Stationary Sequences and The Ergodic Theorem
March 14, 2001

Measure Preserving Transformations. Let \((\Omega, \mathcal{A}, P)\) denote a probability space. Then a measurable transformation \(T : \Omega \rightarrow \Omega\) is said to preserve \(P\) iff

\[P \circ T^{-1} = P;\]

that is, iff \(P[T^{-1}(A)] = P(A)\) for all \(A \in \mathcal{A}\).

**Example 1: Rotations of the Circle \(\mathbb{R}\).** If \(\Omega = (0, 1]\), \(P\) is the restriction of Lebesgue measure to \(\Omega\), and \(0 < \alpha < 1\), then the transformation \(T(\omega) = \langle \alpha + \omega \rangle\), the fractional part of \(\omega + \alpha\), preserves \(P\). For if \(0 < c < 1\), then

\[P \circ T^{-1}(0, c] = P\{\omega : \omega + \alpha \leq c\} + P\{\omega : 1 < \omega + \alpha \leq 1 + c\}\]

If \(c < \alpha\), this is \(0 + 1 + c - \alpha - (1 - \alpha) = c\), and the same result is obtained if \(c \geq \alpha\). So, \(P \circ T^{-1}[0, c] = c = P[0, c]\) for all \(0 \leq c < 1\) and, therefore, \(P \circ T^{-1} = P\).

\(b)\). Alternatively, if \(\Omega = (-\pi, \pi]\) and \(P = \lambda/2\pi\), where \(\lambda\) is the restriction of Lebesgue measure to \(\Omega\), then \(T(\omega) = \alpha + \omega\) (mod \(2\pi\)) preserves \(P\) for any \(-\pi < \alpha \leq \pi\).

**Remarks 1 (Used in the Example).** If \(\mathcal{C}\) is a \(\pi\)-system for which \(\mathcal{A} = \sigma(\mathcal{C})\) and if \(T : \Omega \rightarrow \Omega\) is a transformation for which \(T^{-1}(C) \in \mathcal{A}\) and \(P \circ T^{-1}(C) = P(C)\) for all \(C \in \mathcal{C}\), then \(T\) is measurable and \(T\) preserves \(P\).

2. If \(T\) preserves \(P\) and \(\psi : \Omega \rightarrow \mathbb{R}\) is a non-negative measurable functions, then

\[\int_{\Omega} \psi \circ T dP = \int_{\Omega} \psi dP,\]

by the Transformation Theorem.

3. If \(T\) preserves \(P\), then so do the iterated \(T^k = T \circ \cdots \circ T\), by induction.

4. If \(T\) preserves \(P\) and \(T\) has measurable inverse \(S\), say, then \(S\) preserves \(P\). For, if \(A \in \mathcal{A}\), then \(P(A) = P\{T^{-1}[T(A)]\} = P[T(A)] = P[S^{-1}(A)]\).

**Shift Transformations.** Let \(\mathbb{K} = \mathbb{N}\) or \(\mathbb{Z}\); let \((\mathcal{X}, \mathcal{B}, F)\) be any probability space; and consider the product space, \((\mathcal{X}^{\mathbb{K}}, \mathcal{B}^{\mathbb{K}}, F^{\mathbb{K}})\). Denote elements of \(\mathcal{X}^{\mathbb{K}}\) by \(x = (x_0, x_1, x_2, \cdots)\) or \(x = (\cdots, x_{-1}, x_0, x_1, x_2, \cdots)\) and define \(U : \mathcal{X}^{\mathbb{K}} \rightarrow \mathcal{X}^{\mathbb{K}}\), by

\[U(x)_k = x_{k+1}\]

for \(k \in \mathbb{K}\) and \(\omega \in \Omega\). Then \(U\) preserves \(P\). For if \(R\) is a measurable rectangle, say \(R = \{x : x_k \in B_k\text{ for all } k \in J\}\), where \(J\) is a finite set and \(B_k \in \mathcal{B}\) for all \(k \in J\), then \(U^{-1}(R) = \{x : x_j \in B_{j-1}\text{ for } j \in J + 1\}\) and

\[P \circ U^{-1}(R) = \prod_{j \in J + 1} F\{B_{j-1}\} = \prod_{k \in J} F\{B_k\} = P(R).\]

This transformation is called the left shift transformation. If \(\mathbb{K} = \mathbb{Z}\), then it is easily seen that \(U^{-1}(x)_k = x_{k-1}\) for all \(k\) and \(\omega\), and \(U^{-1}\) is called the right shift transformation. \(\diamond\)
Stationary Sequences. If \( X_k, \; k \in \mathcal{I} \), are random elements with values in a measurable space \((\mathcal{X}, \mathcal{B})\), say, then we may define a measurable transformation \( \mathbf{X} : \Omega \to \mathcal{X}^\mathcal{I} \) by
\[
\mathbf{X}(\omega)_k = X_k(\omega)
\]
for \( k \in \mathcal{I} \) and \( \omega \in \Omega \). Let
\[
Q = P \circ \mathbf{X}^{-1},
\]
the induced distribution of \( \mathbf{X} \). Then the sequence is said to stationary if the shift transformation \( U \) preserves \( Q \); that is, \( X_k, \; k \in \mathcal{I} \), is stationary sequence if \( P\{U(\mathbf{X}) \in B\} = P\{\mathbf{X} \in B\} \) for all measurable \( B \subseteq \mathcal{X}^\mathcal{I} \). Since measurable rectangles generate the product sigma-algebra, this is equivalent to
\[
P\{X_{k+1} \in B_k, \; k = m, \ldots, n\} = P\{X_k \in B_k, \; k = m, \ldots, n\}
\]
for all choices of \( m \leq n \) in \( \mathcal{I} \) and all choices of \( B_k \in \mathcal{B} \).

Examples 2-a) If \( X_k, \; k \in \mathcal{I} \), are i.i.d. with common distribution \( F \), then both sided of (*) are \( \prod_{k=m}^{n} F\{B_k\} \), and the sequence is stationary.

b) If \( T \) preserves \( P \) and \( f : \Omega \to \mathcal{X} \) is measurable, then \( X_k = f \circ T^k, \; k = 0, 1, 2, \ldots \) is a stationary sequence. For if \( B_k \subseteq \mathcal{X} \) are measurable, then
\[
P\{X_{k+1} \in B_k, \; k = m, \ldots, n\} = P\left( \bigcap_{k=m}^{n} T^{-k-1}[f^{-1}(B_k)] \right)
= P\left( \bigcap_{k=m}^{n} T^{-k}[f^{-1}(B_k)] \right)
= P\{X_k \in B_k, \; k = m, \ldots, n\}.
\]

c) If \( T \) is invertible in b), then \( X_k = f \circ T^k, \; k \in \mathbb{Z} \), is a stationary sequence. ◊

Example 1: Continued. If \( \Omega = (-\pi, \pi], \; P = \lambda/2\pi \), and \(-\pi < \alpha < \pi\), then \( X_k(\omega) = \cos(k\alpha + \omega) \) defines a stationary sequence. ◊

Proposition 1. If \( X_k, \; k \in \mathcal{I} \), be a stationary sequence with values in a measurable space \((\mathcal{X}, \mathcal{B})\), let \((\mathcal{Y}, \mathcal{C})\) be a second measurable space, and let \( \phi : \mathcal{X}^\mathcal{I} \to \mathcal{Y} \) be a measurable function. Define \( Y_k, \; k \in \mathcal{I} \), by
\[
Y_k = \begin{cases} 
\phi(X_k, X_{k+1}, \ldots) & \text{if } \mathcal{I} = \mathbb{N} \\
\phi(\ldots, X_{k-1}, X_k, X_{k+1}, \ldots) & \text{if } \mathcal{I} = \mathbb{Z}
\end{cases}
\]
Then \( Y_k, \; k \in \mathcal{I} \), is a stationary sequence.

Proof. Observe that \( Y_k = \phi[U^k(\mathbf{X})], \; k \in \mathcal{I} \), in both cases. Define \( \psi : \mathcal{X}^\mathcal{I} \to \mathcal{Y}^\mathcal{I} \) and \( \psi(x)_k = \phi[U^k(x)] \) for \( k \in \mathcal{I} \) and \( x \in \mathcal{X} \). Then \( \mathbf{Y} = \psi(\mathbf{X}) \). So, if \( Q \) and \( R \) denote the induced distributions of \( \mathbf{X} \) and \( \mathbf{Y} \), then \( R = Q \circ \psi^{-1} \). Next, let \( U \) and \( V \) denote the shift transformations on \( \mathcal{X}^\mathcal{I} \) and \( \mathcal{Y}^\mathcal{I} \). Then
\[
\psi \circ U = V \circ \psi.
\]
So,
\[
R \circ V^{-1} = Q \circ \psi^{-1} \circ V^{-1} = Q \circ U^{-1} \circ \psi^{-1} = Q \circ \psi^{-1} = R,
\]
as required. ◊
Corollary. If $X_k$, $k \in K$, are i.i.d., then $Y_k$, $k \in K$, is a stationary sequence.

**Remark.** The $Y_k$'s are called a moving function of $X_k$.

**Linear Processes.** Let $X_k$, $k \in \mathbb{Z}$, be i.i.d. random variables. If $E|X_0| < \infty$ and $a_k$, $k \in \mathbb{Z}$, are summable, then

$$Y_k = \sum_{j \in \mathbb{Z}} a_j X_{k-j}$$

converges w.p.1 for each $k$ and defines a stationary sequence. The sum also defines a stationary sequence if $E(X_0) = 0$, $E(X_0^2) < \infty$, and $a_j$, $j \in \mathbb{Z}$, are square summable. ◊

**Example 3: Autoregressive processes.** Let $\cdots X_{-1}, X_0, X_1, X_2, \cdots$ be i.i.d. with a finite mean; let $-1 < \alpha < 1$; and let

$$Y_k = \sum_{j=0}^{\infty} \alpha^j X_{k-j}$$

for $k = 0, \pm 1, \pm 2, \cdots$. Then $\cdots Y_{-1}, Y_0, Y_1, Y_2, \cdots$ is a stationary sequence for which

$$Y_k = \alpha Y_{k-1} + X_k$$

for all $k$. This process is called an autoregressive process of order one. ◊

**Invariant Sets.** If $T$ preserves $P$, then a measurable $A \subseteq \Omega$ is said to be invariant if $T^{-1}(A) = A$ and almost invariant if $P[A \Delta T^{-1}(A)] = 0$, where $\Delta$ denotes symmetric difference. Let $\mathcal{I}_0$ denote the collection of all invariant sets, and $\mathcal{I}$ the collection of all almost invariant sets.

**Lemma 1.** $\mathcal{I}_0$ and $\mathcal{I}$ are sigma-algebras. Further, if $B \in \mathcal{I}$, then there is a $A \in \mathcal{I}_0$ for which $P(A \Delta B) = 0$.

**Proof.** The proof that $\mathcal{I}_0$ and $\mathcal{I}$ are sigma-algebras is left as an exercise. If $B \in \mathcal{I}$, then $A = \cap_{m=1}^{\infty} \cup_{n=m}^{\infty} T^{-n}(B)$ has the desired properties. ◊

**Lemma 2.** A measurable function $Y : \Omega \to \mathbb{R}$ is $\mathcal{I}$-measurable iff $Y \circ T = Y$ a.e.

**Proof.** If $Y = Y \circ T$ a.e. and $B \subseteq \mathbb{R}$ is a Borel set, then

$$P\{Y^{-1}(B) \Delta T^{-1}[Y^{-1}(B)]\} \leq P\{Y \neq Y \circ T\} = 0$$

so that $Y^{-1}(B) \in \mathcal{I}$ and, therefore, $Y$ is $\mathcal{I}$-measurable. Conversely, if $Y$ is $\mathcal{I}$-measurable, let

$$B_{nk} = Y^{-1}\left(\frac{k-1}{2^n}, \frac{k}{2^n}\right]$$

for $k \in \mathbb{Z}$ and $n \geq 1$. Then $B_{nk} \in \mathcal{I}$ for all $k$ and $n$, and

$$Y = \lim_{n \to \infty} \sum_{k \in \mathbb{Z}} \frac{k}{2^n} 1_{B_{nk}} = \lim_{n \to \infty} \sum_{k \in \mathbb{Z}} \frac{k}{2^n} 1_{B_{nk} \circ T} = Y \circ T \text{ a.e.},$$

as asserted. ◊

3
Ergodicity and Mixing. A measure preserving transformation $T$ is said to be ergodic (with respect to $P$) iff $T$ is trivial; that is, $T$ is ergodic iff $P(I) = 0$ or 1 for all $I \in \mathcal{I}$. By Lemma 1, this is equivalent to $P(A) = 0$ or 1 for all $A \in \mathcal{I}_0$. Further, $T$ is said to be mixing (with respect to $P$) iff

$$\lim_{n \to \infty} P[A \cap T^{-n}(B)] = P(A)P(B)$$

for all $A, B \in \mathcal{A}$. A stationary sequence $X_k$, $k \in \mathbb{N}$, is said to be mixing or ergodic iff the shift transformation $U$ is mixing or ergodic with respect to $Q = P \circ X^{-1}$.

**Example 1:** Again. If $\alpha$ is rational in Example 1-a), say $\alpha = p/q$, then $\{\omega : \langle q\omega \rangle \leq 1/2\}$ is an invariant set, since $\langle qT(\omega) \rangle = \langle q(\alpha + \omega) \rangle = \langle q\omega \rangle$ for all $\omega \in \Omega$. So, $T$ is not ergodic in this case. If $\alpha$ is irrational, then it can be shown that $T$ is ergodic. See the text.

**Proposition 2.** If $T$ preserves $P$ and $T$ is mixing, then $T$ is ergodic.

**Proof.** If $A \in \mathcal{I}$, then $P(A) = P[A \cap T^{-n}(A)] \to P(A)^2$ as $n \to \infty$ and, therefore, $P(A) = 0$ or 1. \hfill \Box

**Lemma 3.** Let $\mathcal{C}$ be an algebra for which $\mathcal{A} = \sigma(\mathcal{C})$. If $T$ preserves $P$ and $\lim_{n \to \infty} P[A \cap T^{-n}(B)] = P(A)P(B)$ for all $A, B \in \mathcal{C}$, then $T$ is mixing.

**Proof.** Given $A, B \in \mathcal{A}$ and $\epsilon > 0$, there are $A_0, B_0 \in \mathcal{C}$ for which $P(A\Delta A_0) + P(B\Delta B_0) \leq \epsilon$, by the Approximation Theorem; then $P[A\Delta A_0] + P[T^{-n}(B)\Delta T^{-n}(B_0)] \leq \epsilon$ for all $n$. By assumption, there is an $m$ for which $|P[A_0 \cap T^{-n}(B_0)] - P(A_0)P(B_0)| \leq \epsilon$ for all $n \geq m$. Clearly,

$$|P[A \cap T^{-n}(B)] - P(A)P(B)| \leq |P[A \cap T^{-n}(B)] - P[A_0 \cap T^{-n}(B_0)]| + |P[A_0 \cap T^{-n}(B_0)] - P(A_0)P(B_0)| + |P(A_0)P(B_0) - P(A)P(B)|.$$

If $n \geq m$, then the middle term is at most $\epsilon$. For the first and last terms,

$$|P[A \cap T^{-n}(B)] - P[A_0 \cap T^{-n}(B_0)]| \leq P(A\Delta A_0) + P[T^{-n}(B)\Delta T^{-n}(B_0)] \leq \epsilon$$

and

$$|P(A_0)P(B_0) - P(A)P(B)| \leq P(A\Delta A_0) + P(B\Delta B_0) \leq \epsilon.$$

So, $|P[A \cap T^{-n}(B)] - P(A)P(B)| \leq 3\epsilon$ for all $n \geq m$. \hfill \Box

**Proposition 3.** The shift transformation is mixing with respect to any power measure $Q = F^k$.

**Proof.** Let $\mathcal{C}$ be the class of all cylinder sets

$$C = \{x : [x_k, \ldots, x_\ell] \in B\}$$

where $k \leq \ell$ in $\mathbb{N}$ and $B \subseteq \mathcal{X}^{\ell-k}$ is measurable. Then $\mathcal{C}$ is an algebra that generates the product sigma-algebra. If $C_1$ and $C_2$ are any two cylinder sets, then $C_1$ and $T^{-n}(C_2)$ are independent for all sufficiently large $n$ and, therefore, $\lim_{n \to \infty} P[C_1 \cap T^{-n}(C_2)] = P(C_1)P(C_2)$. The theorem now follows directly from the previous lemma. \hfill \Box
**Corollary.** Any moving function $Y_k$ of an i.i.d. sequence $X_k$ is mixing.

**Proof.** Let $Q$ and $R$ denote the induced distributions of $X$ and $Y$ and write $Y = \psi(X)$, where $\psi \circ U = V \circ \psi$. Then

$$
R[A \cap V^{-n}(B)] = Q[\psi^{-1}(A) \cap V^{-n}(B)] \\
= Q[\psi^{-1}(A) \cap U^{-n} \circ \psi^{-1}(B)] \\
\rightarrow Q[\psi^{-1}(A)] \times Q[\psi^{-1}(B)] = R(A) \times R(B)
$$

for all measurable $A, B \subseteq \mathcal{Y}$, as asserted. \hfill \diamond

**Ergodic Theorems.** Suppose that $T$ preserves $P$, and let $f : \Omega \to \mathbb{R}$ be measurable. Let

$$
S_n = f + f \circ T + \cdots + f \circ T^{n-1},
M_n = \max[0, S_1, \ldots, S_n],
$$

and

$$
f^* = \sup_{n \geq 1} \frac{1}{n} S_n.
$$

**Theorem 1** The Maximal Ergodic Theorem. If $f$ is integrable, then

$$
\alpha P\{\omega : f^*(\omega) > \alpha\} \leq \int_{\{f^*>\alpha\}} f dP.
$$

for every $\alpha \in \mathbb{R}$.

**Corollary.** If $\alpha > 0$, then

$$
P\{\omega : f^*(\omega) > \alpha\} \leq \frac{1}{\alpha} \int_{\{f^* > \alpha\}} f dP.
$$

**Proof.** It suffices to prove the theorem in the special case $\alpha = 0$, since the special case can be applied to $f - \alpha$. Further, since $\{f^* > 0\} = \bigcup_{n=1}^{\infty} \{M_n > 0\}$, it suffices to show that

$$
\int_{\{M_n > 0\}} f dP \geq 0. \quad (*)
$$

for each fixed $n$. To prove (*), first observe that $S_k \circ T = S_{k+1} - S_1$ for all $n$. So,

$$
f + M_n \circ T = f + \max[0, S_2 - S_1, \ldots, S_{n+1} - S_1] \geq \max[S_1, \ldots, S_n]
$$

and, therefore,

$$
f \geq \max[S_1, \ldots, S_n] - M_n \circ T = M_n - M_n \circ T
$$
on \( \{M_n > 0\} \). So,

\[
\int_{\{M_n > 0\}} f dP \geq \int_{\{M_n > 0\}} |M_n - M_n \circ T| dP
\]
\[
= \int M_n^+ dP - \int_{\{M_n > 0\}} M_n \circ T dP
\]
\[
\geq \int M_n^+ dP - \int (M_n \circ T)^+ dP = 0,
\]

as required. \(\diamondsuit\)

**The Pointwise Ergodic Theorem.** Now let

\[
\bar{f}_n = \frac{1}{n} S_n = \frac{f + f \circ T + \cdots + f \circ T^{n-1}}{n}
\]

for \( n \geq 1 \) and observe that

\[
\bar{f}_n \circ T = \left( \frac{n+1}{n} \right) f_n - \frac{f}{n}.
\]

**Theorem 2** The Pointwise Ergodic Theorem. If \( T \) preserves \( P \) and \( f \) is integrable, then \( \bar{f} = \lim_{n \to \infty} \bar{f}_n \) exists \( w.p.1 \) and in \( L^1(\Omega, P) \).

**Proof.** For rational \( -\infty < r < s \leq \infty \), let

\[
B_{r,s} = \{ \omega : \liminf_{n \to \infty} \bar{f}_n(\omega) < r < s < \limsup_{n \to \infty} \bar{f}_n(\omega) \}.
\]

Then

\[
\{ \omega : \liminf_{n \to \infty} \bar{f}_n(\omega) < \limsup_{n \to \infty} \bar{f}_n(\omega) \} \subseteq \bigcup_{r < s} B_{r,s}.
\]

It is first shown that \( P(B_{rs}) = 0 \) for all \( r < s \). To see this, we apply the Maximal Ergodic Theorem to \( g = f \times 1_{B_{rs}} \). Observe that \( B_{rs} \) is invariant. So, \( g \circ T^n = f \circ T^n \times 1_{B_{rs}} \) for all \( n \), and, therefore, \( g^* = f^* 1_{B_{rs}} \). It follows that \( B_{rs} = \{ g^* > s \} \) and

\[
sP(B_{rs}) = sP\{g^* > s\} \leq \int_{\{g^* > s\}} g^* dP = \int_{B_{rs}} f dP.
\]

Applying this result to \( -f \) then yields, \( -rP(B_{rs}) \leq \int_{B_{rs}} -fdP \) or

\[
\int_{B_{rs}} fdP \leq rP(B_{rs}).
\]

That \( P(B_{rs}) = 0 \) follows since \( r < s \). So, \( \bar{f} = \lim_{n \to \infty} \bar{f}_n \) exists \( w.p.1 \). That \( \bar{f} < \infty \) \( w.p.1 \) then follows, since \( \int_{\Omega} |\bar{f}| dP \leq \lim_{n \to \infty} \int_{\Omega} |\bar{f}_n| dP \leq \int_{\Omega} |f| dP < \infty \) by Fatou’s Lemma.

That

\[
\lim_{n \to \infty} \int_{\Omega} |\bar{f}_n - \bar{f}| dP = 0
\]
is clear if \( f \) is bounded, say \( |f| \leq c \), since then \( |\bar{f}_n| \leq c \) for all \( n \). Given any \( \epsilon > 0 \), there is a bounded \( g \) for which \( \int_\Omega |g - f|dP \leq \epsilon \) in which case,

\[
\int_\Omega |\bar{g}_n - \bar{f}_n|dP \leq \int_\Omega |g - f|dP \leq \epsilon
\]

for all \( n \) and, therefore,

\[
\int_\Omega |\bar{g} - \bar{f}|dP \leq \liminf_{n \to \infty} \int_\Omega |\bar{g}_n - \bar{f}_n|dP \leq \epsilon,
\]

by Fatou’s Lemma. It follows that

\[
\limsup_{n \to \infty} \int_\Omega |\bar{f}_n - \bar{f}|dP \leq \limsup_{n \to \infty} \left[ \int_\Omega |\bar{f}_n - \bar{g}_n|dP + \int_\Omega |\bar{g}_n - \bar{g}|dP + \int_\Omega |\bar{g} - \bar{f}|dP \right] = 2\epsilon,
\]

and, therefore, that \( \limsup_{n \to \infty} \int_\Omega |\bar{f}_n - \bar{f}|dP = . \)

\textbf{Remark.} If \( T \) is ergodic, then any \( \mathcal{I} \)-measurable function is constant \( w.p.1 \).

\textbf{Corollary 1.} If \( T \) is ergodic, then

\[
\lim_{n \to \infty} \frac{f + f \circ T + \cdots + f \circ T^{n-1}}{n} = \int_\Omega f dP.
\]

\textbf{Proofs.} It is clear that \( \bar{f} \) is almost invariant and, therefore, constant \( w.p.1 \). Moreover, it follows from the Theorem, that \( \int_\Omega \bar{f} dP = \int_\Omega f dP. \)

\textbf{Corollary 2.} \( T \) is ergodic iff

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} P[A \cap T^{-k}(B)] = P(A)P(B)
\]

for all \( A, B \in \mathcal{A} \).

\textbf{Proof.} Given \( A, B \in \mathcal{A} \), let

\[
\bar{N}_n = \frac{1}{n} \sum_{k=1}^{n} 1_B \circ T^k.
\]

If \( T \) is ergodic, then \( \bar{N}_n \to P(B) \) \( w.p.1 \) and in the mean, so that

\[
\frac{1}{n} \sum_{k=1}^{n} P[A \cap T^{-k}(B)] = \int_A \bar{N}_n dP \to P(A)P(B).
\]

Conversely, if (\dagger) holds and \( A \in \mathcal{I} \), then it is easily seen that \( P(A) = P(A)^2 \) and, therefore, that \( P(A) = 0 \) or \( 1. \)
Corollary 3. Let $X_k$, $k \in \mathbb{I}$, are i.i.d.; let $\psi : \mathcal{X}^\mathbb{I} \to \mathbb{R}$ be a measurable function for which $\nu = E[\psi(X)]$ is finite; and let $Y_k = \psi(U^{k-1}(X))$ for $k = 1, 2, \cdots$. Then
\[
\lim_{n \to \infty} \frac{Y_1 + \cdots + Y_n}{n} = \nu \text{ w.p.1.}
\]

Proof. We have
\[
P\{\omega : \lim_{n \to \infty} \frac{Y_1 + \cdots + Y_n}{n} = \nu\} = F^\mathbb{I}\{\mathbf{x} : \lim_{n \to \infty} \frac{\psi(x) + \cdots + \psi(U^{n-1}\mathbf{x})}{n} = \nu\} = 1,
\]
by the Theorem and the first two corollaries.

Example 3: Revisited. In Example 2, suppose that $E(X_k) = 0$. Then
\[
\lim_{n \to \infty} \frac{Y_0^2 + \cdots + Y_{n-1}^2}{n} \to E(Y_0^2) = \frac{\sigma^2}{1 - \alpha^2},
\]
\[
\lim_{n \to \infty} \frac{Y_0Y_1 + \cdots + Y_{n-1}Y_n}{n} \to E(Y_0Y_1) = \frac{\alpha\sigma^2}{1 - \alpha^2}
\]
w.p.1 by the Pointwise Ergodic Theorem. So,
\[
\hat{\alpha}_n = \frac{Y_0Y_1 + \cdots + Y_{n-1}Y_n}{Y_0^2 + \cdots + Y_{n-1}^2} \to \alpha \text{ w.p.1.}
\]

The Range of a Random Walk. Let $\cdots, X_{-1}, X_0, X_1, X_2, \cdots$ be i.i.d. integer valued random variables and let $S_n = X_1 + \cdots + X_n$ for $n \geq 1$. Further, let
\[
R_n = \#\{S_1, \cdots, S_n\}.
\]

Spitzer’s Theorem. With probability one,
\[
\lim_{n \to \infty} \frac{R_n}{n} = P\{S_k \neq 0, \ k = 1, 2, \cdots\}.
\]

Proof. Let
\[
\psi(x) = 1_{\{x_0 + \cdots + x_{-k} \geq 0, \ k = 0, 1, 2, \cdots\}}
\]
and
\[
\psi_m(x) = 1_{\{x_0 + \cdots + x_{-k} \geq 0, \ k = 0, \cdots, m-1\}}.
\]
Then
\[
R_n = \sum_{k=1}^{n} 1_{\{S_k \neq S_j, j=0, \cdots, k-1\}}
\]
\[
= \sum_{k=1}^{n} 1_{\{X_k + \cdots + X_j \neq 0, j=1, \cdots, k-1\}}
\]
\[
= \sum_{k=1}^{n} \psi_k[U^k(X)]
\]
So,

\[ R_n \leq m + \sum_{k=m+1}^{n} \psi_m[U^k(X)], \]

\[ R_n \geq \sum_{1}^{n} \psi[U^k(X)], \]

for any \( 1 \leq m \leq n < \infty \). It follows that

\[ \limsup_{n \to \infty} \frac{R_n}{n} \leq E[\psi_m(X)], \]

\[ \liminf_{n \to \infty} \frac{R_n}{n} \geq E[\psi(X)], \]

\( w.p.1 \) for any \( m \). Finally,

\[ E\{\psi_m[X]\} = P[S_k \neq 0, \ k = 1, \cdots, m] \]
\[ \rightarrow P[S_k \neq 0, \ k = 1, 2 \cdots] \]
\[ = E\{\psi[X]\} \]

as \( m \to \infty \) and the theorem follow. \( \diamond \)