PROBABILITY THEORY

MATH 154

Unit 22: Markov Chains

22.1. A discrete time **Markov process** is a stochastic process X_i such that the outcome of X_{n+1} given the past only depends on X_n for every n. This can be rephrased also in Martingale language. For now we look at an important simple case, where the random variables X_i take only finitely many values S, the set of **states**. A **Markov chain** is a stochastic process with values in S such that the conditional probability $\mathrm{P}[X_{n+1} = x_{n+1}|X_1 = x_1, \dots, X_n = x_n]$ is $\mathrm{P}[X_{n+1} = x_{n+1}|X_n = x_n]$. If the probabilities $P[X_{n+1} = a|P[X_n = b]$ are independent of n, we talk about a **time homogeneous Markov** chain.

22.2. It follows from the definition of a Markov process that X_n satisfies the **elementary Markov property**: for $n > k$,

$$\mathrm{P}[X_n \in B \mid X_1, \dots, X_k] = \mathrm{P}[X_n \in B \mid X_k] \ .$$

This means that the probability distribution of X_n is determined by knowing the probability distribution of X_{n-1}. The future depends only on the present and not on the past. In the time homogeneous case, the stochastic process defines a transformation T on a probability space $(S^{\mathbb{N}}, \mathcal{A} = \mathcal{B}^{\mathbb{N}})$, where $\mathcal{B}$ is the set of all subsets of S. As we will see, there are often measures π on S such that $P = \pi^{\mathbb{N}}$ is invariant. We want to understand such equilibria.

22.3. We now look at a homogeneous Markov chain on a finite state space S with s elements. Probability measures on S are vectors p with entries $p_i \geq 0$ such that $\sum_i p_i = 1$. The Markov chain is now determined by the left stochastic $s \times s$ matrix $M_{ij} = \mathrm{P}[X_{n+1} = j|X_n = x_i]$. [1] The matrix M^T has the eigenvalue 1 with eigenvector $[1, \dots, 1]$. Therefore, M has also an eigenvalues 1. Its eigenvector is a **stationary measure** describing a stable probability distribution. As Oskar Perron in 1907 and Georg Frobenius in 1908 have shown there is one if M has positive entries:

Theorem 1 (Perron-Frobenius)**.** *If all entries of a left stochastic $n \times n$ matrix A are positive, there is a unique eigenvector to the eigenvalue* 1.

Proof. The set $X = \{\sum_i x_i^2 = 1, x_1 \geq 0, \dots, x_n \geq 0\}$ is closed and bounded. If the entries of A are non-negative, the map $T(v) = Av/|Av|$ maps X to itself. The Brouwer fixed point theorem gives then already fixed point and so an eigenvector to the eigenvalue 1. If A has positive entries then TX is even contained in the interior

[1] Sometimes, right stochastic matrices are used. Matrix multiplication is applied to the right to row vectors.

of X. This fixed point is unique because the map is a **contraction**: There exists $0 < \lambda(x) < 1$ such that $d(Tx, Ty) \leq \lambda d(x, y)$, where d is the geodesic sphere distance. If there was a contraction with a uniform λ we could have used the Banach fixed point theorem. We do not need it: assume $T(x) = x$ and $T(y) = y$ are both fixed points, then the contraction property gives $d(x, y) = d(Tx, Ty) \leq \lambda(x)d(x, y) < d(x, y)$ a contradiction. We have now a unique fixed point $Av = \lambda v$ provided v has non-negative entries. Assume $Aw = w$ and w has some negative entry and $||w|| = 1$. Write $|w|$ for the vector with coordinates $|w_j|$. The computation

$$|w|_i = |w_i| = |\sum_j A_{ij} w_j| \leq \sum_j |A_{ij}||w_j| = \sum_j A_{ij}|w_j| = (A|w|)_i$$

shows that $(A|w|)$ is a vector with norm smaller or equal than 1. For any i with $w_i < 0$ we have an inequality so that $||Aw|| < 1$ contradicting $||w|| = 1$. The only eigenvectors to the eigenvalue 1 must be in X where we had a unique one. $\square$

22.4. T is a true contraction with respect to the **Hilbert metric** on X. One can then use directly the **Banach fixed point theorem**. There is also a connection to ergodic theory. Given an initial measure μ_1 on S, the map M defines measures μ_k on S.

Theorem 2. *A Markov chain defines a measure preserving map T on the product probability space $(\Omega, \mathcal{A}) = (S^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, \prod_k \mu_k)$.*

Proof. The product space $(\Omega, \mathcal{A}) = (S^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}})$ has the π-system $\mathcal{C}$ consisting of cylinder-sets $\prod_{n \in \mathbb{N}} B_n$ given by a sequence $B_n \in \mathcal{B}$ such that $B_n = S$ except for finitely many n. The $P = P_\mu$ on $(\Omega, \mathcal{C})$ is the product measure. This measure has a unique extension to the σ-algebra $\mathcal{A}$. The shift map T on Ω is measure preserving. $\square$

22.5. If M has positive entries and μ is the stable distribution, then T is the shift on $(S^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, \mu^{\mathbb{N}})$ and the random variables $X_i(x) = x_i$ are independent.

22.6. For a countable state space S one is in a random walk situation. The transition matrix M_{ij} then now a bounded linear operator on $l^2(S)$. We have seen in the last lecture that if $S = \mathbb{Z}^d$ and M is a scaled version of the adjacency matrix one could use Fourier theory to understand recurrence. In the infinite case like $S = \mathbb{Z}$, there is no equilibrium measure. The probability distribution of the walker diffuses like a solution of the heat equation. We can still look at $M^n \mu$, where μ is an initial probability measure and study its dynamics. On a translational invariant lattice the walk is also a sum of IID random variables $S_n = X_1 + X_2 + \cdots X_n$, where X_i take finitely many values. Since by the central limit theorem, the variance of S_n grows linearly in time n, the standard deviation grows like $\sqrt{n}$.

22.7. Given a finite stochastic matrix M and a point $x \in S$, the measures $P(x, \cdot)$ are the probability vectors, which are the columns of M. It is also denoted **Markov field**. We have $P^n(x, B) = \sum_{y \in B} P^n(x, y)$. We can see the transition probability functions also as elements in $\mathcal{L}(S, M_1(S))$, by thinking about each column as a probability measure in the set $M_1(S)$ of Borel probability measures on S. This point of view is often taken in economics.