PROBABILITY THEORY

MATH 154

Unit 18: Mixing

18.1. T is **ergodic** if all events $A \in \mathcal{A}$ fixed by T have probability 0 or 1. This can be rephrased as **Césaro decay of correlations** of all random variables $X = 1_{T^{-k}(A)}, Y_k = 1_B$ because $\mathrm{E}[X_k] = \mathrm{P}[T^{-k}(A)], \mathrm{E}[Y] = \mathrm{P}[B]$ and $\mathrm{Cov}[X_k, Y] = \mathrm{P}[T^{-k}(A) \cap B]$ and:

Theorem 1. *T ergodic $\Leftrightarrow \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} (\mathrm{P}[T^{-k}(A) \cap B] - \mathrm{P}[A]\mathrm{P}[B]) = 0 \; \forall A, B \in \mathcal{A}$.*

Proof. (i) The ergodic theorem gives for $X \in \mathcal{L}^2$ the limit $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} X(T^{-k}(x)) \to \mathrm{E}[X]$. Given $Y \in \mathcal{L}^2$, we have $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} X(T^{-k}(x))Y(x) \to \mathrm{E}[X]Y(x)$. Now take expectations $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} E[X(T^{-k})Y] \to E[X]\mathrm{E}[Y]$. If we take $X = 1_A, Y = 1_B$ we get the statement.
(ii) Use contraposition. Assume T is not ergodic. Take any B with $T(B) = B$ with $0 < \mathrm{P}[B] < 1$ and chose $A = B$. We want to show that B is trivial. The assumed identity gives $\mathrm{P}[A] = \frac{1}{n} \sum_{k=0}^{n-1} \mathrm{P}[A] = \frac{1}{n} \sum_{k=0}^{n-1} \mathrm{P}[T^{-k}(A) \cap B] \to \mathrm{P}[A]\mathrm{P}[B] = \mathrm{P}[A]^2$. But this implies either $\mathrm{P}[A] = 1$ or $\mathrm{P}[A] = 0$. $\square$

18.2. T is called **weakly mixing** if $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} |\mathrm{P}[A \cap T^{-k}(B)] - \mathrm{P}[A]\mathrm{P}[B]| = 0$. This looks similar than the ergodic property but there are absolute values! We have **absolute Césaro convergence**. A Kronecker system for example is ergodic but not weakly mixing. Also, no permutation on a finite probability space is weakly mixing but it is ergodic if there is a single cycle. The previous theorem shows that if T is weakly mixing, then T is ergodic. Putting the absolute values outside makes it smaller in general by the triangle inequality.

18.3. Define $T \times T(x, y) = (T(x), T(y))$ on the product probability space. If T is not ergodic, then $T^{-1}(A) = A$ then $T^{-1}(A \times \Omega) = (T^{-1}(A) \times \Omega) = (A, \Omega)$ so that $T \times T$ is not ergodic. We see that ergodicity of $T \times T$ is stronger than ergodicity. What does it mean?

Theorem 2. *T is weakly mixing if and only if $T \times T$ is ergodic.*

Proof. We will show this in class. It needs a bit of real analysis (see HW 8). A second proof is spectral theoretic using that U_T has no eigenvalues besides 1. There is not enough space here on two pages. We will show the even stronger statement: T is weakly mixing if and only if $T \times T$ is weakly mixing. $\square$

18.4. Note that if T is ergodic, then the power $T^2(x) = T(T(x))$ is not necessarily ergodic. A simple example is the measure preserving transformation $T(x) = x + 1 \mod 6$ on the finite probability space $\Omega = \mathbb{Z}_6 = \{0, 1, \dots, 5\}$ with $\mathcal{A} = 2^\Omega$ and uniform probability measure $\mathrm{P}[A] = |A|/|\Omega|$. The transformation T is ergodic, but T^2 leaves the set $A = \{0, 2, 4\}$ invariant and $\mathrm{P}[A] = 1/2$ is not in $\{0, 1\}$.

18.5. A measure preserving is called **mixing** if $\mathrm{P}[T^{-n}(A) \cap B] \to \mathrm{P}[A]\mathrm{P}[B]$. For mixing transformations, the events $T^n(A)$ and B become more and more independent. In particular, $T^n(A)$ and A become more and more independent like a colored A of a dough Ω gets mixed by kneading and folding. In probability theory, mixing means that the random variables $X_n = 1_{T^{-n}(A)}$ and $Y = 1_B$ become more and more decorrelated. In particular $\mathrm{Cov}[X_n, X_0] \to 0$. Obviously, if T is mixing, then T is weakly mixing.

18.6. The linear operator $U_T(X) = X(T^{-1})$ on $\mathcal{L}^2$ is called the **Koopman operator** associated with T. The Hilbert space $\mathcal{L}^2$ has the inner product $\langle X, Y\rangle = \mathrm{E}[\overline{X}Y]$. The unitary property $\langle UX, UY\rangle = \langle X, Y\rangle$ follows from the fact that T preserves probabilities. Note that U always has the trivial eigenvalue 1 with constant eigenfunction $X = c$. This eigenvalue is not interesting. The **spectrum** of the dynamical system is defined as the spectrum of U on the orthogonal complement of the constant random variables (which are functions of zero expectation). Every random variable X defines a **spectral measure** $\mu = \mu_X$ on the complex unit circle defined by $\hat{\mu}_n = \langle X, U^n X\rangle - E[X]^2 = \mathrm{E}[XX(T^{-n})] - E[X]E[X(T^{-n})] = \mathrm{Cov}[X, X(T^{-n})]$.

Theorem 3. *T is weakly mixing if and only if U_T has continuous spectrum.*

18.7. Weakly mixing implies $(1/n)\sum_{k=0}^{n-1} \mathrm{P}[A(T^{-k}) \cap A] - \mathrm{P}[A]^2 \to 0$ implying that the Fourier transform of the spectral measure 1_A goes to zero for every 1_A. The Wiener theorem gives the reverse: if $\hat{\mu}_k \to 0$ in a Cesaro sense, then μ has no point spectrum.

Theorem 4 (Wiener Theorem). *If μ is a measure on the circle $\mathbb{T}$ with Fourier coefficients $\hat{\mu}_k$, then for every $x \in \mathbb{T}$, one has $\mu(\{x\}) = \lim_{n\to\infty} \frac{1}{2n+1}\sum_{k=-n}^{n} \hat{\mu}_k e^{ikx}$.*

Proof. The **Dirichlet kernel** $D_n(t) = \sum_{k=-n}^{n} e^{ikt} = \frac{\sin((k+1/2)t)}{\sin(t/2)}$ satisfies $D_n \star f(x) = S_n(f)(x) = \sum_{k=-n}^{n} \hat{f}(k)e^{ikx}$. The functions $f_n(t) := \frac{1}{2n+1}D_n(t-x) = \frac{1}{2n+1}\sum_{k=-n}^{n} e^{-ikx}e^{ikt}$ are bounded by 1 and go to zero uniformly outside any neighborhood of $t = x$. From $\lim_{\epsilon\to 0}\int_{x-\epsilon}^{x+\epsilon} |d(\mu - \mu(\{x\})\delta_x)| = 0$ follows $\lim_{n\to\infty}\langle f_n, \mu - \mu(\{x\})\rangle = 0$. But we also have $\langle f_n, \mu - \mu(\{x\})\rangle = \langle f_n, \mu\rangle - \langle f_n, \mu(\{x\})\rangle = \frac{1}{2n+1}\sum_{k=-n}^{n} \hat{\mu}_k e^{ikx} - \mu(\{x\})$. $\square$

18.8. If U_T has absolutely continuous spectrum, then by the Riemann-Lebesgue lemma, $\hat{\mu}_n \to 0$, so that T is mixing.

18.9. A measure preserving transformation is called **Bernoulli** if it is isomorphic to the shift of a product probability space $\prod_n(\Omega_n, \mathcal{A}_n, \mathrm{P}_n)$ where each $(\Omega_n, \mathcal{A}_n, \mathrm{P}_n)$ is the same finite probability space with $(\Omega = \{1, \dots, m\}, \mathcal{A} = 2^\Omega, \mathrm{P}[\{j\}] = p_j)$.

Theorem 5. *A Bernoulli transformation has absolutely continuous spectrum and so is mixing.*

18.10. We have the following **"chaos levels"**

$$\{\text{Bernoulli}\} \subset \{\text{Mixing}\} \subset \{\text{Weakly mixing}\} \subset \{\text{Ergodic}\}\ .$$