
    - **equivalent (hom., expanded)**: $P(X_0 = i_0, ..., X_n = i_n) = \mu(i_0)\Pi(i_0, i_1)...\Pi(i_{n-1}, i_n)$
    - $P(X_n = i_n) = \sum_{i_0,...,i_{n-1}\in E^n} P(X_n = i_n, ..., X_0 = i_0)$
  - **homogeneity**: the transition probabilities are *independent of the time* t
    - **formally**: the Markov chain $(X_t)_{t\in\mathbb{N}_0}$ is homogeneous if for every $i, j \in E$ and every $n, m \in \mathbb{N}$, if $P(X_{n-1} = i) > 0$ and $P(X_{m-1} = i) > 0$, then $P(X_n = j \mid X_{n-1} = i) = P(X_m = j \mid X_{m-1} = i)$
- **stochastic process**: a sequence $(X_t)_{t\in\mathbb{N}_0}$ of random variables (t: time), all defined on the same probability space $(\Omega, \mathcal{F}, P)$, all taking values in the same (finite or countably infinite) space E
  - **in other words**: a sequence $\{X(t) : t \in T\}$, with t meaning "(discrete) time" here
  - **notation**: $X_t = i$ means that at time t, the process is in the state i
- **conditional probability**: $P(A|B) = \frac{P(A,B)}{P(B)}$
- **independence**: two events A, B independent (of each other) if and only if $P(A, B) = P(A)P(B)$
  - **also**: $P(A|B) = P(A)$ and $P(B|A) = P(B)$ (observed by using above conditional probability definition)
- **stochastic matrix**: matrix where the sum of each row is 1
  - **formally**: $\Pi \in [0, 1]^{E\times E}$ stochastic, if $\sum_{j\in E} \Pi(i, j) = 1$ for all $i \in E$
  - **doubly stochastic matrix**: columns also sum to 1
    - formally: $\Pi \in [0, 1]^{E\times E}$ double stochastic, if Π stochastic and $\sum_{i\in E} \Pi(i, j) = 1$ for all $j \in E$
    - **note¹**: if a stochastic matrix is symmetric, it is also doubly stochastic
    - **note²**: for a double stochastic matrix, the *stationary distribution* is the *uniform distribution*
- **transition matrix**: a square matrix used to describe the transitions of a homogeneous Markov chain
  - **formally**: $\Pi \in [0, 1]^{E\times E}$ stochastic and $\Pi(i, j) = P(X_{n+1} = j \mid X_n = i)$ for all $i, j \in E$ and $n \in \mathbb{N}_0$ with $P(X_n = i) > 0$

Existence, Markov Property

- **initial distribution**: the distribution of X_0
  - **formally**: $\mu : E \to \mathbb{R}$, $\mu(i) := P(X_0 = i)$
    - $\forall i : \mu(i) \geq 0$
    - $\sum_{i\in E} \mu(i) = 1$
  - **existence theorem**: let μ be a *distribution* on E, let $\Pi \in [0, 1]^{E\times E}$ be a *stochastic matrix*; then there exists a *homogeneous Markov chain* $(X_t)_{t\in\mathbb{N}_0}$ with *initial distribution* μ and *transition matrix* Π
  - **lemma**: let $(X_t)_{t\in\mathbb{N}_0}$ be a *stochastic process* on E, let $\Pi \in [0, 1]^{E\times E}$ be a *stochastic matrix*; then $(X_t)_{t\in\mathbb{N}_0}$ is a homogeneous Markov chain with transition matrix Π *if and only if* for all $n \in \mathbb{N}$ and for all $i, j, i_0, ..., i_{n-1} \in E$ such that $P(X_0 = i_0, ..., X_n = i) > 0$, it holds that

$P(X_{n+1} = j \mid X_0 = i_0, ..., X_n = i) = \Pi(i, j)$

- **random-mapping representation**: every homogeneous Markov chain can be realized as $X_{n+1} = f(X_n, Z_{n+1})$
  - **formally**: let $Z_n, n \in N$ iid taking values in F and let E be a countable state space; let $f : E \times F \to E$ be a measurable function and let $X_0 : \Omega \to E$ be a random variable independent of Z_n; set $X_{n+1} = f(X_n, Z_{n+1})\ \forall n \in \mathbb{N}_0$, then $(X_n)_{n \in \mathbb{N}_0}$ is a *homogeneous Markov chain* on E with transition matrix $\Pi(i, j) = P(f(i, Z_1) = j)\ \forall i, j \in E$
    - **expanded**: $X_1 = f(X_0, Z_1);\ \ X_2 = f(X_1, Z_2) = f(f(X_0, Z_1), Z_2)$ and so on...
  - **on** $[0, 1]$: let E be a countable state space and let $\Pi \in [0, 1]^{E \times E}$ be a stochastic matrix; let $Z_n, n \in N$ iid uniform distributed on $[0, 1]$; let $f : E \times [0, 1] \to E$ be a measurable function; set $X_0 = i_0,\ X_{n+1} = f(X_n, Z_{n+1})\ \forall n \in \mathbb{N}_0$, then $(X_n)_{n \in \mathbb{N}_0}$ is a *homogeneous Markov chain* on E with transition matrix Π and $P(X_0 = i_0) = 1$
    - **corollary**: if E is countable and $\Pi \in [0, 1]^{E \times E}$ is a stochastic matrix, there exists a homogeneous Markov chain with transition matrix Π
- **Markov property**: no matter what happened before time m, once we know $X_m = k$, the process *restarts* at k with the same law as the *original chain* started from k, dropping all history before m
  - **in other words**: the future only depends on the present
  - **formally**: let $(X_n)_{n \in \mathbb{N}_0}$ be a Markov chain with transition matrix Π; fix $m \in \mathbb{N}$ and $k \in E$ such that $P(X_m = k) > 0$; then, under $\tilde{P} := P(\cdot \mid X_m = k)$, the sequence $(\tilde{X}_n := X_{n+m})_{n \in \mathbb{N}_0}$ is a Markov chain with transition matrix Π and starting distribution δ_k (Dirac measure), independent of $X_0, ..., X_m$ (?)
    - $\delta_k(i) = \begin{cases} 1 & \text{if } k = i \\ 0 & \text{if } k \neq i \end{cases}$
    - for any past event A before or at X_m and any future event B after X_m, $P(A \cap B \mid X_m = k) = P(A \mid X_m = k)P(B \mid X_m = k)$

Finite Dimensional Distributions

- **what defines a Markov chain?**: let E be a *countable state space*, let $\Pi \in [0, 1]^{E \times E}$ be a *stochastic matrix*, let μ be a distribution on E, let $(X_n)_{n \in \mathbb{N}_0}$ be a stochastic process on E; then the following are equivalent (iff-s):
  1. $(X_n)_{n \in \mathbb{N}_0}$ is a Markov chain with transition matrix Π and initial distribution μ
  2. $\forall n \in \mathbb{N}_0, i_0, ..., i_n \in E : P(X_0 = i_0, ..., X_i = i_n) = \mu(i_0)\Pi(i_0, i_1)...\Pi(i_{n-1}, i_n)$
  3. $\forall n \in \mathbb{N}_0, A_0, ..., A_n \subseteq E : P(X_0 \in A_0, ..., X_n \in A_n) = \sum_{i_0 \in A_0} \mu(i_0) \sum_{i_1 \in A_1} \Pi(i_0, i_1)... \sum_{i_n \in A_n} \Pi(i_{n-1}, i_n)$
  - **uniqueness in distribution**: μ and Π uniquely define the distribution of a Markov chain
    - **formally**: let μ be a distribution on E and $\Pi \in [0, 1]^{E \times E}$ be a stochastic matrix on E; then two Markov chains $(X_n)_{n \in \mathbb{N}_0}$ and $(Y_n)_{n \in \mathbb{N}_0}$, both with μ and Π, have the *same distribution*
- **Markov chain distribution at time** n: let $(X_n)_{n \in \mathbb{N}_0}$ be a Markov chain with initial distribution μ and transition matrix Π; then for all $n \in \mathbb{N}_0$, the distribution of X_n is $\mu^n = \mu\Pi^n$
  - **equivalent**: $\mu^n(i) = P(X_n = i) = (\mu\Pi^n)(i) = \sum_{j \in E} \mu(j)\Pi^n(j, i)$
    - μ^n row vectors (probability distribution of the chain at time n), Π^n n-th power of Π (n-step transition probability matrix)
      - $\mu^0 = \mu$
      - $\Pi^0 = I$
  - **corollary**: let $(X_n)_{n \in \mathbb{N}_0}$ be a Markov chain with transition matrix Π; then for all $m, n \in \mathbb{N}_0, i, j \in E$ such that $P(X_m = i) > 0$, it holds that $P(X_{m+n} = j \mid X_m = i) = \Pi^n(i, j)$ (see Markov property)
  - $P_i(X_n = j) = \Pi^n(i, j) = \sum_{i_1, ..., i_{n-1} \in E^{n-1}} \Pi(i, i_1)...\Pi(i_{n-1}, j)$

Communication and Period

- **reachability**: a state j is *reachable* from i if there exists $n \in \mathbb{N}_0$ such that $\Pi^n(i, j) > 0$
  - **write**: $i \to j$
  - **formally**: $i \to j \iff \exists n \in \mathbb{N}_0 : \Pi^n(i, j) > 0 \iff \sum_{n=0}^{\infty} \Pi^n(i, j) > 0$
    - **tip**: fix $i_1, ..., i_{n-1} \in E$, then $\Pi^n(i, j) \geq \Pi(i, i_1)...\Pi(i_{n-1}, j)$ always; if the RHS is > 0, then you've proven $\Pi^n(i, j) > 0$
  - $\Pi^0(i, i) = 1 \implies i \to i$ always holds
- **communication**: two states i and j *communicate* if $i \to j$ and $j \to i$
  - **write**: $i \leftrightarrow j$
  - **formally**: $i \leftrightarrow j \iff \exists n, m \in \mathbb{N}_0 : \Pi^n(i, j) > 0, \Pi^m(j, i) > 0$
  - $\leftrightarrow$ is an *equivalence relation*; the equivalence classes are called *communication classes* $E/\leftrightarrow$
- **irreducibility**: a Markov chain is called *irreducible* if it only has *one communication class*, otherwise it is *reducible*
  - **in other words**: a Markov chain is irreducible if *every state* communicates with *every other state*

    - **formally**: $\forall i, j \in E : i \to j$
        - $\forall i, j \in E \; \exists n \in \mathbb{N}, i = i_0, i_1, ..., i_{n-1}, i_n = j : \Pi(i_0, i_1)...\Pi(i_{n-1}, i_n) > 0$
    - **equivalent**: $\forall i, j \in E, i \neq j \; \exists n = n(i, j) \in \mathbb{N} : \Pi^n(i, j) > 0$
- **closed set**: a (non-empty?) set $C \subseteq E$ is *closed* if you cannot leave the set
    - **formally**: $C \subseteq E$ closed $\iff \forall i \in C : \sum_{j \in C} \Pi(i, j) = 1$
    - **equivalent**: $\forall i \in C, j \notin C : \Pi(i, j) = 0$
    - **notes**:
        - $E, \emptyset$ are closed
        - if A, B are closed, then $A \cup B, A \cap B$ are also closed
        - every communication class is closed
        - if the Markov chain is irreducible, then the only closed sets are E and $\emptyset$
        - the Markov chain is irreducible iff $\sum_{n=0}^{\infty} \Pi^n$ has no zero entries
- **periodicity**: a state i has period k if k is the *greatest common divisor* of the number of transitions by which i can be reached, starting from i
    - **formally**: let $i \in E$, define $T(i) := \{n \geq 1 : \Pi^n(i, i) > 0\}$, the period of i is defined as $d_i := \gcd(T(i))$
        - $d_i = 1$: the state i is aperiodic
            - all states aperiodic $\implies$ the Markov chain is aperiodic
                - **formally**: $\forall i \in E : \gcd(T(i)) = 1$
        - $d_i > 1$: the state i is periodic
            - at least one state periodic $\implies$ the Markov chain is periodic
                - **formally**: $\exists i \in E : \gcd(T(i)) > 1$
    - **conventionally**: $\gcd(\emptyset) = \infty$
    - **periodicity under Π^n**: $\frac{d(i)}{\gcd(d(i), n)}$
        - n multiple of $d(i) \implies$ period collapses to 1 (i becomes aperiodic under Π^n)
    - **lemma**: if $i \leftrightarrow j$, then $d_i = d_j$
        - if the Markov chain is irreducible, all states have the same period $\to$ "period of the Markov chain"
            - **formally**: $\forall i, j \in E : i \to j \implies \forall i, j \in E : d_i = d_j$
    - **theorem**: let $(X_n)_{n \in \mathbb{N}_0}$ be an *irreducible* Markov chain with period d; then for all $i, j \in E$, there exist $m = m(i, j) \in \mathbb{N}_0$ and $n_0 = n_0(i, j) \in \mathbb{N}_0$ such that for all $n \geq n_0 : \Pi^{m+nd}(i, j) > 0$
        - choose $m = 0$ if $i = j$
        - **special case (corollary)**: let $(X_n)_{n \in \mathbb{N}_0}$ be an *irreducible, aperiodic* Markov chain on a *finite* state space E; then there exists $n_0 \in \mathbb{N}_0$ such that for all $i, j \in E$ and all $n \geq n_0$: $\Pi^n(i, j) > 0$ (i.e. you can get from i to j in *every sufficiently large* number of steps)
        - **lemma**: let $A \subseteq \mathbb{N}$ such that $\gcd(A) = 1$ and if $a, b \in A$, then $a + b \in A$; then there exists $n_0 \in \mathbb{N}$ such that $n \in A$ for all $n \geq n_0$
- **partitioning**: let $(X_n)_{n \in \mathbb{N}_0}$ be an *irreducible* Markov chain with period d; then there exists exactly one partitioning $C_0, ..., C_{d-1}$ of E such that for all $k = \{0, ..., d-1\}$ and $i \in C_k$: $\sum_{j \in C_{k+1}} \Pi(i, j) = 1$, where $C_d = C_0$
    - **in other words**: choose a state i_0, group all states by "distance mod d" from i_0
    - it's possible to rearrange the states to get a block matrix: $\Pi = \begin{array}{c|cccccc} & C_0 & C_1 & C_2 & \dots & C_{d-2} & C_{d-1} \\ \hline C_0 & 0 & \Pi_0 & 0 & \dots & 0 & 0 \\ C_1 & 0 & 0 & \Pi_1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{d-2} & 0 & 0 & 0 & \dots & 0 & \Pi_{d-2} \\ C_{d-1} & \Pi_{d-1} & 0 & 0 & \dots & 0 & 0 \end{array}$
        - Π^n is also a block matrix, Π^{nd} is a block diagonal matrix (i.e. diagonals are square matrices)

Stationary Distributions

- **stationary distribution**: a probability distribution that, once reached, remains *unchanged* over time as the chain evolves
    - **formally**: a probability measure α is called *stationary* for a Markov chain with transition matrix Π if for all $i \in E$: $\alpha(i) = \sum_{j \in E} \alpha(j)\Pi(j, i)$
    - **matrix form**: $\alpha\Pi = \alpha$
    - **theorem**: if the initial distribution μ of a Markov chain $(X_n)_{n \in \mathbb{N}_0}$ is stationary. then for all $n \in \mathbb{N}_0$ and $A \subseteq E$: $P_\mu(X_n \in A) = \mu(A)$
        - **in other words**: $\mu\Pi^n = \mu^n = \mu$

- then, $P_\alpha(X_n = i) = \alpha(i)$ for all $N \in \mathbb{N}_0, i \in E$
- **finding stationary distributions**: solve $\alpha\Pi = \alpha, \sum_{i \in E} \alpha(i) = 1$ ($0/1/\infty$ solutions)
  - if a Markov chain has 2 stationary distributions, then it has *infinitely* many
    - **formally**: if α, β are stationary, then so is every $\mu \in \{\lambda\alpha + (1-\lambda)\beta : \lambda \in (0,1)\}$
  - if $|E| < \infty$ (or countably infinite, positive recurrent):
    - exactly *one closed* communicating class $\iff$ exactly *one* stationary distribution
    - *two* or *more closed* communicating classes $\iff$ *infinitely many* stationary distributions

- **reverse transitions (time-reversed chain)**: let $(X_n)_{n \in \mathbb{N}_0}$ be a Markov chain with transition matrix Π and stationary distribution α such that $\alpha(i) > 0$ for all $i \in E$; define $\Pi'(i,j) := \dfrac{\alpha(j)\Pi(j,i)}{\alpha(i)}$, then for all $i, j \in E, n \in \mathbb{N}_0$: $\Pi'(i,j) = P_\alpha(X_n = j \mid X_{n+1} = i)$ are the *backwards (reverse) transition probabilities*
- **reversibility**: let $(X_n)_{n \in \mathbb{N}_0}$ be a Markov chain with transition matrix Π; a distribution α on E is called *reversible* if $\alpha(i)\Pi(i,j) = \alpha(j)\Pi(j,i)$ for all $i, j \in E$
  - the Markov chain is called reversible if it has a reversible distribution
  - **theorem**: every *reversible* distribution is *stationary*
  - **note**: if $(X_n)_{n \in \mathbb{N}_0}$ reversible and $\alpha(i) > 0$ for all i, then $P_\alpha(X_n = j \mid X_{n+1} = i) = \Pi(i,j) = P_\alpha(X_{n+1} = j \mid X_n = i)$ (so if we start from α, the forwards and backwards transition probabilities are the same)
  - **Kolmogorov's criterion**: an *irreducible, positive recurrent, aperiodic* Markov chain with transition matrix Π is reversible iff $\pi_{i_1 i_2}\pi_{i_2 i_3}...\pi_{i_n i_1} = \pi_{i_1 i_n}\pi_{i_n i_{n-1}}...\pi_{i_2 i_1}$ for all $i_1, ..., i_n \in E$

Strong Markov Property

- **σ-algebra**: given a set X, a collection $\mathcal{A}$ of subsets $A \subseteq P(X)$ is called a σ-algebra if:
  1. **contains the universe**: $X \in \mathcal{A}$ (and, by 2., the empty set $\emptyset \in \mathcal{A}$)
  2. **closed under complementation**: if $A \in \mathcal{A}$, then $X \setminus A = A^c \in \mathcal{A}$
  3. **closed under countable unions**: if $A_n \in \mathcal{A}$ for all $n \in \mathbb{N}$, then $\bigcup_{n=1}^\infty A_n \in \mathcal{A}$
- **filtration**: a growing sequence of information where past information does not get lost over time (accumulates)
  - **formally**: let $(\Omega, \mathcal{F}, P)$ be a probability space; a sequence $\mathcal{F}_n \subseteq \mathcal{F}$ for $n \in \mathbb{N}_0$ is called a *filtration* if $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ for all $n \in \mathbb{N}_0$
  - **natural filtration**: *all the information* generated by the chain up to time n (i.e. the exact states of the chain, whether certain states are in subsets of the state space, complements, functions of the past etc., but not model parameters themselves)
    - **formally**: let $(X_n)_{n \in \mathbb{N}_0}$ be a Markov chain on E; define $\mathcal{F}_n := \sigma(X_0, ..., X_n)$ as the smallest σ-algebra containing all events of type $X_t^{-1}(A)$ for $A \subseteq E$ and $t \in \{0, ..., n\}$; then $(\mathcal{F})_{n \in \mathbb{N}_0}$ is the natural filtration of $(X_n)_{n \in \mathbb{N}_0}$
- **stopping time**: the event of stopping at time n only depends on what happened up to that time
  - **formally**: a random variable $\tau : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ is called a *stopping time* with respect to the natural filtration $(\mathcal{F})_{n \in \mathbb{N}_0}$ if for all $n \in \mathbb{N}_0$: $\{\tau = n\} \in \mathcal{F}_n$ (i.e. can you rewrite it in a way that depends on X_i only up to n?)
  - **stopped Markov chain**: let $(X_n)_{n \in \mathbb{N}_0}$ be a Markov chain, let τ be a stopping time w.r.t. the natural filtration $(\mathcal{F})_{n \in \mathbb{N}_0}$; define $a \wedge b = \min(a,b)$ for $a, b \in \mathbb{R}$; the stopped Markov chain is $(X_{n \wedge \tau})_{n \in \mathbb{N}_0}$ with $X_{n \wedge \tau} = \begin{cases} X_n & \text{if } n \leq \tau \\ X_\tau & \text{if } n \geq \tau \end{cases}$
- **strong Markov property**: generalization of the Markov property to random stopping times; the Markov property still holds even if you *restart* the process at a random stopping time τ
  - **formally**: let $(X_n)_{n \in \mathbb{N}_0}$ be a Markov chain with transition matrix Π, let τ be a stopping time w.r.t. the natural filtration $(\mathcal{F})_{n \in \mathbb{N}_0}$; fix $k \in E$ such that $P(\tau < \infty, X_\tau = k) > 0$; then, under $\tilde{P} := P(\cdot \mid X_\tau = k)$, the sequence $(\tilde{X}_n := X_{n+\tau})_{n \in \mathbb{N}_0}$ is a Markov chain with transition matrix Π and starting distribution δ_k (Dirac measure), independent of $(X_{n \wedge \tau})_{n \in \mathbb{N}_0}$
    - if $\tau = \infty$, choose $(\tilde{X}_n)$ arbitrarily

Recurrence & Transience

- **recurrence and transience**: a state $i \in E$ is called...
  - **...recurrent**, if $P_i(T_i < \infty) \overset{!}{=} 1$
    - **in other words**: starting from i and from wherever you can go, there is always a way (path) of returning to i (finite)
    - **positive recurrent**: recurrent and $E_i[T_i] < \infty$
    - **null recurrent**: recurrent and $E_i[T_i] = \infty$
    - **theorem**: a state $i \in E$ is *recurrent* if and only if $\sum_{n=0}^\infty \Pi^n(i,i) = \infty$
  - **...transient**, if $P_i(T_i < \infty) < 1$

- **in other words**: starting from i, there is at least one path such that, if you take it, you will never be able to return to i (finite)
- i transient, E finite $\implies \alpha(i) = 0$

- **reminders**:
  - $P_i(\cdot) = P(\cdot \mid X_0 = i)$
  - $E_i[\cdot] = E[\cdot \mid X_0 = i]$
  - $T_i := \inf\{n \geq 1 : X_n = i\}$ (first return time)
  - $P_i(T_i < \infty) = P(\text{ever return to } i \mid X_0 = i)$
    - $P_i(T_i < \infty) = \sum_{n=1}^{\infty} P_i(T_i = n)$
      - $P_i(T_i = n) = P_i(X_1 \neq i, ..., X_{n-1} \neq i, X_n = i) = \sum_{i_1, ..., i_{n-1} \neq i} P_i(X_1 = i_1) P_i(X_2 = i_2 \mid X_1 = i_1) ... P_i(X_n = i \mid X_{n-1} = i_{n-1})$
  - $E_i[T_i] = \text{expected number of steps to go back to } i \text{ from } i$
    - $E_i[T_i] = \sum_{n=0}^{\infty} P_i(T_i \geq n)$
      - $P_i(T_i \geq n) = \sum_{k=n}^{\infty} P_i(T_i = k)$
- if a Markov chain returns to state i with probability 1 (*recurrent state*), it visits i *infinitely* many times
  - **formally**: $E_i[N_i] = \infty \iff P_i(T_i < \infty) = 1 \iff P_i(N_i = \infty) = 1$
    - $N_i = \sum_{n=1}^{\infty} 1_{\{X_n = i\}}$ (number of visits to i starting at time 1)
- if a Markov chain returns to state i with probability < 1 (*transient state*), it visits i *finitely* many times
  - **formally**: $E_i[N_i] < \infty \iff P_i(T_i < \infty) < 1 \iff P_i(N_i < \infty) = 1$
- **communication class theorem**: if $i \leftrightarrow j$, then either both are *recurrent* or both are *transient*
  - **appendix**: an *irreducible* Markov chain is called *recurrent* if all states are *recurrent* and *transient* if all states are *transient*

- **(*) different notation**:
  - $f_{ij} := P_i(T_j < \infty)$
  - $f_{ii} := P_i(T_i < \infty) = \sum_{n=1}^{\infty} f_{ii}$
  - $f_{ii}^{(n)} = P_i(T_i = n)$
- **hitting time**: the first time the chain enters a set $A \subseteq E$
  - **formally**: $H^A := \inf\{n \geq 0 : X_n \in A\}$
  - **hitting probability vector**: $h^A = (h_i^A)_{i \in E}$
    - **for each starting state** i: $h_i^A = P_i(H_A < \infty) = P(\text{chain ever visits } A \mid X_0 = i)$
    - $\begin{cases} h_i^A = 1 & \text{if } i \in A \\ h_i^A = \sum_{j \in E} \Pi(i,j) h_j^A & \text{if } i \notin A \end{cases}$ (*smallest* non-negative solution!)
      - $h_i^A = 0$ if A is unreachable from i (absorbing state, state outside of closed set)
  - **mean hitting time vector**: $k^A = (k_i^A)_{i \in E}$
    - **for each starting state** i: $k_i^A = E_i[H_A] = \text{expected number of steps to hit } A \text{ from } i$
    - $\begin{cases} k_i^A = 0 & \text{if } i \in A \\ k_i^A = 1 + \sum_{j \notin A} \Pi(i,j) k_j^A & \text{if } i \notin A \end{cases}$ (*smallest* non-negative solution!)
      - $k_i^A = \infty$ if A is unreachable from i (i.e. $P_i(H^A < \infty) = 0$)
- **invariant distribution**: a function $\alpha : E \to \mathbb{R}$ is called an invariant distribution for a Markov chain $(X_n)_{n \in \mathbb{N}_0}$ with transition matrix Π if:
  1. $\forall i \in E : \alpha(i) \in [0, \infty)$
  2. $\exists i \in E : \alpha(i) > 0$ (i.e. α is not the null function)
  3. $\alpha(i) = \sum_{j \in E} \alpha(j) \Pi(j, i)$
     - **matrix form**: $\alpha \Pi = \alpha \; (= \alpha \Pi^n)$
  - if, additionally, $\sum_{i \in E} \alpha(i) = 1$, then this is the *stationary (invariant) distribution*
  - Markov chain *recurrent* $\implies$ it has an *invariant measure*
  - Markov chain *recurrent* + *irreducible* $\implies$ it has a ***unique*** *invariant measure* (up to multiplication by a constant)
  - **existence theorem + construction**: let $(X_n)_{n \in \mathbb{N}_0}$ be an *irreducible, recurrent* Markov chain with transition matrix Π and state space E; then, $(X_n)_{n \in \mathbb{N}_0}$ has an *invariant measure* which can be *constructed* as follows:
    1. pick an element $0 \in E$ (any element, call it "0")
    2. let $T_0 := \inf\{n \geq 1 : X_n = 0\}$ (first return time to 0)
    3. write $\alpha(i) = E_0[\sum_{n=1}^{\infty} 1_{\{X_n = i\}} 1_{\{n \leq T_0\}}] = E_0[\sum_{n=1}^{T_0} 1_{\{X_n = i\}}]$ (expected number of visits to i between two visits to 0)

        - $\alpha(i) = \sum_{n=1}^{\infty} P_0(X_n = i, n \leq T_0)$
      - then, α is invariant
        - $\alpha(0) = 1$
        - $\sum_{i \in E} \alpha(i) = E_0[T_0]$
  - *any invariant measure* of an *irreducible* Markov chain is *(strictly) positive everywhere*
    - **formally**: $(X_n)_{n \in \mathbb{N}_0}$ irreducible $\implies \forall i \in E : \alpha(i) > 0$
  - **uniqueness up to a constant**: let $(X_n)_{n \in \mathbb{N}_0}$ be an *irreducible, recurrent* Markov chain with transition matrix Π, let α, β be invariant measures for $(X_n)_{n \in \mathbb{N}_0}$; then $\exists C > 0 : \alpha = C\beta$ (*unique* invariant measure up to *multiplication by a constant*)
  - **theorem**: let $(X_n)_{n \in \mathbb{N}_0}$ be an *irreducible, recurrent* Markov chain; then for all $i, j \in E$: $P_i(T_j < \infty) = 1$
  - **theorem**: $\sum_{i \in E} \alpha(i) < \infty \iff (X_n)_{n \in \mathbb{N}_0}$ (irreducible and) positive recurrent
    - **corollary**: an *irreducible* Markov chain is *positive recurrent* if and only if it has a *stationary distribution*
      - then, this stationary distribution is *unique* with $\alpha(i) > 0$ for all $i \in E$
  - **theorem**: let $(X_n)_{n \in \mathbb{N}_0}$ be an *irreducible, positive recurrent* Markov chain; then $\alpha(i) = \frac{1}{E_i[T_i]}$ for all $i \in E$
  - **theorem**: every *irreducible* Markov chain with a *finite state space* is *positive recurrent*
    - **formally**: $(X_n)_{n \in \mathbb{N}_0}$ *irreducible*, $|E| < \infty \implies (X_n)_{n \in \mathbb{N}_0}$ is *positive recurrent*
- **reversed transition matrix**: let $(X_n)_{n \in \mathbb{N}_0}$ be a Markov chain with transition matrix Π, let α be an invariant measure for $(X_n)_{n \in \mathbb{N}_0}$; define the α-*reversed* transition matrix as $\Pi^\alpha(i, j) = \dfrac{\alpha(j)}{\alpha(i)} \Pi(j, i)$
  - Π^α is a *stochastic matrix*
  - if $(X_n)_{n \in \mathbb{N}_0}$ is recurrent, then so is a Markov chain with transition matrix Π^α
  - let $(X_n)_{n=0}^N$ be a Markov chain with initial distribution δ_0, conditioned on $X_N = 0$; then, $(Y_n := X_{N-n})_{n=0}^N$ is a Markov chain with transition matrix Π^α, initial distribution δ_0, conditioned on $Y_N = 0$
    - **in other words**: $(Y_n)_{n=0}^N$ is the time reversal of the Markov chain $(X_n)_{n=0}^N$, where we start at 0 and return to 0 at time N
    - **corollary**: assume (X_n) is recurrent, let (Z_n) have transition matrix Π^α, then (Z_n) is also recurrent, and if you run both with initial distribution δ_0 from time 0 to time T_0, then $(Z_n)_{n=0}^{T_0}$ has the same distribution as $(X_{T_0 - n})_{n=0}^{T_0}$
  - let $(X_n)_{n \in \mathbb{N}_0}$ be a Markov chain with initial distribution δ_0 and transition matrix Π; let $(Y_n)_{n \in \mathbb{N}_0}$ be a Markov chain with initial distribution δ_0 and transition matrix Π^α; if $P(T_0 < \infty) = 1$, then $(Y_0, ..., Y_{T_0})$ has the *same distribution* as $(X_{T_0}, ..., X_0)$

Convergence

- **total variation metric**: notion of convergence; the "distance" between α and β in total variation
  - **formally**: let α, β be distributions on E (countable), then the total variation distance is defined as $d_{TV} = \dfrac{1}{2} \sum\limits_{i \in E} |\alpha(i) - \beta(i)|$ (i.e. half of the L1 norm)
- **convergence theorem**: let $(X_n)_{n \in \mathbb{N}_0}$ be an *irreducible, aperiodic, positive recurrent* Markov chain with *invariant distribution* α; then $\lim\limits_{n \to \infty} P_i(X_n = j) = \lim\limits_{n \to \infty} \Pi^n(i, j) = \alpha(j)$
  - **recipe**: if the Markov chain is *not irreducible*, split Markov chain into equivalence classes and look at "restricted" transition matrices