WEAK CONVERGENCE

CHALMERS MSF500, SPRING 2019

MICHAEL BJÖRKLUND

1. Borel measures on metric spaces

1.1. Borel sets versus closed sets

Let (X, d) be a metric space. Given a subset $A \subseteq X$, we define $d(\cdot, A) : X \to [0, \infty)$ by

$$d(x, A) = \inf_{y \in A} d(x, y), \quad \text{for } x \in X. \tag{1.1}$$

The triangle inequality for d shows that

$$\big|d(x, A) - d(y, A)\big| \leqslant d(x, y), \quad \text{for all } x, y \in X,$$

so in particular, $x \mapsto d(x, A)$ is continuous, whence

$$\overline{A} = \{x \in X \mid d(x, A) = 0\},$$

and the set $A_r \subset X$ defined by

$$A_r = \{x \in X \mid d(x, A) < r\} \tag{1.2}$$

is open for every $r > 0$. Hence,

$$\overline{A} = \bigcap_{n \in \mathbb{N}} A_{1/n}. \tag{1.3}$$

Let us fix a Borel probability measure μ on X. Our aim here is to show that μ is completely determined by its values on closed sets. More specifically:

Lemma 1.1. *For every Borel set* $B \subset X$,

$$\begin{aligned} \mu(B) &= \sup\{\mu(C) \mid C \subset B \textit{ is closed}\} \\ &= \inf\{\mu(U) \mid U \supset B \textit{ is open}\}. \end{aligned}$$

Proof. Let $\mathcal{A}_\mu$ denote the set of all Borel sets $B \subset X$ for which

$$\begin{aligned} \mu(B) &= \sup\{\mu(C) \mid C \subset B \text{ is closed}\} \\ &= \inf\{\mu(U) \mid U \supset B \text{ is open}\}. \end{aligned}$$

We claim that $\mathcal{A}_\mu$ is a σ-algebra, containing all closed sets, whence must be equal to the Borel σ-algebra of X. To prove that $\mathcal{A}_\mu$ is a σ-algebra, first note that $\emptyset, X \in \mathcal{A}_\mu$ trivially, and that $\mathcal{A}_\mu$ is closed under complements. Hence it suffices to show that if $B_1, B_2, \ldots$ belong to $\mathcal{A}_\mu$, then so does $B := \cup_k B_k$. Fix $\varepsilon > 0$ and a sequence (ε_k) such that $\sum_k \varepsilon_k < \varepsilon/2$. Since $B_k \in \mathcal{A}_\mu$ for every k, we can find closet sets $C_k \subset X$ and open sets $U_k \subset X$ such that

$$C_k \subset B_k \subset U_k \quad \text{and} \quad \mu(U_k \setminus C_k) < \varepsilon_k, \quad \text{for all } k.$$

Set $U = \cup_k U_k$ and note that U is an open set with $B \subset U$ and

$$\mu(U \setminus B) \leqslant \sum_k \mu(U_k \setminus B_k) < \varepsilon.$$

Set $F = \cup_k C_k$. We stress that F might no longer be closed, but since μ is σ-additive, there exists $N \geqslant 1$ such that

$$\mu(F \setminus \bigcup_{k=1}^{N} C_k) < \varepsilon/2.$$

Set $C = \bigcup_{k=1}^{N} C_k$ and note that C is a closed subset of B with

$$\mu(B \setminus C) \leqslant \mu(B \setminus F) + \mu(F \setminus C) < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we conclude that $B \in \mathcal{A}_\mu$.

It remains to show that every closed subset $C \subset X$ belongs to $\mathcal{A}_\mu$, which amounts to showing that

$$\mu(C) = \inf\{\mu(U) \mid U \supset C \text{ is open}\}.$$

By (1.3),

$$C = \bigcap_{n \geqslant 1} C_{1/n},$$

where each $C_{1/n}$ is open, whence $\mu(C) = \lim_n \mu(C_{1/n})$, and we are done.

□

1.2. **Narrow convergence and the Portmanteau Lemma**

A sequence (μ_n) of bounded and positive Borel measures on X converges *narrowly* to a bounded and positive Borel measure μ on X if

$$\int_X f \, d\mu_n \to \int_X f \, d\mu, \quad \text{for all } f \in C_b(X).$$

Our aim here is to show that narrow convergence can be characterized in terms of the values of μ_n on *closed* subsets of X (the equivalence $(i) \iff (ii)$ below).

Lemma 1.2 (Portmanteau Lemma). *Let $\mu, \mu_1, \mu_2, \ldots$ be Borel probability measures on X. TFAE,*

(i) $\mu_n(f) \to \mu(f)$ *for every* $f \in C_b(X)$.
(ii) $\overline{\lim}_n \mu_n(C) \leqslant \mu(C)$ *for every closed set* $C \subset X$.
(iii) $\overline{\lim}_n \mu_n(f) \leqslant \mu(f)$ *for every upper semicontinuous function* $f : X \to \mathbb{R}$ *which is bounded from above.*
(iv) $\underline{\lim}_n \mu_n(U) \geqslant \mu(U)$ *for every open set* $U \subset X$.
(v) $\underline{\lim}_n \mu_n(f) \geqslant \mu(f)$ *for every lower semicontinuous function* $f : X \to \mathbb{R}$ *which is bounded from below.*

Proof. The implications

$$(ii) \iff (iv) \quad \text{and} \quad (iii) \iff (v)$$

are trivial. Furthermore, if $C \subset X$ is a closed set, then $f = \chi_C$ is an upper semicontinuous function (which is clearly bounded from above), and if $U \subseteq X$ is an open set, then $f = \chi_U$ is a lower semicontinuous function (which is clearly bounded from below). In view of this,

$$(iii) \implies (ii) \quad \text{and} \quad (v) \implies (iv)$$

are immediate. Since every $f \in C_b(X)$ is both lower and upper semicontinuous, (iii) and (v) together imply (i). It remains to prove

$$(i) \implies (ii) \quad \text{and} \quad (iv) \implies (v).$$

We begin with $(i) \implies (ii)$. Fix a closed set $C \subset X$ and define

$$\psi_N(x) = 1 - \left(\frac{d(x,C)}{1+d(x,C)}\right)^{\frac{1}{N}}, \quad \text{for } N \geqslant 1.$$

Note that ψ_N is continuous for every N and $1 \leqslant \psi_N(x) \searrow \chi_C(x)$ as $N \to \infty$ for all $x \in X$. In particular,

$$\mu(C) = \lim_N \int_X \psi_N \, d\mu = \lim_N \lim_n \int_X \psi_N \, d\mu_n \geqslant \overline{\lim_n} \int_X \chi_C \, d\mu_n = \overline{\lim_n}\, \mu_n(C),$$

which finishes the proof.

To prove $(iv) \implies (v)$, let us fix a lower semicontinuous function $f : X \to \mathbb{R}$ which is bounded from below by some constant M. Then $f - M \geqslant 0$, and thus

$$\int_X f \, d\mu_n - M = \int_0^\infty \mu_n(\{f - M > t\}) \, dt.$$

Since f, and thus $f - M$, is lower semicontinuous, the set $U_t = \{f - M > t\}$ is open. By our assumption (iv), we know that

$$\underline{\lim_n}\, \mu_n(U_t) \geqslant \mu(U_t), \quad \text{for all } t,$$

whence, by Fatou's Lemma,

$$\underline{\lim_n} \int_X f \, d\mu_n - M \geqslant \int_0^\infty \underline{\lim_n}\, \mu_n(\{f - M > t\}) \, dt \geqslant \int_0^\infty \mu(\{f - M > t\}) \, dt = \int_X f \, d\mu - M.$$

□

1.3. **Tightness**

A subset $M \subset \mathcal{P}(X)$ is *tight* if for every $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subset X$ such that

$$\inf_{\mu \in M} \mu(K_\varepsilon) \geqslant 1 - \varepsilon.$$

Theorem 1.3. *Suppose that (X, d) is separable. Then every tight subset $M \subset \mathcal{P}(X)$ is sequentially precompact.*

Towards the proof, let us fix a countable dense sequence (x_n) in X once and for all.

Step I: Embedding X into $[0,1]^{\mathbb{N}}$

Set $Z := [0,1]^{\mathbb{N}}$ and define the map $\varphi : X \to Z$ by

$$\varphi(x)_n = \frac{d(x, x_n)}{1 + d(x, x_n)}, \quad \text{for } x \in X.$$

It is easy to see that φ is continuous. We claim that φ is also injective. Indeed, if $\varphi(x) = \varphi(y)$, then

$$\frac{d(x, x_n)}{1 + d(x, x_n)} = \frac{d(y, x_n)}{1 + d(y, x_n)}, \quad \text{for all } n,$$

whence $d(x, x_n) = d(y, y_n)$ for all n. Fix $\varepsilon > 0$ and pick x_n such that $d(x, x_n) < \varepsilon/2$. Then, by the triangle inequality,

$$d(x, y) \leqslant d(x, x_n) + d(x_n, y) = 2d(x, x_n) < \varepsilon.$$

Since ε is arbitrary, we conclude that $d(x, y) = 0$, and thus $x = y$.

Step II: The closure of M in $\mathcal{P}([0,1]^{\mathbb{N}})$

Let (μ_n) be a sequence in $\mathcal{P}(X)$, and define $\nu_n \in \mathcal{P}(Z)$ by

$$\nu_n(C) = \mu_n(\varphi^{-1}(C)), \quad \text{for } C \subset Z \text{ Borel}.$$

Since Z is compact and metrizable, $\mathcal{P}(Z)$ is sequentially compact, and hence we can extract a subsequence (ν_{n_k}) which converges narrowly to a Borel probability measure ν on Z.

Step III: Pulling back from $\mathcal{P}([0,1]^{\mathbb{N}})$

Since M is tight, there exists for every integer $N \geqslant 1$ a compact set $K_N \subset X$ such that

$$\mu_n(K_N) \geqslant 1 - 1/N \quad \text{and} \quad \text{for all } n.$$

In particular, the set $Y := \bigcup_N K_N$ is σ-compact and satisfies $\mu_n(Y) = 1$ for all n. Since $\varphi : Y \to Z$ is continuous and injective, Exercise 8 shows that $\varphi(Y)$ is a Borel set in Z. We claim that $\nu(\varphi(Y)) = 1$. Indeed, since $\varphi(K_N)$ is compact (and hence closed) in Z for every N, it follows from Lemma (1.2) that

$$\nu(\varphi(Y)) \geqslant \nu(\varphi(K_N)) \geqslant \varlimsup_k \nu_{n_k}(\varphi(K_N)) = \varlimsup_k \mu_{n_k}(K_N) \geqslant 1 - \frac{1}{N}$$

for every N, whence $\nu(\varphi(Y)) = 1$. Set

$$\mu(B) = \nu(\varphi(B \cap Y)), \quad \text{for } B \subset X \text{ Borel}.$$

By Exercise 8, μ defines a Borel probability measure on X. It remains to show that the sequence μ_{n_k} narrowly converges to μ. By Lemma 1.2 it suffices to show that

$$\varlimsup_k \mu_{n_k}(C) \leqslant \mu(C), \quad \text{for every closed set } C \subset X.$$

Pick a closed set $C \subset X$. Then, for every N,

$$\mu_{n_k}(C) = \mu_{n_k}(C \cap K_N) + \mu_{n_k}(C \cap K_N^c) \leqslant \nu_{n_k}(\varphi(C \cap K_N)) + \frac{1}{N}.$$

Since $\nu_{n_k} \to \nu$ narrowly and $\varphi(C \cap K_N) \subset Z$ is compact, and hence closed, we know from Lemma 1.2 that $\varlimsup_k \nu_{n_k}(\varphi(C \cap K_N)) \leqslant \nu(\varphi(C \cap K_N)) = \mu(C \cap K_N)$, and thus

$$\varlimsup_k \mu_{n_k}(C) \leqslant \mu(C \cap K_N) + \frac{1}{N} \leqslant \mu(C) + \frac{1}{N}.$$

Since $N \geqslant 1$ is arbitrary, we are done.

2. Tightness through Fourier transforms

Let μ be a Borel measure on $\mathbb{R}^d$ with finite measure. The *Fourier transform* $\widehat{\mu}$ of μ is defined by

$$\widehat{\mu}(\xi) = \int_{\mathbb{R}} e^{i\xi, x}\, d\mu(x), \quad \text{for } \xi \in \mathbb{R}^d.$$

It is easy to see that $\xi \mapsto \widehat{\mu}(\xi)$ is uniformly continuous. Furthermore, by Cauchy-Schwarz,

$$\begin{aligned} |\widehat{\mu}(\xi + \eta) - \widehat{\mu}(\xi)| &\leqslant \Big(\int_{\mathbb{R}} |e^{i\eta x} - 1|^2\, d\mu(x) \Big)^{1/2} \\ &= \Big(\int_{\mathbb{R}} 2(1 - \cos(\xi x))\, d\mu(x) \Big)^{1/2} \\ &\leqslant \big(2(1 - \operatorname{Re} \widehat{\mu}(\xi)\big)^{1/2} \leqslant \big(2(1 - |\widehat{\mu}(\xi)|)\big)^{1/2} \end{aligned}$$

2.1. **Levy's Continuity Theorem and Bochner's Theorem**

2.2. **Infinitely divisible probability measures and Levy-Khinchin's representation**

3. Decorrelation of Wiener sequences and Hellinger distances

Let (X, d) be a metric space. Let μ and ν be bounded and positive Borel measures on X and fix a bounded and positive Borel measure on X such that $\mu \ll \lambda$ and $\nu \ll \lambda$; for instance, $\lambda = \mu + \nu$. Set

$$d\mu = u\, d\lambda \quad \text{and} \quad d\nu = \nu\, d\lambda,$$

for some non-negative Borel functions $u, \nu : X \to [0, \infty]$. The *Hellinger distance* $\mathrm{dist}_H(\mu, \nu)$ is defined as

$$\mathrm{dist}_H(\mu, \nu) = \frac{1}{2} \int_X \left(\sqrt{u} - \sqrt{\nu}\right)^2 d\lambda = \frac{1}{2}\big(\mu(X) + \nu(X)\big) - H(\mu, \nu)\big), \tag{3.1}$$

where

$$H(\mu, \nu) = \int_X \sqrt{u\nu}\, d\lambda.$$

You are encouraged to check that this definition is independent of the choice of λ (Exercise 1).

Our aim here is to show:

Theorem 3.1. *The Hellinger distance is (sequentially) lower semicontinuous with respect to narrow convergence, i.e. if μ and ν are bounded and positive Borel measures on X, and (μ_n) and (ν_n) are two sequences of bounded and positive Borel measures on X such that $\mu_n \to \mu$ and $\nu_n \to \nu$ in the narrow topology, then*

$$\varliminf_n \mathrm{dist}_H(\mu_n, \nu_n) \geqslant \mathrm{dist}_H(\mu, \nu),$$

or equivalently,

$$\varlimsup_n H(\mu_n, \nu_n) \leqslant H(\mu, \nu).$$

The key to Theorem 3.1 is the following lemma.

Lemma 3.2. *Let μ and ν be bounded and positive Borel measures on X. There exists a function*

$$\gamma : [0, 1) \to [0, 1), \quad \textit{with} \quad \lim_{t \to 0} \gamma(t) = 0,$$

such that for every $\kappa > 0$ and for every $\varepsilon > 0$, which is small enough, there are $N = N_{\kappa,\varepsilon}$ and (Lipschitz) continuous functions $f_0, f_1, \ldots, f_N : X \to [0, 1]$ with

$$\sum_{k=0}^{N} f_k(x) = 1, \quad \textit{for all } x \in X,$$

such that

$$H(\mu, \nu) \geqslant \frac{1}{1+\varepsilon} \sum_{k=0}^{N} \Big(\int_X f_k\, d\mu\Big)^{\frac{1}{2}} \Big(\int_X f_k\, d\nu\Big)^{\frac{1}{2}} - \gamma(\varepsilon) - \kappa.$$

3.0.1. *Proof of Theorem 3.1 assuming Lemma 3.2*

Let μ and ν be bounded and positive Borel measures on X, and let (μ_n) and (ν_n) be two sequences of bounded and positive Borel measures on X such that $\mu_n \to \mu$ and $\nu_n \to \nu$ in the narrow topology. We want to show that

$$H(\mu, \nu) \geqslant \varlimsup_n H(\mu_n, \nu_n).$$

By Lemma 3.2 we can find a function $\gamma : [0,1) \to [0,1)$ with $\lim_{t\to 0}\gamma(t) = 0$, such that for every $\kappa > 0$ and for all small enough $\varepsilon > 0$, there are $N = N_\varepsilon$ and continuous functions $f_0, f_1, \ldots, f_N$ with

$$\sum_{k=0}^{N} f_k(x) = 1, \quad \text{for all } x \in X$$

such that

$$H(\mu,\nu) \geqslant \frac{1}{1+\varepsilon}\sum_{k=0}^{N}\Big(\int_X f_k\,d\mu\Big)^{\frac{1}{2}}\Big(\int_X f_k\,d\nu\Big)^{\frac{1}{2}} - \gamma(\varepsilon) - \kappa.$$

Let η be positive and bounded Borel measure on X such that $\mu_n \ll \eta$ and $\nu_n \ll \eta$ for all n. For instance, $\eta = \sum_{n\geqslant 1}\frac{1}{2^n\mu_n(X)}\mu_n$ will do. We write

$$d\mu_n = u_n\,d\eta \quad \text{and} \quad d\nu_n = \nu_n\,d\eta, \quad \text{for all } n,$$

where $u_n, \nu_n : X \to [0,\infty]$ are Borel measurable functions. Since $\mu_n \to \mu$ and $\nu_n \to \nu$ in the narrow topology, and each f_k is continuous, we see that

$$\int_X f_k\,d\mu_n \to \int_X f_k\,d\mu \quad \text{and} \quad \int_X f_k\,d\nu_n \to \int_X f_k\,d\nu, \quad \text{for all } k = 0,1,\ldots,N,$$

and thus

$$\begin{aligned}
H(\mu,\nu) &\geqslant \frac{1}{1+\varepsilon}\sum_{k=0}^{N}\Big(\int_X f_k\,d\mu\Big)^{\frac{1}{2}}\Big(\int_X f_k\,d\nu\Big)^{\frac{1}{2}} - \gamma(\varepsilon) - \kappa \\
&= \lim_n \frac{1}{1+\varepsilon}\sum_{k=0}^{N}\Big(\int_X f_k\,d\mu_n\Big)^{\frac{1}{2}}\Big(\int_X f_k\,d\nu_n\Big)^{\frac{1}{2}} - \gamma(\varepsilon) - \kappa \\
&= \lim_n \frac{1}{1+\varepsilon}\sum_{k=0}^{N}\Big(\int_X f_k\,u_n\,d\eta\Big)^{\frac{1}{2}}\Big(\int_X f_k\,\nu_n\,d\eta\Big)^{\frac{1}{2}} - \gamma(\varepsilon) - \kappa \\
&\geqslant \overline{\lim_n}\, \frac{1}{1+\varepsilon}\sum_{k=0}^{N}\int_X f_k\sqrt{u_n\nu_n}\,d\eta - \gamma(\varepsilon) - \kappa \\
&= \overline{\lim_n}\, \frac{1}{1+\varepsilon}\int_X \sqrt{u_n\nu_n}\,d\eta - \gamma(\varepsilon) - \kappa \\
&= \frac{1}{1+\varepsilon}\,\overline{\lim_n}\, H(\mu_n,\nu_n) - \gamma(\varepsilon) - \kappa,
\end{aligned}$$

where we in the second inequality used Cauchy-Schwarz inequality. Since $\varepsilon > 0$ and $\kappa > 0$ are arbitrary and $\gamma(\varepsilon) \to 0$ when $\varepsilon \to 0$, we are done.

3.0.2. *Proof of Lemma 3.2*

Fix $\varepsilon > 0$ and a bounded and positive Borel measure λ on X such that $\mu \ll \lambda$ and $\nu \ll \lambda$. We write

$$d\mu = u\,d\lambda \quad \text{and} \quad d\nu = \nu\,d\lambda,$$

and set

$$C = \{x \in X \mid \nu(x) = 0\} \quad \text{and} \quad D = \{x \in X \setminus C \mid u(x) = 0\}.$$

We define

$$A_k = \Big\{x \in X \setminus C \mid (1+\varepsilon)^k \leqslant \frac{u(x)}{\nu(x)} < (1+\varepsilon)^{k+1}\Big\}, \quad \text{for } k \in \mathbb{Z},$$

and note that $X = C \sqcup D \sqcup \left(\sqcup_{k\in\mathbb{Z}} A_k \right)$. In particular, $\sum_k \nu(A_k) \leqslant \nu(X) < \infty$, so for any $\delta > 0$ we can find $M_\delta \geqslant 1$ such that

$$\sum_{|k|\geqslant M_\delta} \nu(A_k) < \delta.$$

Given δ, we set

$$B_o = C \sqcup \left(\sqcup_{|k|\geqslant M_\delta} A_k \right) \quad \text{and} \quad B_1 = D \quad \text{and} \quad B_k = A_{k-1-M_\delta},$$

for $k = 2, \ldots, 2M_\delta$, so that $B_0, B_1, \ldots, B_{2M_\delta}$ is a partition of X into Borel sets, and

$$\nu(B_o) < \delta \quad \text{and} \quad \mu(B_1) = 0. \tag{3.2}$$

By Exercise 3 we can find a function $\beta : [0,1) \to [0,1)$ with $\lim_{t\to 0} \beta(t) = 0$ and continuous functions $f_0, f_1, \ldots, f_{2M_\delta} : X \to [0,1]$ such that

$$\sum_{k=0}^{2M_\delta} f_k(x) = 1 \quad \text{for all } x \in X,$$

and with the following properties:

(i) If $\mu(B_k) = 0$, then $\int_X f_k \, d\mu < \varepsilon$.
(ii) If $\nu(B_k) = 0$, then $\int_X f_k \, d\nu < \varepsilon$.
(iii) If $\mu(B_k)\nu(B_k) > 0$, then

$$\beta(\varepsilon) \leqslant \int_X f_k \, d\mu \leqslant (1+\varepsilon)^{\frac{1}{2}} \mu(B_k) + \beta(\varepsilon), \tag{3.3}$$

and

$$\beta(\varepsilon) \leqslant \int_X f_k \, d\nu \leqslant (1+\varepsilon)^{\frac{1}{2}} \nu(B_k) + \beta(\varepsilon). \tag{3.4}$$

We further note that for $k \geqslant 2$,

$$\mu(B_k) = \int_{B_k} u \, d\lambda = \int_{B_k} \frac{u}{\nu} \nu \, d\lambda \leqslant (1+\varepsilon)^{k-M_\delta} \nu(B_k),$$

and thus

$$\begin{aligned} \int_{B_k} \sqrt{u\nu} \, d\lambda &= \int_{B_k} \sqrt{\frac{u}{\nu}} \nu \, d\lambda \\ &\geqslant (1+\varepsilon)^{(k-1-M_\delta)/2} \nu(B_k) \\ &\geqslant (1+\varepsilon)^{-\frac{1}{2}} \mu(B_k)^{\frac{1}{2}} \nu(B_k)^{\frac{1}{2}}. \end{aligned}$$

By (3.3), we can now conclude that if $\mu(B_k)\nu(B_k) > 0$, then

$$\begin{aligned} \int_{B_k} \sqrt{u\nu} \, d\lambda &\geqslant (1+\varepsilon)^{-1} \Big(\int_X f_k \, d\mu - \beta(\varepsilon) \Big)^{\frac{1}{2}} \Big(\int_X f_k \, d\nu - \beta(\varepsilon) \Big)^{\frac{1}{2}} \\ &\geqslant (1+\varepsilon)^{-1} \Big(\Big(\int_X f_k \, d\mu \Big)^{\frac{1}{2}} - \sqrt{\beta(\varepsilon)} \Big) \Big(\Big(\int_X f_k \, d\nu \Big)^{\frac{1}{2}} - \sqrt{\beta(\varepsilon)} \Big) \\ &\geqslant (1+\varepsilon)^{-1} \Big(\int_X f_k \, d\mu \Big)^{\frac{1}{2}} \Big(\int_X f_k \, d\nu \Big)^{\frac{1}{2}} - 2\sqrt{\beta(\varepsilon)}, \end{aligned}$$

if ε is small enough. Hence,

$$\begin{aligned}
H(\mu,\nu) &\geqslant \sum_{k=2}^{2M_\delta} \int_{B_k} \sqrt{uv}\, d\lambda \\
&\geqslant (1+\varepsilon)^{-1} \sum_{k=2}^{2M_\delta} \Big(\int_X f_k\, d\mu\Big)^{\frac{1}{2}} \Big(\int_X f_k\, d\nu\Big)^{\frac{1}{2}} - 4M_\delta \sqrt{\beta(\varepsilon)} \\
&= (1+\varepsilon)^{-1} \sum_{k=0}^{2M_\delta} \Big(\int_X f_k\, d\mu\Big)^{\frac{1}{2}} \Big(\int_X f_k\, d\nu\Big)^{\frac{1}{2}} - 4M_\delta \sqrt{\beta(\varepsilon)} \\
&- (1+\varepsilon)^{-1}\Big(\Big(\int_X f_0\, d\mu\Big)^{\frac{1}{2}} \Big(\int_X f_0\, d\nu\Big)^{\frac{1}{2}} + \Big(\int_X f_1\, d\mu\Big)^{\frac{1}{2}} \Big(\int_X f_1\, d\nu\Big)^{\frac{1}{2}}\Big) \\
&\geqslant (1+\varepsilon)^{-1} \sum_{k=0}^{2M_\delta} \Big(\int_X f_k\, d\mu\Big)^{\frac{1}{2}} \Big(\int_X f_k\, d\nu\Big)^{\frac{1}{2}} - 4M_\delta \sqrt{\beta(\varepsilon)} \\
&- C\max(\sqrt{\varepsilon}, \sqrt{\delta}),
\end{aligned}$$

for some constant C. Fix $\kappa > 0$ and choose $\delta = \delta_{\kappa,\varepsilon} > 0$ so that the last term is less than κ. By setting

$$N = 2M_{\delta,\varepsilon} \quad \text{and} \quad \gamma(\varepsilon) = 4M_\delta \sqrt{\beta(\varepsilon)},$$

we are done.

3.1. **An application to decorrelation of Wiener sequences**

Let (a_n) be a bounded sequence of complex numbers and *assume* that the limit

$$\gamma_a(n) := \lim_N \frac{1}{N} \sum_{m=0}^{N-1} a_m \overline{a}_{m+n}$$

exists for all $n \geqslant 0$. If this is the case, then we say that (a_n) is a *Wiener sequence* and we refer to γ_o as its *autocorrelation*. If we extend γ_o to a function $\gamma_a : \mathbb{Z} \to \mathbb{C}$ by setting $\gamma_a(-n) = \overline{\gamma_a(n)}$ for $n < 0$, then γ_a is a positive definite function (Definition A.1) on $\mathbb{Z}$ (Exercise 2). Let θ_{γ_a} denote the spectral measure associated to γ_a (Definition (A.3). By Exercise 2, the sequence $(\theta_N^{(a)})$ of bounded positive Borel measure measures on $\mathbb{T}$ defined by

$$\int_{\mathbb{T}} f\, d\theta_N^{(a)} = \int_{\mathbb{T}} f(x) \Big| \frac{1}{N} \sum_{n=0}^{N-1} a_n e^{-2\pi i n x} \Big|^2 d\lambda(x), \quad \text{for } f \in C(\mathbb{T}), \tag{3.5}$$

converges in the narrow topology to θ_{γ_a}, where λ denotes the Lebesgue probability measure on $\mathbb{T}$.

Let us now fix two Wiener sequences (a_n) and (b_n) with autocorrelations (γ_a) and γ_b respectively, with associated spectral measures θ_{γ_a} and θ_{γ_b}. Let $(\theta_N^{(a)})$ and $(\theta_N^{(b)})$ be the sequences of bounded positive Borel measures on $\mathbb{T}$ defined in (3.5), which narrowly converge to θ_{γ_a} and θ_{γ_b} respectively. Note that both $\theta_N^{(a)}$ and $\theta_N^{(b)}$ are absolutely continuous with respect to the Lebesgue probability measure λ on $\mathbb{T}$, whence

$$\begin{aligned} H(\theta_N^{(a)},\theta_N^{(b)}) &= \int_{\mathbb{T}} \frac{1}{N}\Big|\sum_{m=0}^{N-1} a_m e^{-2\pi i m x}\Big|\Big|\sum_{n=0}^{N-1} b_n e^{-2\pi i n x}\Big|\, d\lambda(x) \\ &\geqslant \frac{1}{N}\Big|\int_{\mathbb{T}} \sum_{m,n=0}^{N-1} a_m \overline{b}_n e^{2\pi i (n-m)x}\, d\lambda(x)\Big| \\ &= \frac{1}{N}\Big|\sum_{n=0}^{N-1} a_n \overline{b}_n\Big|. \end{aligned}$$

By Theorem 3.1, we now have the following result, due to Coquet, Kamae and Mendes France.

Theorem 3.3. *For any two Wiener sequences (a_n) and (b_n), we have*

$$\overline{\lim_N}\Big|\frac{1}{N}\sum_{n=0}^{N-1} a_n \overline{b}_n\Big| \leqslant H(\theta_{\gamma_a},\theta_{\gamma_b}),$$

where θ_{γ_a} and θ_{γ_b} denote the spectral measures associated to the autocorrelations γ_a and γ_b of (a_n) and (b_n) respectively.

4. A crash course in ergodic theory

4.1. **Measurable aspects**

Let (X,μ) be a Borel probability measure space. A measurable map $T: X \to X$ is said to *preserve* the measure μ if

$$\mu(T^{-1}B) = \mu(B), \quad \text{for every Borel set } B \subset X.$$

If T preserves μ, we say that (X,μ,T) is a *probability measure preserving system*. In this case, T induces (for every $p \geqslant 1$) an isometric linear map $T: L^p(X,\mu) \to L^p(X,\mu)$ by $f \mapsto f \circ T$ (see exercises below). We say that (X,μ,T) is *ergodic* if there is no Borel set $B \subset X$ with $0 < \mu(B) < 1$ such that $T^{-1}(B) = B$. It is not hard to show (exercise below) that (X,μ,T) is ergodic if and only if $U_T : L^p(X,\mu) \to L^p(X,\mu)$ (for some p) does not have a non-constant fixed point. Let us consider two examples.

Example 4.1. Let $X = \mathbb{R}/\mathbb{Z}$ and $Tx = x + \alpha \pmod 1$ for some *irrational* α. It is plain to see that T preserves the Lebesgue probability measure μ on X (we can think of X as the interval $[0,1]$ with the end points identified). We claim that (X,μ,T) is ergodic. Indeed, suppose that $f \in L^2(X,\mu)$ satisfies $U_T f = f$. We wish to show that f is essentially constant. To do this, expand f in a Fourier series,

$$f = \sum_{n\in\mathbb{Z}} c_n e^{2\pi i n \cdot},$$

where the convergence of the series is taken in the Hilbert space sense. The equation $U_T f = f$ now translates to

$$\sum_n c_n e^{2\pi n \alpha} e^{2\pi i n \cdot} = \sum_n c_n e^{2\pi i n \cdot},$$

whence $c_n e^{2\pi i n \alpha} = c_n$ for all n. We conclude that for every n, we either have $c_n = 0$ or $e^{2\pi i n \alpha} = 1$. Since α is irrational, the second case only happens for $n = 0$, and thus $f = c_o$, i.e. f is (essentially) a constant.

Example 4.2. Let $X = \mathbb{R}/\mathbb{Z}$ and $Tx = 2x \pmod 1$. We leave it is an exercise (see below) to show that T preserves the Lebesgue measure μ on X. We claim that (X, μ, T) is ergodic, and we will argue as in the previous example: Pick $f \in L^2(X, \mu)$ such that $U_T f = f$, and expand in a Fourier series as above. Then $c_{2n} = c_n$ for all n, whence $c_{2^k n} = c_n$ for all n and $k \geqslant 1$. In particular, for every $n \neq 0$, by Parseval's Theorem,

$$\infty > \int_{\mathbb{T}} |f|^2 \, d\mu = \sum_m |c_m|^2 \geqslant \sum_{k \geqslant 1} |c_{2^k n}|^2 = \sum_{k \geqslant 1} |c_n|^2,$$

which forces $c_n = 0$. Hence $f = c_o$, and we are done.

Our aim is to show the following classical theorem of George Birkhoff.

Theorem 4.1 (The Pointwise Ergodic Theorem)**.** *Let* (X, T, μ) *be an ergodic probability measure preserving system. Then, for every* μ*-integrable* $f : X \to \mathbb{R}$*,*

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) = \int_X f \, d\mu, \quad \textit{for } \mu\textit{-almost every } x \in X.$$

The proof will be broken down into two lemmas. We will use the notations,

$$(A_n f)(x) = \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x)$$

and

$$(M_N f)(x) = \sup_{1 \leqslant n \leqslant N} (A_n f)(x) \quad \text{and} \quad (Mf)(x) = \sup_n (A_n f)(x),$$

for $x \in X$.

Lemma 4.2 (Maximal inequality)**.** *For every* $\lambda > 0$ *and* $f \in L^1(X, \mu)$*,*

$$\mu\big(\big\{x \in X \mid |Mf(x)| > \lambda\big\}\big) \leqslant \frac{1}{\lambda} \int_X |f| \, d\mu.$$

Lemma 4.3 (Approximation)**.** *Set* $C = \operatorname{span}\{g - g \circ T \mid g \in L^1(X, \mu)\big\}$. *Then,*

$$L^1(X, \mu) = \mathbb{R}1 \oplus \overline{C}.$$

Proof of Theorem 4.1 assuming Lemma 4.2 and Lemma 4.3. The theorem is trivial for $f \in \mathbb{R} \oplus C$. Indeed, it is trivial for constants, if $f = g - g \circ g$, then by telescoping,

$$A_n f = \frac{g - g \circ T^n}{n},$$

which clearly goes to zero as $n \to \infty$ (for the second term, use Borel-Cantelli). Set

$$\mathcal{P} = \Big\{f \in L^1(X, \mu) \mid A_n f(\cdot) \to \int_X f \, d\mu, \ \ \mu\text{-almost everywhere}\Big\}.$$

We wish to prove that $\mathcal{P} = L^1(X, \mu)$. The argument above shows that $\mathcal{P}$ is dense (since it contains $\mathbb{R} \oplus C$, which is dense by Lemma 4.3), so it suffices to show that $\mathcal{P}$ is closed in the L^1-norm topology. Let (f_m) be a sequence in $\mathcal{P}$ which converges to f in the L^1-norm. Set

$$\Delta(x) := \overline{\lim_n} \Big|A_n f(x) - \int_X f \, d\mu\Big|.$$

We wish to prove that $\Delta = 0$ μ-almost everywhere. Note that for all m,

$$\begin{aligned}\Delta(x) &\leqslant \overline{\lim_n}\Big(\big|A_n f_m(x) - \int_X f_m\,d\mu\big| + \|f - f_m\|_1 + |M(f-f_m)(x)|\Big)\\ &\leqslant \|f - f_m\|_1 + |M(f-f_m)(x)|,\end{aligned}$$

since $f_m \in \mathcal{P}$, whence, for all $\lambda > 0$ and m,

$$\mu\big(\{\Delta(x) > \lambda\}\big) \leqslant \mu\big(\{M(f-f_m)(x) > \lambda - \delta_m\}\big) \leqslant \frac{\delta_m}{\lambda - \delta_m},$$

where $\delta_m = \|f - f_m\|_1$, and where we in the last inequality used Lemma 4.2. Since $\delta_m \to 0$, we conclude that $\Delta = 0$ μ-almost everywhere. □

4.1.1. *Proof of Lemma 4.2*

By replacing f with $f - \lambda$, it suffices to show that

$$\int_{Mf>0} f\,d\mu \geqslant 0.$$

Since $Mf(x) > 0$, if and only if $Q_N f(x) = \max_{1\leqslant n\leqslant N} \sum_{k=0}^{n-1} f(T^n x) > 0$ for large enough N, it is enough to show

$$\int_{Q_N f>0} f\,d\mu \geqslant 0, \quad \text{for large enough } N.$$

Since

$$Q_N f(x) \leqslant Q_{N+1} f(x) = \max(0, f(x) + Q_N f(Tx)), \quad \text{for all } N,$$

we see that

$$\int_X Q_N f(x)\,d\mu(x) \leqslant \int_{Q_N f>0} f\,d\mu + \int_X Q_N f(Tx)\,d\mu(x),$$

and thus $\int_{Q_N f>0} f\,d\mu$ since T preserves μ.

4.1.2. *Proof of Lemma 4.3*

If $\mathbb{R}1 \oplus C$ is not dense in $L^1(X,\mu)$, then by Hahn-Banach's Theorem, there exists a non-zero $h \in L^\infty(X,\mu) = L^1(X,\mu)^*$ such that

$$\int_X h\,d\mu = 0 \quad \text{and} \quad \int_X (g - g\circ T)h\,d\mu = 0, \quad \text{for all } g \in L^1(X,\mu).$$

The second set of inequalities is clearly equivalent to saying $U_T^* h = h$, where $U_T^* : L^\infty(X,\mu) \to L^\infty(X,\mu)$ denotes the transpose of U_T. By Exercise 13, we conclude that h is constant, hence identically zero (since its integral is zero).

4.2. **Topological aspects**

Let us now consider the case when X is a compact metrizable space. Suppose that $T : X \to X$ is a continuous map which preserves a Borel probability measure μ. Let us further assume that (X,μ,T) is ergodic. By the Pointwise Ergodic Theorem, there exists for every $f \in C(X)$, a conull set $X_f \subset X$ such that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) = \int_X f\,d\mu, \quad \text{for all } x \in X_f.$$

In one of the exercises below you are asked to prove that $C(X)$ is separable. Assuming this for now, we can pick a countable dense set (f_j) in $C(X)$, and define the μ-conull subset X_{gen} of *μ-generic points* by

$$X_{gen} = \bigcap_j X_{f_j} \subset X.$$

Lemma 4.4. *For every $f \in C(X)$, we have*

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) = \int_X f \, d\mu, \quad \textit{for all } x \in X_{gen}.$$

Proof. Fix $f \in C(X)$ and $\varepsilon > 0$. Pick f_j such that $\|f - f_j\|_\infty < \varepsilon/2$. Then, if $x \in X_{gen}$,

$$\varlimsup_n \Big| \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) - \int_X f \, d\mu \Big| \leqslant 2\|f - f_j\|_\infty + \varlimsup_n \Big| \frac{1}{n} \sum_{k=0}^{n-1} f_j(T^k x) - \int_X f_j \, d\mu \Big| < \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we are done. □

Corollary 4.5. *Suppose that μ and ν are two different ergodic and T-invariant Borel probability measures on X. Then $\mu \perp \nu$. In fact, if X_μ denotes the set of μ-generic points in X, then $\mu(X_\mu) = 1$, while $\nu(X_\mu) = 0$.*

Proof. Let X_μ and X_ν denote the set of μ-generic points and the set of ν-generic points respectively, so that $\mu(X_\mu) = 1$ and $\nu(X_\nu) = 1$. We claim that $X_\mu \cap X_\nu = \emptyset$, whence $\mu \perp \nu$. Since $\mu \neq \nu$, there exists $f \in C(X)$ such that $\mu(f) \neq \nu(f)$. For this f, we have for every $x \in X_\mu \cap X_\nu$,

$$\int_X f \, d\mu = \lim_n \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) = \int_X f \, \nu,$$

which is clearly impossible. □

We say that a continuous map $T : X \to X$ is *uniquely ergodic* if there exists exactly one T-invariant Borel probability measure on X.

Lemma 4.6. *Suppose that $T : X \to X$ is uniquely ergodic, and let μ denote the unique T-invariant Borel probability measure on X. Then, for every $f \in X$,*

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) = \int_X f \, d\mu, \quad \textit{uniformly in } x.$$

Proof. Let (x_n) be a sequence in X such that $x_n \to x$. Consider the sequence of probability measures,

$$\nu_n = \frac{1}{n} \sum_{k=0}^{n-1} \delta_{T^k x_n}.$$

By weak*-compactness, we can extract a convergent sub-sequence (n_N). We claim that the limit measure $\nu = \lim_N \nu_{n_N}$ is T-invariant, whence equal to μ, which would finish the proof. To prove invariance, note that

$$T_* \nu_N - \mu_N = \frac{1}{n_N} \sum_{k=0}^{n_N - 1} \big(\delta_{T^{k+1} x} - \delta_{T^k x} \big) = \frac{1}{n_N} (\delta_{T^{n_N} x} - \delta_x) \to 0,$$

in the weak*-topology. □

The fact that the limit measure ν is T-invariant does not assume unique ergodicity, and the same argument as in the proof above can be applied to prove the following classical fact.

Scholium 4.7 (Krylov-Bogliouboff)**.** *For every continuous map* $T : X \to X$*, there exists at least one* T*-invariant probability measure on* X*.*

Example 4.3. Let (X, T) be as in Example 4.1. We claim that T is uniquely ergodic (Lebesgue measure μ is the unique T-invariant probability measure) To prove this, suppose that ν is a T-invariant probability measure on $\mathbb{R}/\mathbb{Z}$. We shall prove that its Fourier transform satisfies $\widehat{\nu}(n) = 0$ for all $n \geqslant 0$, which means that $\nu = \mu$. Since $T_*\mu = \mu$, we have

$$\widehat{\nu}(n)e^{2\pi i n\alpha} = \widehat{\nu}(n), \quad \text{for all } n.$$

If $\widehat{\nu}(n) \neq 0$, we must have $e^{2\pi i n\alpha} = 1$, which clearly forces $n = 0$ since α is irrational. We are done. As a corollary, we get that

$$\lim_n \frac{1}{n}\sum_{k=0}^{n-1} f(x + k\alpha) = \int_X f\, d\mu, \quad \text{for all } x \in X \text{ and } f \in C(X).$$

4.3. **Existence of ergodic measures**

Let X be a compact metrizable space and let $T : X \to X$ be a continuous map. By the Scholium above, there exists at least one T-invariant probability measure on X. In this subsection, we shall prove that there exists in fact always at least one *ergodic* T-invariant probability measure. We begin by formulating another characterization of ergodic measures.

Lemma 4.8. *A* T*-invariant Borel probability measure* μ *on* X *is* ergodic *if and only if it cannot be written on the form*

$$\mu = \alpha\mu_1 + (1-\alpha)\mu_2,$$

for some $0 < \alpha < 1$ *and two different* T*-invariant Borel probability measures* μ_1 *and* μ_2 *on* X*.*

Proof. If μ is not ergodic, we can find a T-invariant Borel set $B \subset X$ with $0 < \mu(B) < 1$, and thus

$$\mu = \mu(B)\frac{\mu(\cdot \cap B)}{\mu(B)} + (1 - \mu(B))\frac{\mu(\cdot \cap B^c)}{\mu(B^c)},$$

where both measures on the right hand side are T-invariant (and clearly different). Conversely, suppose that μ can be written as

$$\mu = \alpha\mu_1 + (1-\alpha)\mu_2,$$

for some $0 < \alpha < 1$, where μ_1 and μ_2 are T-invariant. Then μ_1 and μ_2 are absolutely continuous with respect to μ. Their Radon-Nikodym derivatives $\frac{d\mu_1}{d\mu}$ and $\frac{d\mu_2}{d\mu}$ are clearly T-invariant, and define non-constant T-invariant Borel functions on X, whence μ is not ergodic. □

Proposition 4.9. *For every continuous map* $T : X \to X$*, there exists at least one ergodic* T*-invariant Borel probability measure.*

Proof. We shall show that there exists a T-invariant Borel probability measure on X which cannot be written on the form

$$\mu = \alpha\mu_1 + (1-\alpha)\mu_2, \tag{4.1}$$

for some $0 < \alpha < 1$ and two different T-invariant Borel probability measures μ_1 and μ_2. To do this, let us fix a countable dense subset (f_j) of $C(X)$ (which exists by one of the exercises below). By another exercise, the set $\mathcal{P}_T(X)$ of T-invariant Borel probability measures on X is weak*-compact. Set

$$M_o = \big\{\nu \in \mathcal{P}_T(X) \mid \sup_{\mu \in \mathcal{P}_T(X)} \mu(f_o) = \nu(f_o)\big\}.$$

Since $\mu \mapsto \mu(f)$ is weak*-continuous for every $f \in C(X)$ and $\mathcal{P}_T(X)$ is weak*-compact and non-empty, M_o is non-empty. Now inductively define for $j \geqslant 1$,

$$M_j = \big\{\nu \in \mathcal{P}_T(X) \mid \sup_{\mu \in M_{j-1}} \mu(f_j) = \nu(f_j)\big\}.$$

By the same argument, M_j is non-empty, so by weak*-compactness,

$$M_\infty = \bigcap_j M_j,$$

is non-empty as well. We claim that no $\mu \in M_\infty$ can be written on the form (4.1), and is thus ergodic. Indeed, suppose that $\mu \in M_\infty$ can be written on this form. Then,

$$\mu(f_o) = \alpha\mu_1(f_o) + (1-\alpha)\mu_2(f_o) = \sup_{\nu \in P_T(X)} \nu(f_o),$$

whence $\mu(f_o) = \mu_1(f_o) = \mu_2(f_o)$, so $\mu_1, \mu_2 \in M_o$ as well. Also,

$$\mu(f_1) = \alpha\mu_1(f_1) + (1-\alpha)\mu_2(f_1) = \sup_{\nu \in M_o} \nu(f_1),$$

whence $\mu(f_1) = \mu_1(f_1) = \mu_2(f_2)$, and thus $\mu_1, \mu_2 \in M_1$ as well. We can continue like this, and thus $\mu_1(f_j) = \mu_2(f_j)$ for all j. Since (f_j) is dense, we conclude that $\mu_1 = \mu_2$, which contradicts our assumption that μ_1 and μ_2 are different. ☐

4.4. **Skew products**

Given an ergodic probability measure preserving system (X, μ, T) and a measurable map $c : X \to \mathbb{R}/\mathbb{Z}$, we can form a new probability measure preserving system $(\widehat{X}, \mu \otimes \lambda, \widehat{T})$ by

$$\widehat{X} = X \times \mathbb{R}/\mathbb{Z} \quad \text{and} \quad \widehat{T}(x,t) = (Tx, t + c(x)), \quad \text{for } (x,t) \in \widehat{X},$$

where λ denotes the Lebesgue measure on $\mathbb{R}/\mathbb{Z}$. One readily checks that $\mu \otimes \lambda$ is $\widehat{T}$-invariant. In what follows, we shall assume that X is compact, whence $\widehat{X}$, and that $T : X \to X$ and $c : X \to \mathbb{R}/\mathbb{Z}$ are continuous, whence $\widehat{T}$. Note that for all $n \geqslant 1$,

$$\widehat{T}^n(x,t) = (T^n x, t + c(x) + \ldots + c(T^{n-1}x)), \quad \text{for } (x,t) \in \widehat{X}.$$

Proposition 4.10 (Furstenberg)**.** *If* T *is uniquely ergodic and* $(\widehat{X}, \mu \otimes \lambda, \widehat{T})$ *is ergodic, then* $\widehat{T}$ *is uniquely ergodic.*

Proof. Since $(\widehat{X}, \mu \otimes \lambda, \widehat{T})$ is ergodic, there exists by Lemma 4.4 a $\mu \otimes \lambda$-conull subset $\widehat{X}_{\text{gen}}$ such that

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} f(\widehat{T}^k(x,t)) = \int_X f\, d\mu, \quad \text{for all } (x,t) \in \widehat{X}_{\text{gen}} \text{ and } f \in C(\widehat{X}).$$

We claim that if $(x,t) \in \widehat{X}_{\text{gen}}$, then $(x,s) \in \widehat{X}_{\text{gen}}$ for all $s \in \mathbb{R}/\mathbb{Z}$. Indeed, for all n,

$$\frac{1}{n} \sum_{k=0}^{n-1} f(\widehat{T}^k(x,s)) = \frac{1}{n} \sum_{k=0}^{n-1} f_{s-t}(\widehat{T}^k(x,t)),$$

where $f_u(x,t) = f(x, t+u)$. Since $f_u \in C(\widehat{X})$ and $(x,t) \in \widehat{X}_{\mathrm{gen}}$, we must have

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} f_{s-t}(\widehat{T}^k(x,t)) = \int_{\widehat{X}} f_{s-t}\, d\mu\, d\lambda = \int_{\widehat{X}} f\, d\mu\, d\lambda,$$

since λ is invariant under translations, whence $(x,s) \in \widehat{X}_{\mathrm{gen}}$. We conclude that

$$\widehat{X}_{\mathrm{gen}} = X_\mu \times \mathbb{R}/\mathbb{Z},$$

for some μ-conull Borel set $X_\mu \subset X$. In particular, if we denote by π the projection from $\widehat{X}$ to X, then $\widehat{X}_{\mathrm{gen}} = \pi^{-1}(X_\mu)$ for some μ-conull subset $X_\mu \subset X$.

By Proposition 4.9, it suffices to show that $\mu \otimes \lambda$ is the unique $\widehat{T}$-invariant Borel probability measure on $\widehat{X}$. So, for the sake of argument, let us assume that there exists an ergodic $\widehat{T}$-invariant Borel probability measure ν different from $\mu \otimes \lambda$. Then, by Corollary 4.5, we know that

$$0 = \nu(\widehat{X}_{\mathrm{gen}}) = \nu(\pi^{-1}(X_\mu)) = \pi_*\nu(X_\mu) = \mu(X_\mu) = 1,$$

where the second to last identity follows from unique ergodicity of T, and the fact that $\pi_*\nu$ is a T-invariant probability measure on X, whence equal to μ. $\square$

4.5. **Autocorrelations revisited and a theorem of Weyl**

Recall that a sequence (a_n) of complex numbers is *Wiener* if its autocorrelation

$$\theta_a(n) = \lim_N \frac{1}{N} \sum_{k=0}^{N-1} a_k \overline{a}_{n+k}$$

exists for every $n \geqslant 0$. Let us now consider a sequence of the form $a_n^x = f(T^n x)$, where (X, μ, T) is an ergodic probability measure preserving system and $f : X \to \mathbb{C}$ is μ-integrable (or X is compact and $T : X \to X$ is uniquely ergodic, with the unique T-invariant probability measure μ, and $f : X \to \mathbb{C}$ is continuous). In either of these two cases, we ask whether the limits

$$\lim_N \frac{1}{N} \sum_{k=0}^{N-1} f(T^k x)\overline{f(T^{n+k}x)} = \lim_N \frac{1}{N} \sum_{k=0}^{N-1} g_n(T^k x),$$

where $g_n(x) = f(x)\overline{f(T^n x)}$, exist for all n. By either applying Birkhoff's ergodic theorem, or Lemma 4.6, we can conclude that

$$\theta_{a^x}(n) = \int_X f\, \overline{f \circ T^n}\, d\mu, \quad \text{for } \mu\text{-almost every } x,$$

in the first case, and

$$\theta_{a^x}(n) = \int_X f\, \overline{f \circ T^n}\, d\mu, \quad \text{for all } x \in X,$$

in the second case.

When we introduced autocorrelations in class, we discussed the ("linear phase")-case when $a_n = e^{2\pi i n \alpha}$ for some $\alpha \in [0,1)$. In this case the corresponding spectral measure equals δ_α. We shall now address the (superficially similar) case

$$a_n = e^{2\pi i n^2 \alpha}, \quad \text{with } \alpha \notin \mathbb{Q}. \tag{4.2}$$

The theory that we have developed above will give that the spectral measure in this case is the Lebesgue measure, so radically different from the "linear phase"-setting. This was first observed by Hermann Weyl.

To get the machine we have constructed going, we need to represent the sequence in (4.2) on a "dynamical" form. In what follows, let $X = \mathbb{R}/\mathbb{Z}$ and $Tx = x + \alpha \pmod 1$, where $\alpha \notin \mathbb{Q}$. Set $c : X \to \mathbb{R}/\mathbb{Z}$

$$c(x) = 2x + \alpha \mod 1,$$

and define $\widehat{T}$ and $\widehat{X}$ as above. It is pretty straightforward to check that

$$\widehat{T}^n(0,0) = (n\alpha, n^2\alpha), \quad \text{for } n \geqslant 1,$$

so if we set $f(x,t) = e^{2\pi i t}$, then $f \in C(\widehat{X})$ and

$$f(\widehat{T}^n(0,0)) = e^{2\pi i n^2 \alpha}.$$

More generally,

$$\widehat{T^n}(x,t) = (x + n\alpha, t + d_n x + n^2\alpha), \quad \text{for some } d_n \geqslant 1.$$

Hence, if we can show that $\widehat{T}$ is uniquely ergodic, then it follows from above that

$$\theta_a(n) = \int_{\widehat{X}} f(x,t)\overline{f(x + n\alpha, t + d_n x + n^2\alpha}\, d\mu(x)\, d\lambda(t) = 0,$$

for all $n \geqslant 1$, and $\theta_a(0) = 1$; in other words, the corresponding spectral measure is the Lebesgue measure.

It thus remains to show that $\widehat{T}$ is uniquely ergodic; or (since T is uniquely ergodic) - by Furstenberg's observation above, that $(\widehat{X}, \mu \otimes \lambda, \widehat{T})$ is ergodic. You are encouraged to prove this - it will be useful to write $\widehat{T}$ in a slightly different form. Note that $\widehat{X} = (\mathbb{R}/\mathbb{Z})^2$, and

$$\widehat{T}(x,t) = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x \\ t \end{pmatrix} + \begin{pmatrix} \alpha \\ \alpha \end{pmatrix} =: A \begin{pmatrix} x \\ t \end{pmatrix} + b_\alpha \tag{4.3}$$

Appendix A. Fourier analysis on $\mathbb{T}$

Definition A.1. A function $\gamma : \mathbb{Z} \to \mathbb{C}$ is called *positive definite* if for every $N \geqslant 1$ and for every $n_1, \ldots, n_N \in \mathbb{Z}$ and $c_1, \ldots, c_N \in \mathbb{C}$ we have

$$\sum_{k,l=1}^{N} \phi(n_k - n_l) c_k \bar{c}_l \geqslant 0.$$

If μ is a bounded Borel measure on $\mathbb{T}$, we define its *Fourier transform* $\hat{\mu}$ by

$$\widehat{\mu}(n) = \int_X e^{-2\pi n x}\, d\mu(x), \quad \text{for } n \in \mathbb{Z}.$$

Theorem A.2 (Herglotz's Theorem). *If $\phi : \mathbb{Z} \to \mathbb{C}$ is positive definite, then there exists a unique bounded positive measure θ_ϕ on $\mathbb{T}$ such that $\phi = \widehat{\theta}_\phi$.*

Definition A.3. We refer to the measure θ_ϕ in Theorem A.2 as the *spectral measure* associated to the positive definite function ϕ.

Proof. Since ϕ is positive definite, the function $F_N : \mathbb{T} \to \mathbb{C}$

$$F_N(x) = \frac{1}{N} \sum_{m,n=1}^{N} \phi(m-n) e^{2\pi i (m-n) x},$$

is non-negative for every $N \geqslant 1$, whence the Borel measure $d\theta_N = F_N(x)\,d\lambda(x)$, where λ denotes the Lebesgue probability measure on $\mathbb{T}$, is bounded and positive. Furthermore,

$$\theta_N(\mathbb{T}) = \frac{1}{N}\sum_{m=1}^{N} \phi(0) = \phi(0), \quad \text{for all } N,$$

so by narrow sequential compactness, we can extract a subsequence (N_k) such that $\theta_{N_k} \to \theta$ for some bounded positive Borel measure θ on $\mathbb{T}$ in the narrow topology. In particular, for every $n \in \mathbb{Z}$,

$$\lim_k \int_{\mathbb{T}} e^{-2\pi i n x}\, d\theta_{N_k}(x) = \int_{\mathbb{T}} e^{-2\pi i n x}\, d\theta(x).$$

It is not hard to check that the left hand side always converges to $\phi(n)$, whence $\int_{\mathbb{T}} e^{-2\pi i n x}\, d\theta(x) = \phi(n)$ for all n. Since trigonometric polynomials are dense in $C(\mathbb{T})$, we see that if θ' is any other bounded positive measure on $\mathbb{T}$ such that $\phi(n) = \int_{\mathbb{T}} e^{-2\pi i n x}\, d\theta'(x)$ for all n, then $\theta' = \theta$, whence the notation θ_ϕ makes sense. □

A.1. **Properties of spectral measures**

Lemma A.4 (Wiener's Lemma)**.** *If μ is a bounded and positive Borel measure on $\mathbb{T}$, then*

$$\lim_N \frac{1}{2N+1} \sum_{|n|\leqslant N} |\widehat{\mu}(n)|^2 = \sum_{x\in\mathbb{T}} |\mu(\{x\})|^2.$$

In particular, if $\lim_N \frac{1}{2N+1}\sum_{|n|\leqslant N} |\widehat{\mu}(n)|^2 = 0$, then μ is non-atomic.

Proof. Let $\Delta = \{(x,x) \mid x \in \mathbb{T}\} \subset \mathbb{T}\times\mathbb{T}$ and note that

$$\frac{1}{N}\sum_{|n|\leqslant N} e^{2\pi n(x-y)} \to \chi_\Delta(x,y), \quad \text{for all } (x,y) \in \mathbb{T}\times\mathbb{T},$$

whence

$$\begin{aligned} \frac{1}{N}\sum_{|n|\leqslant N} |\widehat{\mu}(n)|^2 &= \int_{\mathbb{T}}\int_{\mathbb{T}} \frac{1}{N}\sum_{|n|\leqslant N} e^{2\pi n(x-y)}\, d\mu(x)\, d\mu(y) \\ &\to \mu\otimes\mu(\Delta) = \sum_{x\in\mathbb{T}} |\mu(\{x\})|^2. \end{aligned}$$

□

Lemma A.5 (Riemann-Lebesgue's Lemma)**.** *If μ is a bounded Borel measure on $\mathbb{T}$, which is absolutely continuous with respect to the Lebesgue measure λ, then $\widehat{\mu}(n) \to 0$ as $n \to \infty$.*

Proof. Write $\mu = u\,d\lambda$ with $u \in L^1(\lambda)$. If $u \in L^2(\lambda)$, then the assertion is trivial by Parseval's Theorem. Since $L^2(\lambda)$ is dense in $L^1(\lambda)$, we are done. □

Appendix B. Around the theorem of Arzela and Ascoli

Let (X, d_X) and (Y, d_Y) be metric spaces. Denote by $C(X,Y)$ the space of continuous maps from X to Y, equipped with the topology of uniform convergence on compact sets. This topology can be described as follows. Fix $f \in C(X,Y)$ and a compact set $K \subset X$. We say that a sequence (f_n) in $C(X,Y)$ *converge compactly* to f if

$$\lim_n \sup_{x\in K} d_Y(f(x), f_n(x)) = 0, \quad \text{for every compact set } K \subset X.$$

Furthermore, we define the (f,K)-*modulus of continuity* $\omega_{f,K} : [0\infty) \to [0,\infty)$ by

$$\omega_{f,K}(t) = \sup\big\{d_Y(f(x), f(x')) \mid x, x' \in X \text{ and } d_X(x,x') \leqslant t\big\}, \quad \text{for } t \geqslant 0.$$

Note that $\omega_{f,K}(t) \to 0$ as $t \to 0$. More generally, if $A \subset C(X,Y)$, we define the (A,K)-*modulus of continity* $\omega_{A,K}$ by

$$\omega_{A,K}(t) = \sup_{f \in A} \omega_{f,K}(t).$$

We say that A is *totally bounded* if for every $x \in X$, the set

$$Y_x = \big\{f(x) \mid f \in A\big\} \subset Y$$

is sequentially pre-compact, and *equicontinuous* if $\lim_{t\to 0} \omega_{A,K}(t) \to 0$.

Theorem B.1 (Arzela-Ascoli). *Suppose that (X, d_X) is separable and (Y, d_Y) is complete. If $A \subset C(X,Y)$ is totally bounded and equicontinuous, then it is sequentially pre-compact.*

Proof. In class. □

The proofs of the following corollaries are left as exercises (see below).

Corollary B.2. *Let $\beta > 0$, and fix $Q > 0$ and a function $\gamma : (0,\infty) \to (0,\infty)$. Then the set $A_{Q,\gamma} \subset C([0,\infty), \mathbb{R}^N)$ defined by*

$$A_{Q,\gamma} = \Big\{f \in C([0,\infty), \mathbb{R}^N) \mid |f(0)| \leqslant Q \textit{ and for every } T > 0, \sup_{0 \leqslant s < t \leqslant T} \frac{|f(s) - f(t)|}{|s-t|^\beta} \leqslant \gamma(T)\Big\}$$

is sequentially pre-compact.

Corollary B.3. *Suppose that $M \subset \mathcal{P}(C([0,\infty), \mathbb{R}^N))$ satisfies*

$$\lim_{Q\to\infty} \inf_{\mu \in M} \mu\big(\{f \in C([0,\infty), \mathbb{R}^N) \mid |f(0)| \leqslant Q\big\}\big) = 1$$

and, for every $T > 0$,

$$\lim_{R\to\infty} \inf_{\mu \in M} \mu\big(\{f \in C([0,\infty), \mathbb{R}^N) \mid \sup_{0 \leqslant s < t \leqslant T} \frac{|f(s) - f(t)|}{|s-t|^\beta} \leqslant R\big\}\big) = 1.$$

Then M is sequentially pre-compact in in the narrow topology.

B.1. Besov saves the day

Let $\Phi : [0,\infty) \to (0,\infty)$ and $\omega : (0,\infty) \to (0,\infty)$ be strictly increasing functions, and suppose that $\lim_{t\to 0^+} \omega(t) = 0$. Fix $T > 0$ and define for $g \in C([0,T], \mathbb{R}^N)$, the *Besov norm*

$$B_T(g) = \int_0^T \int_0^T \Phi\Big(\frac{|g(s) - g(t)|}{\omega(|s-t|)}\Big)\, ds\, dt.$$

Proposition B.4. *If $B_T(g) < \infty$, then*

$$|g(s) - g(t)| \leqslant 8 \int_0^{t-s} \Phi^{-1}\Big(\frac{4B_T(g)}{u^2}\Big)\, d\omega(u), \quad \textit{for all } 0 \leqslant s < t \leqslant T,$$

where $d\omega$ denotes the Stiltjes measures associated to ω.

The proof of the following corollary is left as an exercise.

Corollary B.5. *Suppose that* $M \subset \mathcal{P}(C([0,\infty),\mathbb{R}^N))$ *satisfies*

$$\lim_{Q\to\infty} \inf_{\mu\in M} \mu\big(\{f \in C([0,\infty),\mathbb{R}^N) \mid |f(0)| \leqslant Q\big\} = 0$$

and, for some $\alpha, r > 0$*, there exist, for every* $T > 0$ *a constant* C_T *such that*

$$\int |f(s) - f(t)|^r \, d\mu(f) \leqslant C_T|s-t|^{1+\alpha}, \quad \textit{for all } 0 \leqslant s < t \leqslant T.$$

Then M *is sequentially pre-compact.*

Let us now turn to the proof of Proposition B.4. We begin by showing that it suffices to consider the case when $T = t = 1$ and $s = 0$. Given $f \in C([0,1],\mathbb{R}^N)$, we set

$$A_f(s) = \int_0^1 \Phi\Big(\frac{|f(s)-f(t)|}{\omega(|s-t|)}\Big)\, dt, \quad \text{for } 0 \leqslant s \leqslant 1,$$

and

$$B_f = \int_0^1 A_f(s)\, ds.$$

Lemma B.6. *For every* $f \in C([0,1],\mathbb{R}^N)$*,*

$$|f(1) - f(0)| \leqslant 8 \int_0^1 \Phi^{-1}\Big(\frac{4B_f}{u^2}\Big)\, d\omega(u).$$

Proof of Proposition B.4 assuming Lemma B.6. Given $g \in C([0,T],\mathbb{R}^N)$ and $0 \leqslant s < t \leqslant T$, define

$$f(\tau) = g(s + \tau(t-s)) \quad \text{and} \quad \overline{\omega}(u) = \omega((t-s)u), \quad \text{for } 0 \leqslant \tau \leqslant 1 \text{ and } u > 0.$$

Apply Lemma B.6 to f and $\overline{\omega}$, and note that $f(1) = g(t)$ and $f(0) = g(s)$. □

B.1.1. *Proof of Lemma B.6*

Fix $f \in C([0,1],\mathbb{R}^N)$ such that $B_f < \infty$. We shall prove the following lemma.

Lemma B.7. *For every* $s_o \in (0,1)$ *such that* $A_f(s_o) \leqslant B_f$*, we have*

$$|f(s_o) - f(0)| \leqslant 4 \int_0^1 \Phi^{-1}\Big(\frac{4B_f}{u^2}\Big)\, d\omega(u).$$

To see how Lemma B.6 follows from this lemma, write $\check{f}(s) = f(1-s)$, and note that $A_{\check{f}}(s) = A_f(1-s)$ and $B_{\check{f}} = B_f$. If $A_f(s_o) \leqslant B_f$, then $A_{\check{f}}(1-s_o) \leqslant B_f$, whence, by the lemma above,

$$|f(1) - f(s_o)| = |\check{f}(1-s_o) - \check{f}(0)| \leqslant 4 \int_0^1 \Phi^{-1}\Big(\frac{4B_f}{u^2}\Big)\, d\omega(u).$$

We conclude that

$$|f(1) - f(0)| \leqslant |f(1) - f(s_o)| + |f(s_o) - f(0)| \leqslant 8 \int_0^1 \Phi^{-1}\Big(\frac{4B_f}{u^2}\Big)\, d\omega(u),$$

which finishes the proof of Lemma B.6.

Let us now turn to the proof of Lemma B.7. We shall need:

Lemma B.8. *For every* $\alpha, \beta \in (0,1)$*, there exists* $u \in (0,\alpha)$ *such that*

$$A_f(u) \leqslant \frac{2B_f}{\alpha} \quad \textit{and} \quad \frac{|f(\beta) - f(u)|}{\omega(|\beta - u|)} \leqslant \Phi^{-1}\Big(\frac{2A_f(\beta)}{\alpha}\Big).$$

Assuming this lemma, let us see how Lemma B.7 follows. Fix $s_o \in (0,1)$ such that $A_f(s_o) \leqslant B_f$, and define $t_o \in (0,1)$ such that $\omega(t_o) = \frac{\omega(s_o)}{2}$. Since $\omega(t_o) < \omega(s_o)$ and ω is strictly increasing, we see that $t_o < s_o$. Apply Lemma B.8 with $\alpha = t_o$ and $\beta = s_o$, to find $0 < s_1 < t_o$ such that

$$A_f(s_1) \leqslant \frac{2B_f}{t_o} \quad \text{and} \quad \frac{|f(s_1) - f(s_o)|}{\omega(|s_1 - s_o|)} \leqslant \Phi^{-1}\Big(\frac{2A_f(s_o)}{t_o}\Big).$$

Now define $0 < t_1 < s_1$ by $\omega(t_1) = \frac{\omega(s_1}{2}$, and use Lemma B.8 with $\alpha = t_1$ and $\beta = s_1$ to find $0 < s_2 < t_1$ such that

$$A_f(s_2) \leqslant \frac{2B_f}{t_1} \quad \text{and} \quad \frac{|f(s_2) - f(s_1)|}{\omega(|s_2 - s_1|)} \leqslant \Phi^{-1}\Big(\frac{2A_f(s_1)}{t_1}\Big).$$

We can continue this construction to find a sequence $s_o > t_o > s_1 > t_1 > s_2 > \dots$, where $\omega(t_n) = \frac{\omega(s_n)}{2}$ such that

$$A_f(s_{n+1}) \leqslant \frac{2B_f}{t_n} \quad \text{and} \quad \frac{|f(s_{n+1}) - f(s_n)|}{\omega(|s_{n+1} - s_n|)} \leqslant \Phi^{-1}\Big(\frac{2A_f(s_n)}{t_n}\Big).$$

Since $\omega(t_n) \leqslant \frac{\omega(s_o)}{2^n}$, we see that $t_n \to 0$, and thus $s_n \to 0$. Furthermore,

$$\omega(|s_n - s_{n+1}|) \leqslant \omega(s_n) = 2\omega(t_n) = 4(\omega(t_n) - \frac{1}{2}\omega(t_n)) \leqslant 4(\omega(t_n) - \omega(t_{n+1}))$$

Since f is continuous, we conclude that

$$\begin{aligned}
|f(s_o) - f(0)| &= \Big|\sum_{n=0}^{\infty}\big(f(s_n) - f(s_{n+1})\big)\Big| \\
&\leqslant \sum_{n=0}^{\infty}\Big(\frac{|f(s_n) - f(s_{n+1})|}{\omega(|s_n - s_{n+1}|)}\Big)\,\omega(|s_n - s_{n+1}|) \\
&\leqslant 4\sum_{n=0}^{\infty}\Phi^{-1}\Big(\frac{2A_f(s_n)}{t_n}\Big)\,(\omega(t_n) - \omega(t_{n+1})) \\
&\leqslant 4\sum_{n=0}^{\infty}\Phi^{-1}\Big(\frac{4B_f}{t_n t_{n-1}}\Big)\,(\omega(t_n) - \omega(t_{n+1})) \\
&\leqslant 4\sum_{n=0}^{\infty}\Phi^{-1}\Big(\frac{4B_f}{t_n^2}\Big)\,(\omega(t_n) - \omega(t_{n+1})) \\
&\leqslant 4\sum_{n=0}^{\infty}\int_{t_{n+1}}^{t_n}\Phi^{-1}\Big(\frac{4B_f}{u^2}\Big)\,d\omega(u) \\
&\leqslant 4\int_0^1\Phi^{-1}\Big(\frac{4B_f}{u^2}\Big)\,d\omega(u),
\end{aligned}$$

which finishes the proof.

It remains to prove Lemma B.8. Set

$$I = \{u \in [0,1] \mid A_f(u) \leqslant \frac{2B_f}{\alpha}\}$$

and

$$J = \{u \in [0,1] \mid \frac{|f(\beta) - f(u)|}{\omega(|\beta - u|)} \leqslant \Phi^{-1}\Big(\frac{2A_f(\beta)}{\alpha}\Big)\}.$$

By Markov's inequality,

$$\lambda(I) > 1 - \frac{\alpha}{2} \quad \text{and} \quad \lambda(J) > 1 - \frac{\alpha}{2},$$

where λ denotes the Lebesgue measure on $[0,1]$, whence

$$\lambda(I \cap J) > 1 - \alpha,$$

and thus $I \cap J$ intersects the interval $(0, \alpha)$, which finishes the proof.

Exercises

Exercise 1 (3 points)**.** Let (X, d) be a metric space and let μ and ν be bounded positive Borel measures on X. Fix a bounded positive Borel measure λ on X such that $\mu \ll \lambda$ and $\nu \ll \lambda$, and write $d\mu = u\, d\lambda$ and $d\nu = v\, d\lambda$. Define the *Hellinger affinity* $H(\mu, \nu)$ by

$$H(\mu, \nu) = \int_X \sqrt{uv}\, d\lambda.$$

Show that $H(\mu, \nu)$ is independent of the choice of λ, and that

$$\mu \perp \nu \implies H(\mu, \nu) = 0.$$

Exercise 2 (3 points)**.** Suppose that (a_n) is a Wiener sequence and $\gamma_a : \mathbb{N}_o \to \mathbb{C}$ its auto-correlation. Show that

$$\gamma_a(n) := \begin{cases} \gamma_a(n) & \text{if } n \geqslant 0 \\ \overline{\gamma_a(-n)} & \text{if } n < 0 \end{cases}$$

is a positive definite function on $\mathbb{Z}$, and that the sequence of probability measures $(\theta_N^{(a)})$ on $\mathbb{T}$ defined by

$$\int_{\mathbb{T}} f\, d\theta_N^{(a)} = \frac{1}{N} \int_{\mathbb{T}} f(x) \Big| \sum_{n=0}^{N-1} a_n e^{-2\pi i n x} \Big|^2 dx, \quad \text{for } f \in C(\mathbb{T}), \tag{B.1}$$

converges narrowly to the spectral measure θ_{γ_a} associated to γ_a.

Exercise 3 (5 points)**.** Let (X, d) be a metric space and let $B_0, \ldots, B_N$ be a partition of X into Borel sets. Let μ and ν be positive and bounded tight Borel measures on X. Construct a function

$$\beta : [0,1) \to [0,1), \text{ with } \lim_{t \to 0} \beta(t) = 0,$$

such that for all small enough $\varepsilon > 0$, there are continuous functions $f_0, f_1, \ldots, f_N : X \to [0,1]$ such that

$$\sum_{k=0}^{N} f_k(x) = 1, \quad \text{for all } x \in X,$$

with the following properties:

(i) If $\mu(B_k) = 0$, then $\int_X f_k\, d\mu < \varepsilon$.
(ii) If $\nu(B_k) = 0$, then $\int_X f_k\, d\nu < \varepsilon$.
(iii) If $\mu(B_k)\nu(B_k) > 0$, then

$$\beta(\varepsilon) \leqslant \int_X f_k\, d\mu \leqslant (1+\varepsilon)^{\frac{1}{2}} \mu(B_k) + \beta(\varepsilon),$$

and

$$\beta(\varepsilon) \leqslant \int_X f_k\, d\nu \leqslant (1+\varepsilon)^{\frac{1}{2}} \nu(B_k) + \beta(\varepsilon).$$

Exercise 4 (10 points)**.** Given a non-negative integer n, let S_n denote the sum of the digits in the binary expansion of n. For instance,

$$S_0 = 0, \;\; S_1 = 1, \;\; S_2 = 1, \;\; S_3 = 2, \;\; S_4 = 1, \;\; S_5 = 2 \ldots.$$

The *Thue-Morse sequence* is defined as $a_n = (-1)^{S_n}$.

a) Show that $S_{2n} = S_n$ and $S_{2n+1} = S_n + 1$ for all n, whence

$$a_{2n} = a_n \quad \text{and} \quad a_{2n+1} = -a_n, \quad \text{for all } n. \tag{B.2}$$

b) Show that the autocorrelation $\gamma : \mathbb{N}_0 \to \mathbb{R}$,

$$\gamma(m) = \lim_N \frac{1}{N+1} \sum_{n=0}^{N} a_n a_{n+m} \quad \text{exists for all } m \geqslant 0,$$

and satisfies

$$\gamma(2m) = \gamma(m) \quad \text{and} \quad \gamma(2m+1) = -\frac{1}{2}\big(\gamma(m) + \gamma(m+1)\big), \quad \text{for all } m \geqslant 0. \tag{B.3}$$

c) Let θ_γ denote the spectral measure associated to γ. Show that θ_γ is both non-atomic *and* singular with respect to the Lebesgue measure on $\mathbb{T}$.

d) Let $\xi \in \mathbb{C}$ with $|\xi| = 1$ and let (ε_k) be a sequence of real-valued i.i.d random variables with zero means and finite variances. Show that

$$\lim_{N\to\infty} \frac{1}{N+1} \sum_{n=0}^{N} a_n \xi^n = 0$$

and

$$\lim_{N\to\infty} \frac{1}{N+1} \sum_{n=0}^{N} a_n \varepsilon_n = 0, \quad \text{almost surely.}$$

Exercise 5 (3 points)**.** Let μ be a Borel probability measure on on a metric space (X, d). Show that for every $x \in X$ there exists a countable set $S_x \subset (0, \infty)$ such that $B_r(x)$ is μ-Jordan measurable for all $r \notin S_x$, where $B_r(x)$ denotes the closed ball around x of radius r.

Exercise 6 (4 points)**.** Let (X, d) be a metric space. Recall that the *variation metric* d_{Var} between two Borel probability measure μ and ν on X is defined by

$$d_{\mathrm{Var}}(\mu, \nu) = \sup\big\{\big|\mu(B) - \nu(B)\big| \mid B \subset X \text{ is Borel}\big\}.$$

Is $(\mu, \nu) \mapsto d_{\mathrm{Var}}(\mu, \nu)$ sequentially lower semi-continuous with respect to narrow convergence?

Exercise 7 (5 points)**.** Let (X_1, d_1) and (X_2, d_2) be separable metric spaces, and let (X, d) denote the metric space $(X_1 \times X_2, d_1 + d_2)$. Given Borel probability measures μ_1 and μ_2 on X_1 and X_2 respectively, a *coupling* μ of μ_1 and μ_2 is a Borel probability measure on X such that

$$\mu_1(B_1) = \mu(B_1 \times X_2) \quad \text{and} \quad \mu_2(B_2) = \mu(X_1 \times B_2)$$

for all Borel sets $B_1 \subset X_1$ and $B_1 \subset X_2$.

Let $\mathcal{C}(\mu_1, \mu_2)$ denote the set of all couplings of μ_1 and μ_2. Show that $\mathcal{C}(\mu_1, \mu_2)$ is sequentially compact in the narrow topology, and deduce the following fundamental principle in the field of optimal transport: If $c : X_1 \times X_2 \to \mathbb{R}$ is lower semicontinuous and bounded from below, then the map

$$\mu \mapsto \int_{X_1}\int_{X_2} c(x_1, x_2)\, d\mu(x_1, x_2)$$

attains a minimum in $\mathcal{C}(\mu_1, \mu_2)$.

Exercise 8 (5 points)**.** Let Y and Z be metrizable spaces with Y σ-compact (meaning that Y can be exhausted by an increasing union of countably many compact sets). Suppose that $\varphi : Y \to Z$ is an injective and continuous map. Show that:

(i) $\varphi(Y) \subset Z$ is σ-compact, hence Borel measurable.
(ii) $\mathcal{A} := \{B \subset Y \mid \varphi(B) \text{ is Borel}\}$ is a σ-algebra.
(iii) Every closed subset of Y is contained in $\mathcal{A}$.

Deduce from (i)-(iii) that the φ-image of every Borel set in Y is Borel measurable in Z, and show that if ν is a Borel probability measure on Z, then $\mu(B) = \nu(\varphi(B))$, for $B \subset Y$ Borel, is a Borel probability measure on Y.

Remark B.9. Henri Lebesgue, in his original expose of integration theory, claimed that images of Borel sets under continuous maps are always Borel measurable. Later, Mikhail Suslin showed that this is not true without the hypothesis that the map is injective.

Exercise 9 (3 points)**.** Let m denote the Lebesgue probability measure on $[0,1]$. Define the sequence (ν_n) of Borel probability measures on $[0,1]$ by

$$\nu_n(f) = \frac{1}{n+1}\sum_{k=0}^{n} f\Big(\frac{k}{n}\Big), \quad \text{for } f \in C([0,1]).$$

a) Prove the fundamental theorem in Riemann integration, namely that ν_n narrowly converges to m.
b) For every $\varepsilon > 0$, there exists an open set $U \subset [0,1]$ such that $m(U) < \varepsilon$ and $\nu_n(U) = 1$ for all n. In particular, $\nu_n(U) \not\to m(U)$.

Exercise 10 (15 points)**.** Prove Corollaries B.2, B.3 and B.5.

Exercise 11 (5 points)**.** Let $\varepsilon_1, \varepsilon_2, \ldots$ be centered i.i.d. random variables with finite fourth moments on some probability measure space $(\Omega, \mathbb{P})$, and set $S_n = \varepsilon_1 + \ldots + \varepsilon_n$ (with the convention $S_o = 0$). Define the sequence $\pi_n : \Omega \to C([0,1], \mathbb{R})$ by

$$\pi_n(\cdot, t) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(S_{i-1}(\cdot) + n\Big(t - \frac{i-1}{n}\Big)\varepsilon_i(\cdot)\Big)\chi_{\left(\frac{i-1}{n}, \frac{i}{n}\right]}(t), \quad \text{for } t \in [0,1],$$

and set $\mu_n = (\pi_n)_*\mathbb{P}$ (the push-forward of $\mathbb{P}$ to a probability measure on $C([0,1], \mathbb{R})$. Show that (μ_n) is tight.

Exercise 12 (3 points)**.** Let μ be a Borel probability measure on $\mathbb{R}$. Show that for every positive integer N, we have

$$\big|\widehat{\mu}(\xi) - 1\big| \leqslant N\sqrt{2|\widehat{\mu}(\xi/N) - 1|}, \quad \text{for all } \xi.$$

Exercise 13 (7 points)**.** Let $T : X \to X$ be a measurable map which preserves a Borel probability measure μ. Fix $p \geqslant 1$. Show that the map $U_T : L^p(X,\mu) \to L^p(X,\mu)$, $f \mapsto f \circ T$ is well-defined, and satisfies $\|U_T f\|_p = \|f\|_p$ for all $f \in L^p(X,\mu)$. Also show that (X,μ,T) is ergodic if and only if there is no non-constant $f \in L^p(X,\mu)$ such that $U_T f = f$, if and only if there is no non-constant solution to $U_T^* g = g$, where $U_T^* : L^q(X,\mu) \to L^q(X,\mu)$ denotes the transpose map (where $\frac{1}{p} + \frac{1}{q} = 1$) [for the last equivalence, you can use the next exercise].

Exercise 14 (3 points)**.** Let (X, T, μ) be an ergodic probability measure preserving system. Show that if $f \in L^p(X,\mu)$ for some $1 < p < \infty$,

$$\frac{1}{N}\sum_{n=0}^{N-1} f \circ T^n \to \int_X f \, d\mu, \quad \text{in the weak-topology.}$$

Exercise 15 (4 points)**.** Show that the map $T : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ preserves the Lebesgue measure μ on $\mathbb{R}/\mathbb{Z}$. *Possible approach: Show first that* $\mu(T^{-1}(B)) = \mu(B)$ *for every* interval $B \subset \mathbb{T}$, *then use magic from your integration theory course.*

Exercise 16 (5 points)**.** Prove Rajchman's Lemma: Let (f_k) be a sequence of bounded measurable functions on a probability measure space (X, μ) with

$$\sup_k \|f_k\|_\infty < \infty \quad \text{and} \quad \langle f_j, f_k \rangle_{L^2(X,\mu)} = 0, \quad \text{for all } j, k.$$

Then $\frac{1}{n} \sum_{k=1}^{n} f_k(x) \to 0$ μ-almost everywhere. *Hint: Show that*

$$\sum_{n \geqslant 1} \Big\| \frac{1}{n^2} \sum_{k=1}^{n^2} f_k \Big\|_2^2 < \infty,$$

use Borel-Cantelli's Lemma, and use a "sandwich"-argument.

Exercise 17 (5 points)**.** A real number $x \in [0, 1)$ is called *2-normal* if the limits

$$d_i(x) := \lim_N \frac{|\{n \in [1, N] \mid x_n = i\}|}{N}, \quad \text{i=0,1,}$$

where x_n denotes the n'th digit in the binary expansion of x exists and equals $1/2$ for $i = 1, 2$. Show that Lebesgue almost every number in $[0, 1)$ is 2-normal. *Hint: Show that* $x_n = i$ *depending on whether* $2^n x$ *mod* 1 *ends up in* $[0, 1/2)$ *or* $[1/2, 1)$, *and use the ergodicity of the map* $x \mapsto 2x$ *(mod* 1*), together with Birkhoff's Ergodic Theorem.*

Exercise 18 (3 points)**.** Show that if X is compact and metrizable, and $T : X \to X$ is continuous, then the set $\mathcal{P}_T(X)$ of T-invariant Borel probability measures on X is weak*-compact.

Exercise 19 (4 points)**.** Show that the system $(\widehat{X}, \mu \otimes \lambda, \widehat{T})$ defined in (4.3) is ergodic.

Exercise 20 (4 points)**.** Show that if X is a compact metrizable space, then $C(X)$ is separable in the uniform norm (Is the same true if the assumption that X is compact is dropped?).

DEPARTMENT OF MATHEMATICS, CHALMERS, GOTHENBURG, SWEDEN
Email address: micbjo@chalmers.se