since ergodicity implies

$$\frac{1}{N}\sum_{n=0}^{N-1}\mu\otimes\mu\left((T\times T)^{-n}(A\times A)\cap B\times B\right)\xrightarrow{N\to\infty}\mu\otimes\mu(A\times A)\mu\otimes\mu(B\times B)=\mu(A)^2\mu(B)^2$$

and

$$\mu(A)\mu(B)\cdot\frac{1}{N}\sum_{n=0}^{N-1}\mu(T^{-n}A\cap B)\xrightarrow{N\to\infty}\mu(A)^2\mu(B)^2.$$

End of lecture 21

(ii) $\Rightarrow$ (v): Suppose $f\in L^2(\mu)$ is an eigenfunction of T with eigenvalue $\lambda\in\mathbb{C}$. Since U_T is an isometry, we have

$$|\lambda|\,\|f\|_2=\|\lambda f\|_2=\|U_T f\|_2=\|f\|_2,$$

i.e., $\lambda\in\mathbb{S}^1$. Define $\varphi\in L^2(\mu\otimes\mu)$ by

$$\varphi(x_1,x_2)=f(x_1)\overline{f(x_2)}.$$

Then

$$U_{T\times T}\varphi(x_1,x_2)=U_Tf(x_1)\overline{U_Tf(x_2)}=\varphi(x_1,x_2),$$

i.e., φ is an eigenfunction for eigenvalue 1/ invariant. Since $T\times T$ is ergodic by assumption, φ is constant and thus so is f (by Fubini).

(v) $\Rightarrow$ (ii) is more difficult (i.e., deferred)

Excursion to compact operators

Definition

A Banach space is a complete normed vector space over $\mathbb{C}$.

Definition (Compact operator)

Let V,W Banach spaces and $A:V\longrightarrow W$ linear. A is a compact operator if $\overline{A(B_1^V(0))}\subseteq W$ is compact.

Example Let $V=C^1([0,1])$ with $\|f\|:=\|f\|_\infty+\|f'\|_\infty$. Let $W=C^0([0,1])$ with norm $|||f|||:=\|f\|_\infty$. Then the "identity" map

$$i:(C^1([0,1]),\|\cdot\|)\hookrightarrow(C^0([0,1]),|||\cdot|||)$$

is compact by Arzela-Ascoli. To this end note that

$$i(B_1^V(0))\subseteq\{f\in W:|||f|||\le 1,\ f\text{ is }1\text{-Lipschitz}\}.$$

Definition

Let V a Banach space. V is a Hilbert space if the norm is induced by an inner product.

Definition

Let V a Hilbert space. A linear operator $A:V\to V$ is self-adjoint if $\forall v_1,v_2\in V\ \langle v_1,Av_2\rangle=\langle Av_1,v_2\rangle$.

Theorem (without proof – spectral theorem for self-adj compact operators)

Let V be a separable Hilbert space and let $A:V\to V$ self-adjoint compact. Then there exists an orthonormal basis consisting of eigenvectors of A, i.e., A is (in particular) completely diagonalizable. The eigenvalues are real and for every non-zero eigenvalue λ of A, the eigenspace for λ is finite dimensional. Moreover, the eigenvalues are bounded and if $|\lambda_1|\ge|\lambda_2|\ge\cdots$, then $\lambda_n\xrightarrow{n\to\infty}0$.

Note: $\sigma:\ell^2(\mathbb{Z})\longrightarrow\ell^2(\mathbb{Z})$ has no eigenvectors.

$(x_n)_{n\in\mathbb{N}}\longmapsto(x_{n+1})_{n\in\mathbb{N}}$

Proposition (Hilbert-Schmidt)

Let $(X,\mathcal{B},\mu)$ and $(Y,\mathcal{C},\sigma)$ σ-finite. Let $k\in L^2(\mu\otimes\sigma)$. Then the Hilbert-Schmidt integral operator $K:L^2(\mu)\longrightarrow L^2(\sigma)$ defined by

$$K(f)(y)=\int_X f(x)k(x,y)\,d\mu(x)\qquad\sigma\text{-a.s.}$$

is compact. It is self-adjoint if and only if $k(x_1,x_2)=\overline{k(x_2,x_1)}$ $\mu\otimes\mu$-a.s.

Proof that (v) $\Rightarrow$ (ii):

Suppose $T\times T$ is not ergodic. Then there exists $B\subseteq X\times X$ s.t. $\mu\otimes\mu(B)\in(0,1)$ and $(T\times T)^{-1}B=B$. Define

$$k_1(x,y)=\frac{1}{2}\left(\mathbb{1}_B(x,y)+\mathbb{1}_B(y,x)\right)$$

$$k_2(x,y)=\frac{1}{2i}\left(\mathbb{1}_B(x,y)-\mathbb{1}_B(y,x)\right).$$

Then $k_1(x,y)=\overline{k_1(y,x)}$ and $k_2(x,y)=\overline{k_2(y,x)}$. Moreover

$$\mathbb{1}_B=k_1+ik_2$$

implies that at least one of the k_i is non-constant (since $\mathbb{1}_B$ is non-constant).

Let k be a non-constant element in $\{k_i-\int k_i\,d\mu\otimes\mu : i=1,2\}$

Note that $U_{T\times T}k=k$, $k\neq 0$, $\int k=0$. Let $A:L^2(\mu)\longrightarrow L^2(\mu)$,

$$f\longmapsto\int_X k(x,y)f(x)\,d\mu(x).$$

Then

$$A(U_Tf)(y)=\int_X k(x,y)f(Tx)\,d\mu(x)=\int_X k(Tx,Ty)f(Tx)\,d\mu(x)$$

$$=\int_X k(x,Ty)f(x)\,d\mu(x)=U_T(Af)(y),$$

i.e., A and U_T commute.

Applying the spectral theorem and Hilbert-Schmidt, A is diagonalizable. If 0 is the only eigenvalue of A, then $k=0$ $\mu\otimes\mu$-a.s., which is absurd. So let $\lambda\neq 0$ be an eigenvalue for A. Then V_λ is finite-dimensional and $U_T V_\lambda \subseteq V_\lambda$. In particular, U_T has an eigenvector for eigenvalue λ, say, f. Since $f\neq 0$, and $\int k \, d\mu\otimes\mu = 0$, f is non-constant, which contradicts (v).

<u>Theorem</u>

Suppose that $(S^{\mathbb{N}_0}, \mathcal{B}(S^{\mathbb{N}_0}), \mu, \sigma)$ is a fully supported stationary (p, P)-Markov chain. The following are equivalent.

(i) σ is weak-mixing.

(ii) σ is mixing.

(iii) P is aperiodic, i.e., $\exists n_0 \in \mathbb{N}$ such that $P^{n_0} > 0$.

(iv) $\forall s,t\in S \quad (P^n)_{st} \longrightarrow p(t)$ as $n\longrightarrow\infty$.

The proof relies on the following observation.

<u>Proposition</u> ("part" of Perron-Frobenius)

Let P be a stochastic matrix. Suppose that $P>0$. Then 1 is the only eigenvalue on the unit circle, i.e.,

$$\forall \lambda \in \operatorname{spec}(P) \qquad \lambda\neq 1 \Rightarrow |\lambda|<1.$$

<u>Proof</u>: Let $\lambda \in \operatorname{spec}(P)-\{1\}$, $Pv=\lambda v$. Then v isn't a multiple of a positive vector by uniqueness of the positive eigenvector; cf. Exercise Sheet 5.

Assume w.l.o.g. $\|v\|_\infty = 1$. Then

$$|\lambda| = \|Pv\|_\infty = \max\left\{ \left|\sum_{j=1}^d P_{ij} v_j\right| : 1\le i\le d\right\}.$$

<u>Lemma</u>

Let π be a fully supported probability vector, $v\in\mathbb{C}^d$. Then

$$|\pi\cdot v| = \|v\|_\infty \quad\Rightarrow\quad v \text{ is constant}.$$

Indeed, $|\pi\cdot v| \le \sum_{j=1}^d \pi_j |v_j| \le \|v\|_\infty$ with equality if and only if $\forall 1\le j\le d \;\; |v_j| = \|v\|_\infty$.

$\therefore \sum_{j=1}^d \pi_j |v_j| = \|v\|_\infty \sum_{j=1}^d \pi_j z_j$ for some $(z_1,\dots,z_d)\in(\mathbb{S}^1)^d$.

and, clearly, averaging points on the circle has the desired property.

It follows that

$$|\lambda| = \max\left\{ \left|\sum_{j=1}^d P_{ij} v_j\right| : 1\le i\le d\right\} < \|v\|_\infty = 1.$$

$\square$

<u>Proof of the theorem</u>

(ii) $\Rightarrow$ (i): Clear

(i) $\Rightarrow$ (iii): By weak-mixing, for all $s,t\in S$

$$\frac{1}{N}\sum_{n=0}^{N-1} \Big|\underbrace{\mu(\mathcal{C}(s)\cap T^{-n}\mathcal{C}(t))}_{p(s)(P^n)_{st}} - \underbrace{\mu(\mathcal{C}(s))\mu(\mathcal{C}(t))}_{p(s)p(t)}\Big| \xrightarrow{N\to\infty} 0$$

$$\Rightarrow \frac{1}{N}\sum_{n=0}^{N-1} |P^{(n)}_{st} - p(t)| \xrightarrow{N\to\infty} 0$$

$\Rightarrow (P^n)_{st} \xrightarrow{n\to\infty} p(t)$ along a set $K_{st} = \mathbb{N} - J_{st}$, where J_{st} has density 0.

Since $\bigcup_{s,t\in S} J_{st}$ has density zero, there is $n_0\in\mathbb{N}$ s.t.

$$\forall s,t\in S \qquad |P^{(n_0)}_{st} - p(t)| < \frac{\min\{p(\sigma):\sigma\in S\}}{2},$$

i.e., $P^{n_0} > 0$.

(iii) $\Rightarrow$ (iv): Using the proposition, we know that

$$P^{n_0} \sim \begin{pmatrix} 1 & & & \\ & J_2(\lambda_2) & & \\ & & \ddots & \\ & & & J_k(\lambda_k)\end{pmatrix},$$

where $\operatorname{spec}(P^{n_0}) = \{1,\lambda_2,\dots,\lambda_k\}$ and $\forall 2\le i\le k \;\; |\lambda_i|<1$. Hence, since 1 is an eigenvalue of P, we have

$$P \sim \begin{pmatrix} 1 & & & \\ & J_2(\mu_2) & & \\ & & \ddots & \\ & & & J_\ell(\mu_\ell)\end{pmatrix} =: J \qquad (\mu\in\operatorname{spec}(P) \Rightarrow \mu^{n_0}\in\operatorname{spec}(P^{n_0}))$$

where $\operatorname{spec}(P) = \{1,\mu_2,\dots,\mu_\ell\}$ and $\forall 2\le i\le \ell \;\; |\mu_i|<1$. The right-hand side satisfies $J^n \longrightarrow \begin{pmatrix} 1 & 0 \\ 0 & 0\end{pmatrix}$ as $n\to\infty$ (since $J_i(\mu_i)^n \longrightarrow 0$). Hence

$$P^n \longrightarrow Q'$$

for some Q'. But, in particular, $Q' = \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} P^n = Q$, where Q is as in the proof of ergodicity for irreducible stationary Markov chains. Hence, since aperiodic $\Rightarrow$ irreducible, we have $Q'_{st} = p(t)$.

(iv)=>(ii): Let $\underline{s},\underline{t}\in S^{k+1}$. It suffices to show that

$$\mu(\mathcal{C}(\underline{s})\cap T^{-n}\mathcal{C}(\underline{t}))\longrightarrow\mu(\mathcal{C}(\underline{s}))\mu(\mathcal{C}(\underline{t}))$$

as $n\to\infty$, since this implies the characterization of strong mixing for simple functions, which are dense in $L^2(\mu)$.

But

$$\mu(\mathcal{C}(\underline{s})\cap T^{-n}\mathcal{C}(\underline{t}))=p(s_0)\cdot\prod_{k=0}^{\ell-1}P_{s_k s_{k+1}}\cdot(P^n)_{s_\ell t_0}\cdot\prod_{k=0}^{\ell-1}P_{t_k t_{k+1}}$$

$$\xrightarrow{n\to\infty}p(s_0)\prod_{k=0}^{\ell-1}P_{s_k s_{k+1}}\cdot p(t_0)\cdot\prod_{k=0}^{\ell-1}P_{t_k t_{k+1}}=\mu(\mathcal{C}(\underline{s}))\mu(\mathcal{C}(\underline{t})).$$

$\boxtimes$

End of lecture 22

Invariant Measures for Continuous Transformations

In what follows, we will focus on the following setup. (X,d) will be a compact metric space and $\mathcal{B}=\mathcal{B}(X)$ is the Borel σ-algebra on X. We denote by $\mathcal{M}_1(X)$ the set of Borel probability measures on X. We recall that any $\mu\in\mathcal{M}_1(X)$ is regular, i.e.,

$$\forall B\in\mathcal{B}\ \forall\varepsilon>0\ \exists C_\varepsilon\subseteq B \text{ compact } \exists U_\varepsilon\supseteq B \text{ open } \quad \mu(U_\varepsilon-C_\varepsilon)<\varepsilon.$$

In particular,

$$\forall B\in\mathcal{B}\quad \mu(B)=\inf_{\substack{B\subseteq U\\ U \text{ open}}}\mu(U)=\sup_{\substack{C\subseteq B\\ C\text{ compact}}}\mu(C).$$

We can map $\mathcal{M}_1(X)$ into $C(X)^*$ by defining

$$\forall\mu\in\mathcal{M}_1(X)\ \forall f\in C(X)\qquad \langle f,\mu\rangle:=\int f\,d\mu.$$

Then $\|\mu\|_{op}=1$, since $\langle \mathbb{1}_X,\mu\rangle=1$ and $|\langle f,\mu\rangle|\le\|f\|_\infty$.

This mapping is injective:

Theorem (without proof)

Let $\mu,\sigma\in\mathcal{M}_1(X)$. Then

$$\forall f\in C(X)\quad \langle f,\mu\rangle=\langle f,\sigma\rangle\Longrightarrow\mu=\sigma.$$

cf. Walters, §6.1.

In fact, this mapping is onto in some sense, i.e., $C(X)^*$ identifies with the signed measures on $\mathcal{B}$. The main step in this direction and for us the crucial result is the following.

Theorem (Riesz-Kakutani-Markov, without proof)

Let $\ell: C(X)\longrightarrow\mathbb{C}$ a bounded linear functional. Suppose that ℓ is positive, i.e., $\forall f\in C(X)\ f\geq 0\Rightarrow \ell(f)\geq 0$, and normalized, i.e., $\ell(\mathbb{1}_X)=1$.

Then there exists $\mu\in\mathcal{M}_1(X)$ s.t.

$$\forall f\in C(X)\quad \ell(f)=\langle f,\mu\rangle.$$

Definition

The weak-* topology on $C(X)^*$ is the smallest topology s.t. for all $f\in C(X)$ the map

$$\mathrm{ev}_f: C(X)^*\longrightarrow\mathbb{C},\qquad \ell\longmapsto\ell(f)$$

is continuous. The weak-* topology on $\mathcal{M}_1(X)$ is the restriction to $\mathcal{M}_1(X)$ of the weak-* topology on $C(X)^*$.

Remark: The weak-* topology on $\mathcal{M}_1(X)$ is generated by the open neighbourhoods of the form

$$N_{f_1,\dots,f_s;\varepsilon}(\mu)=\bigcap_{i=1}^{s}\left\{\sigma\in\mathcal{M}_1(X):\left|\int f_i\,d\mu-\int f_i\,d\sigma\right|<\varepsilon\right\},$$

where $s\in\mathbb{N}$, $\varepsilon>0$, $f_1,\dots,f_s\in C(X)$, $\mu\in\mathcal{M}_1(X)$ vary.

Theorem (exercise)

Let $(f_n)_{n\in\mathbb{N}}$ a dense sequence in $C(X)-\{0\}$ and define

$$D:\mathcal{M}_1(X)\times\mathcal{M}_1(X)\longrightarrow[0,\infty)$$
$$(\sigma,\mu)\longmapsto\sum_{n=1}^{\infty}\frac{|\langle f_n,\sigma\rangle-\langle f_n,\mu\rangle|}{2^n\|f_n\|_\infty}.$$

Then D is a metric inducing the weak-* topology on $\mathcal{M}_1(X)$.

Theorem (Special case of Banach-Alaoglu)

$B_1^{C(X)^*}(0)\subseteq C(X)^*$ is compact in the weak-* topology.

More generally: $(V,\|\cdot\|)$ Banach, then $\overline{B_1^{V^*}(0)}\subseteq V^*$ is compact.