Stationarity and ergodicity

A sequence \((X_n)_{n \geq 0}\) of random variables taking values in some measurable space is called stationary if, for every \(k \geq 1\),

\[(X_0, X_1, X_2, \ldots) \overset{d}{=} (X_{k}, X_{k+1}, X_{k+2}, \ldots)\]

(this is the same as requiring that for every \(n \geq 0\) and \(k \geq 1\),

\[(X_0, \ldots, X_n) \overset{d}{=} (X_{k}, \ldots, X_{k+n})\]

Examples: 1) \((X_n)\) are IID.

2) Markov chain in the stationary distribution.

To avoid technicalities, focus on the finite-state space case. \(S\)-state space, \(P\) is the transition probabilities matrix.

Let \(\pi\) be a stationary distribution.

Take \(X \sim \pi\), \((X_0, X_1, \ldots)\) be a sample of the Markov chain.

\[P(X_0 = x_0, \ldots, X_n = x_n) = \pi(x_0)P(x_0, x_1)P(x_1, x_2) \cdots P(x_{n-1}, x_n)\]

Since \(\pi\) is stationary, the same holds for \((X_k, \ldots, X_{k+n})\).

3) If \((X_n)\) is stationary, taking values in some Borel\(s\), then for any measurable function \(f: S^{\mathbb{N}} \rightarrow S'\), it holds that \((X'_0, X'_1, \ldots)\) is stationary, where \(X'_n := f(X_n, X_{n+1}, \ldots)\).

Measure-preserving maps: Suppose \((\mathcal{B}, \mathbb{P})\)
Measure-preserving maps: Suppose $(\Omega, \mathcal{F}, P)$ is a probability space. Suppose $\varphi: \Omega \rightarrow \Omega$ is measure preserving if $\varphi$ is measurable and $\forall E \in \mathcal{F}$, $P(\varphi^{-1}E) = P(E)$.

It is equivalent that for every random variable $X$ on $(\Omega, \mathcal{F}, P)$, $X \overset{d}{=} X'$

where $X'(w) = X(\varphi(w))$.

Now, given a rv $X$, can define

$X_n(w) := X(\varphi^n(w))$, $n \geq 0$

(so that $X_0(w) = X(w)$)

and then $(X_0, X_1)$ is stationary.

Proof: Suppose $X$ takes values in $S$.

Let $n \geq 0$, and $A$ a measurable set in $S$.

Then, $\forall k \geq 1$,

$P(\{X_k, X_{k+1}, \ldots, X_{k+n} \} \in A) =$

$= P(\{X(\varphi^k(w)), X(\varphi^{k+n}(w)) \} \in A) =$

$= P(\{X(w), X(\varphi^n(w)) \} \in \varphi^{-k}A) =$

$= P(\varphi^{-k}A) = P(A)$

since $\varphi$ is measure preserving.

Example: (rotation of the circle)

Let $\Lambda = [0, 1)$, $\mathcal{F}$ - Borel sets, $\varrho$ - Lebesgue measure.

Fix $\theta \in (0, 1)$ and let $\varphi: [0, 1) \rightarrow [0, 1)$

be $\varphi(w) = (w + \theta) \mod 1$

Fractional part of $w + \theta$. 

\[ \theta \varphi(w) \]
It is simple that \( \Psi \) is measure preserving. Define \( X(w) = w \). \( X_n(w) := w + n \mod 1 \). So \( (X_0, X_1, \ldots) \) is stationary.

**Every stationary sequence can be realized by a measurable map.**

Let \( (Y_n) \) be stationary, taking values in a Borel space \( S \). We will define \((\mathcal{A}, \mathcal{F}, P)\) and \( \Psi, \chi \) so that \( X_n = X_0 \psi^n \) satisfies that \( (Y_0, \ldots) \overset{d}{=} (X_0, \chi(X_1), \ldots) \).

Indeed, let \( \mathcal{A} = S \times 0, 1, \ldots \) and let \( P \) be the distribution of \( (Y_0, Y_1, \ldots) \).

Let \( X(w_0, w_1, \ldots) = w_0 \).

\[ \Psi(w_0, w_1, \ldots) = (w_1, w_2, \ldots) \text{ the left shift.} \]

So that \( X_n(w) = w_n \).

\( \Psi \) is measure preserving since, under \( P \),

\[ (w_0, w_1, \ldots) \overset{d}{=} (Y_0, Y_1, \ldots) \overset{d}{=} \]

\[ (Y_1, Y_2, \ldots) \overset{d}{=} (w_1, w_2, \ldots). \]

It also follows that

\[ (X_0, \ldots, X_n) \overset{d}{=} (Y_0, \ldots, Y_n). \]

**Invariant Events and Ergodicity**

From now on, fix \((\mathcal{A}, \mathcal{F}, P)\) a prob. space, \( \Psi: \mathcal{A} \to \mathcal{A} \) measure-pres. and \( X \) a RV.

Then let \( X_n = X_0 \psi^n \). \((\mathcal{A}, \mathcal{F}, P, \Psi)-\text{Dynamical System}\)

An event \( E \in \mathcal{F} \) is called **invariant**

\[ \psi(E) = E \]
An event \( I \in \mathcal{F} \) is called in\textit{variant} if \( \gamma^{-1}I = I \).

An event \( I \in \mathcal{F} \) is called \textit{almost-invariant} if \( \mathbb{P}(\gamma^{-1}I \Delta I) = 0 \). 

\[ A \Delta B = (A \setminus B) \cup (B \setminus A) \]

\[ \text{Exercise: } 1) \text{ The invariant events form a sigma algebra } \mathcal{F}. \text{ The same for the almost-invariant events.} 

2) A RV \( Y \) is measurable with respect to \( \mathcal{F} \) iff \( Y(w) = Y(Y(w)) \) \( \forall \) \( w \in \Omega \).

Similarly, it is measurable wrt. the almost-invariant \( \sigma \)-algebra if \( \mathbb{P}(Y(w) \neq Y(Y(w))) = 0 \).

3) If \( I \) is an almost-invariant event then there exists an invariant \( I' \) such that \( \mathbb{P}(I \Delta I') = 0 \), or \( (I, \gamma, \mathbb{P}) \) and \( \gamma \) is \textit{ergodic}.

The sequence \((X_n)\) is called \textit{ergodic} if \( \mathbb{P}(I) \in \{0, 1\} \) for all \( I \in \mathcal{F} \).

\[ \text{Examples: } 1) (X_n) \text{ are } i.i.d. \text{ with product measure.} \]

Realize \((X_n)\) on sequence space \((\omega_0, \omega, \sigma)\) with the shift \( \gamma \). Then this construction is \textit{ergodic} (i.e., random variables which are functions of the \((X_n)\) and are invariant to shifts or the \((X_n)\) must be constant almost surely).

Indeed, suppose \( f(X_0, X_1, \ldots) \) is invariant, i.e., \( f(X_0, X_1, \ldots) = f(X_1, X_2, \ldots) \).

It follows that \( f(X_0, \ldots) = f(X_n, \ldots) \), \( \forall n \geq 0 \).

Defining \( C_n := \sigma(X_n, X_{n+1}, \ldots) \) we see that \( f(X_0, \ldots) \in C_n \) for all \( n \geq 0 \) and
that \( \mathbb{P}(X_0 = -1) \in G_n \) for all \( n \geq 0 \) and hence measurable with the tail \( \sigma \)-algebra \( \mathcal{T}_n' = \mathcal{G}_n' \). Kolmogorov's 0-1 law then implies that \( \mathbb{P} \) is constant a.s.

2) **Markov Chain at Stationarity**

- **Finite State Space**, \( S \)
- **Transition Prob. Matrix**, \( P \)
- **Stationary Dist.**, \( \pi \)
- \( (X_0, X_1, \ldots) \) sample of the Markov chain.

Realize on seg. space with \( \mathcal{T} = \text{shift} \), without assuming irreducibility, this might not be ergodic. E.g., \( S = \{0, 1\}, P(0, 0) = 1 \) and \( \pi = (\frac{1}{2}, \frac{1}{2}) \). Then \( \mathbb{P}(X_0 = 0, X_1 = 0) > 0 \) is invariant but \( \mathbb{P}(F = 1) = \mathbb{P}(F = 0) = \frac{1}{2} \).

Assume \( P \) is irreducibility and let us show that \( (X_n) \) is ergodic.

Suppose \( A \) is an invariant event, i.e.,

\[
1_A(X_0, X_1, \ldots) = 1_A(X_1, X_2, \ldots)
\]

We need to show \( \mathbb{P}(A) \in \{0, 1\} \).

Define \( h(X) = \mathbb{P}_X(A) \). Start the Markov chain at \( \mathbf{x} \).

By the Markov property, invariance

\[
\mathbb{P}(A \mid \mathcal{F}_n) = \mathbb{E}(1_A(X_0, \ldots, X_n) \mid \mathcal{F}_n) = \mathbb{E}(1_A(X_n, X_{n+1}, \ldots) \mid \mathcal{F}_n) = \mathbb{P}_X(A) \quad \text{(Markov prop.)}
\]

\[
\mathcal{F}_n = \sigma(X_0, \ldots, X_n)
\]
\( X_n = o(X_0, X_n) \) = \( \mathbb{E}(X_n | X_{n+1}, ..., X_0) \cdot X_n = \mathbb{E}(X_n | X_{n+1}) = h(X_n) \).

By Lévy's 0-1 law (upward martingale conv. thm.)

\[ P(A | X_n) \xrightarrow{n \to \infty} 1_A, \text{ almost surely.} \]

Since \( S \) is finite and the chain is irreducible, \( \forall x \in S \), \( P(X_n = x \text{ infinitely often}) = 1 \).

It follows from the fact that \( h(X_n) \) converges a.s. that \( h \) is constant. Moreover, the limit \( 1_A \) must also be constant, almost surely. So \( P(A) \in \{0, 1\} \).

3) Rotation of the circle is ergodic when \( \theta \) (the rotation amount) is irrational – exercise.

**Remark:** Suppose \( P \) is the transition prob. matrix of an irreducible Markov chain.

As we saw, the invariant \( \sigma \)-alg. is trivial.

What is the tail \( \sigma \)-alg.?

\( \tau \) if for every \( n \) there exists \( F_n \) such that \( 1_{\tau} (X_0, X_1, ..., X_n) = F_n (X_n, X_{n+1}, ...) \).

For instance, if \( S = \{0, 1, 2\}, P = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \)

Then \( A = 1 \{X_0 = 0 \} \in \tau \). \( \pi = \begin{pmatrix} 1/3 & 1/3 & 1/3 \end{pmatrix} \)

\( P(A) = 1/2 \).

More generally, it turns out that \( \tau = \sigma \{ \tau \text{-periodic classes} \} \) (up to measure 0).
Note also that in the example, 
\((X_0, X_1, \ldots)\) is ergodic
but \((X_0, X_2, \ldots)\) is not ergodic.
(So \(\varphi = \text{shift is ergodic but } \varphi^2 \text{ is not ergodic} \).)

**Birkhoff Ergodic Theorem (1937)**

Still work in a measure-pres. setup. 
\((\Omega, \mathcal{F}, \mathbb{P})\) - prob. space
\(\varphi: \Omega \to \Omega\) - meas. preserving.

**Theorem:** For each RV \(X: \Omega \to \mathbb{R}\) with \(\mathbb{E}|X| < \infty\),
\[
\frac{1}{n} S_n(w) := \frac{1}{n} \sum_{k=0}^{n-1} X(\varphi^k w) \to \mathbb{E}(X | I) \quad \text{almost surely} \quad \text{and in } L^1.
\]

**Example:** In the IID setup,
\[X_n = X(y^n w)\] are IID.

\[
\frac{1}{n} \sum_{k=0}^{n-1} X_k \xrightarrow{\text{a.s.}} \mathbb{E}(X | I) = \mathbb{E}(X) = \mathbb{E}(X_0)
\]
is the strong law of large numbers (and \(\mathcal{L}^1\) conv.)
(also the \(L^1\) conv.)

2) Taking \(X = 1_A\) in the ergodic theorem,
we get that in ergodic systems,
the "fraction of time" spent in \(A\) equals \(\mathbb{P}(A) = \text{"part of space that } \varphi \text{ takes"} \).

Physics: Space average = time average.

Start the proof with a lemma due to
Start the proof with a lemma due to Yosida-Kakutani (1939); we show a proof of Garsia (1965).

**Lemma (Maximal Ergodic Lemma):**

Define \( X_n(w) = X(Y^n w), \quad (X_0 = X) \)
\[ S_n(w) = X_0(w) + \cdots + X_{n-1}(w) , \]
\[ M_n(w) = \max(0, S_n(w), \ldots, S_{n-1}(w)). \]

Then \( \mathbb{E}(X I \{ M_k > 0 \}) \geq 0. \)

**Proof:** The idea is to relate \( M_k(w) \) and \( M_k(Y(w)) \).

Observe that for all \( 1 \leq i \leq k \),
\[ M_k(Y(w)) \geq S_i(Y(w)). \]

Thus \( X(w) + M_k(Y(w)) \geq X(w) + S_i(Y(w)) = S_i+1(w) \).

Rearranging, \( X(w) \geq S_{i+1}(w) - M_k(Y(w)) \) for \( 1 \leq i \leq k \).

This holds trivially for \( i = 0 \) since \( S_1 = X \) and \( M_k(Y(w)) \geq 0 \).

Therefore
\[ \mathbb{E}(X I \{ M_k > 0 \}) = \int X(w) \, d\mathbb{P}(w) \geq \left\{ \begin{array}{l} \mathbb{E}(M_k > 0) \\ \mathbb{E}(M_k = 0) \end{array} \right\} \]
\[ \geq \int \left[ \max \{ S_1(w), \ldots, S_k(w) \}^3 - M_k(Y(w)) \right] \, d\mathbb{P}(w) = \]
\[ \int M_k(w) - M_k(Y(w)) \, d\mathbb{P}(w) \geq \left\{ \begin{array}{l} \mathbb{E}(M_k > 0) \\ \mathbb{E}(M_k = 0) \end{array} \right\} \]

On \( \{ M_k > 0 \} \cap \{ M_k = 0 \} \), we have \( M_k = 0, M_k = 0, Y \geq 0 \).
$\sum_{k=1}^{\infty} (M_k^*(w) - M_{k+1}^*(w)) \Delta P(w) = \sum_{k=1}^{\infty} (M_k^* - M_{k+1}^*) \Delta P = 0$

since $\Delta P$ is measure-preserving.

**Proof of the ergodic theorem:**

Observe that $1\mathbb{E}(X|\mathcal{I})$ is an invariant RV, i.e., $1\mathbb{E}(X|\mathcal{I})(w) = 1\mathbb{E}(X|\mathcal{I})(\mathcal{I}w)$.

Replace $X$ by $X - 1\mathbb{E}(X|\mathcal{I})$ and then we may assume without loss of generality that $1\mathbb{E}(X|\mathcal{I}) = 0$ a.s.

Let $\bar{X} = \limsup_{n \to \infty} \frac{1}{n} S_n$, let $\varepsilon > 0$ and

\[ D = \{ \bar{X} > \varepsilon \} \leftarrow \bar{X} \text{ is invariant and hence } \mathbb{P}(D) = 0. \]

**Goal:** $\mathbb{P}(D) = 0$. Then, as $\varepsilon$ is arbitrary, we get $\bar{X} \leq 0$ a.s., applying the same reasoning to $-\bar{X}$ gives $\liminf_{n \to \infty} S_n \geq 0$ a.s.

and then $\lim_{n \to \infty} \frac{S_n}{n} = 0$ a.s., proving the a.s. convergence.

Define a new seq. of RVs,

\[ X^*_n(w) = (X(w) - \varepsilon) \mathbb{1}_D(w) \]

\[ X^*_n(w) = (X^*_n(w) - \varepsilon) \mathbb{1}_D(w) \]

\[ S^*_n(w) = X^*_n(w) + X^*_n(w) + \cdots + X^*_n(w) \]

\[ M^*_n(w) = \max \{ 0, S^*_n(w), \ldots, S^*_n(w) \} \]

\[ F^*_n(w) = \{ M^*_n > 0 \} \quad \text{an increasing sequence of events} \]

Consider $F = \bigcup_{n} F_n = \{ \sup_{k \geq 1} S^*_k(w) > \varepsilon \}

Notice that $\left( \sup_{k \geq 1} S^*_k(w) > \varepsilon \right) = 0 \lor \left( \sup_{k \geq 1} S_k(w) > \varepsilon \right) \lor \left( \sup_{k \geq 1} S_k(w) > \varepsilon \right)$.
Notice that \( \{ \sup_{k \geq 1} \frac{S^*_k(w)}{k} > 0 \} = \emptyset \cap \{ \sup_{k \geq 1} \frac{S_k(w)}{k} > \varepsilon \} \).

Since \( \frac{S^*_k(w)}{k} = 1 \_D \left( \frac{S_k(w)}{k} - \varepsilon \right) \).

Thus \( F = \emptyset \cap \{ \sup_{k \geq 1} \frac{S_k(w)}{k} > \varepsilon \} = \emptyset \).

The maximal ergodic lemma says that

\[ IE(X^*_IF) \geq 0 \quad \forall n \geq 1. \]

This implies that \( IE(X^*_IF) \geq 0 \).

Since \( F_n \rightarrow F \) (since \( F_n \) increases)

and \( |X^*_k| \leq |X| + \varepsilon \) so we can use dominated convergence.

Thus \( 0 \leq IE(X^*_IF) = IE(X^*_IF) = IE((X-\varepsilon)1_F) = IE((X-\varepsilon)1_D) = IE(X1_D) - \varepsilon \Phi(\phi) = -\varepsilon \Phi(\phi) \)

\[ = IE(IE(X1_D | X)) = IE(I_D | IE(X | X)) = 0 \quad \forall \varepsilon \in \mathbb{R} \]

We conclude that \( \Phi(\phi) = 0 \), finishing the proof of the a.s. conv.

It remains to prove conv. in \( L_1 \).

The idea is to truncate, and this is relatively routine.

For \( M > 0 \),

write \( X = X1_{|X| \leq M} + X1_{|X| > M} \)

\( = \bar{X} + \bar{X} \)

Define \( (X'_n) \) and \( (\bar{X}'_n) \) using these.

By the a.s. conv.

\[ 1_{\bar{X}'_n} \rightarrow IE(X'1_X) \quad \text{a.s.} \]
\[
\frac{1}{n} S_n' \rightarrow \mathbb{E}(X'/X) \quad \text{a.s.}
\]

bounded by \( M \).

By the bdd. conv. theorem,
\[
\frac{1}{n} S_n' \rightarrow \mathbb{E}(X'/X) \quad \text{in } L_0.
\]

Next, note that
\[
\left| \mathbb{E} \left( \frac{1}{n} S_n'' \right) \right| = \left| \mathbb{E} \left( \sum_{k=0}^{n-1} x''(y^k w) \right) \right| \leq \sum_{k=0}^{n-1} \frac{1}{n} \mathbb{E} |x''(y^k w)| = \mathbb{E} |x''|.
\]

\( y \) is meas. pres.

Additionally,
\[
\mathbb{E} / \mathbb{E}(X''/X) / \leq \mathbb{E}(\mathbb{E}(x''(1/X)) = \mathbb{E}(1/X').
\]

Putting the last two eqs. together,
\[
\left| \mathbb{E} \left( \frac{1}{n} S_n'' - \mathbb{E}(X''/X) \right) \right| \leq 2 \mathbb{E} |x''|.
\]

It remains to note that \( \mathbb{E}(x'') \rightarrow 0 \)

by the dominated conv. theorem.

Thus,
\[
\left| \mathbb{E} \left( \frac{1}{n} S_n' - \mathbb{E}(X/X) \right) \right| \rightarrow 0.
\]