INTRODUCTION: Some Basic Concepts

Jesús Gonzalo

Universidad Carlos III, Madrid

jesus.gonzalo@uc3m.es

October 19, 2021

Some Useful References:

- A.N. Shiryayev (2016). Probability (a complete book)
- Peter Phillips. Lecture Notes (great notes)
- K. Saxe (2001). Beginning Functional Analysis (easy book)

Time Series Econometrics requires a knowledge of Probability Theory. This Introduction is based on the above references.

Table of Contents

Basic Concepts

- Take $\Omega = [0, 1)$ and consider the problem of choosing points at random from this set.
- Symmetry: all the points are equiprobable.
- The set $[0, 1)$ is uncountable, and if we suppose its probability is 1, then

$$P(\omega) = 0 \quad \forall \omega \in \Omega$$

- This approach doesn't lead very far. For instance, if $P(A)$ is defined by

$$P(A) = \sum_{\omega \in A} P(\omega)$$

  The above assignment of probabilities $(P(\omega) = 0, \omega \in \Omega)$ doesn't let us to define the probability that a point chosen at random from $[0, 1)$ belongs to the set $A = [0, \frac{1}{2})$.
- These remarks suggest that in constructing probabilistic models for uncountable spaces Ω we must ASSIGN PROBABILITIES not to individual elements of Ω but to <u>subsets</u>

Basic Concepts

Definition (1.1)

Let Ω be a set of points ω. A system $\mathcal{A}$ of subsets of Ω is called an algebra if

- $\Omega \in \mathcal{A}$.
- $A, B \in \mathcal{A} \implies A \cup B \in \mathcal{A}$, and $A \cap B \in \mathcal{A}$ (by Morgan's law).
- $A \in \mathcal{A} \implies \overline{A} \in \mathcal{A}$.

Definition (1.2)

Let $\mathcal{A}$ be an algebra of subsets of Ω. A set function $\mu = \mu(A)$, $A \in \mathcal{A}$ taking values in $[0, \infty)$ is called a finitely additive measure defined on $\mathcal{A}$ if

$$\mu(A + B) = \mu(A) + \mu(B)$$

for every disjoint sets A, $B \in \mathcal{A}$.

When $\mu(\Omega) < \infty$, μ is called finite.
When $\mu(\Omega) = 1$, μ is called finitely additive probability measure.

Basic Concepts

Definition (1.3)

An ordered triple $(\Omega, \mathcal{A}, P)$, where

- Ω is a set of point ω
- $\mathcal{A}$ is an algebra of subsets of Ω
- P is a finitely additive probability on A

is a probabilistic model

However, this definition is too broad to lead to a fruitful mathematical theory.

Definition (1.4)

A system $\mathcal{F}$ of subsets of Ω is a σ-algebra if it is an algebra and satisfies
If $\mathcal{A}_n \in \mathcal{F}$, $n = 1, 2, \ldots$ then $\cup \mathcal{A}_n \in \mathcal{F}$, and $\cap \mathcal{A}_n \in \mathcal{F}$ (by Morga's law)

Definition (1.5)

The space Ω together with a σ-algebra $\mathcal{F}$ of its subsets is a measurable space, and is denoted by $(\Omega, \mathcal{F})$.

Definition (1.6)

A finitely additive measure μ defined on an algebra $\mathcal{A}$ of subsets of Ω is countably additive, or simply a measure, if for all pairwise disjoint subsets $A_1, A_2, \ldots$ of $\mathcal{A}$,

$$\mu\left(\sum_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

If $P(\Omega) = 1$, it is called a probability measure.

Probability measures have the following properties:

- If $\emptyset$ is the empty set, then $P(\emptyset) = 0$
- If $A, B \in \mathcal{A}$ then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- If $A, B \in \mathcal{A}$ and $B \subseteq A$, then $P(B) \leq P(A)$.
- If $A_n \in \mathcal{A}$, $n = 1, 2, \ldots$ and $\cup A_n \in \mathcal{A}$, then $P(A_1 \cup A_2 \cup \ldots) \leq P(A_1) + P(A_2) + \ldots$

Basic Concepts

Definition (1.7)

An ordered triple $(\Omega, \mathcal{F}, P)$ where

- Ω is a set of points ω.
- $\mathcal{F}$ is a σ-algebra of subsets of Ω.
- P is a probability on $\mathcal{F}$.

is called a probability model or probability space.

Example $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

Let $\mathbb{R} = (-\infty, \infty)$, and $(a, b] = \{x \in \mathbb{R} : a < x \leq b\}$. Let $\mathcal{A}$ be the system of subsets of $\mathbb{R}$ which are finite sums of disjoint intervals of the form $(a, b]$:

$$A \in \mathcal{A} \quad if \quad A = \sum_{i=1}^{n} (a_i, b_i] \quad n < \infty$$

This system is an algebra but not a σ-algebra: if $A_n = (0, 1 - \frac{1}{n}] \in \mathcal{A}$ we have $\cup_n A_n = (0, 1) \notin \mathcal{A}$

Basic Concepts

Example $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

...Let $\mathcal{B}(\mathbb{R})$ be the smallest σ-algebra $\sigma(\mathcal{A})$ containing $\mathcal{A}$. This σ-algebra is called the Borel algebra of subsets of the real line, and its sets are Borel sets.

Check that $\mathcal{B}(\mathbb{R})$ contains any type of interval you can imagine of the real line: (a, b), $[a, b)$, $\{a\}$, etc.

Example $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$

Define $\mathbb{R}^n = \mathbb{R} \times ... \times \mathbb{R}$ and $I = I_1 \times ... \times I_n$ where $I_k = (a_k, b_k]$. Let $\mathcal{F}$ be the set of all rectangles I. The smallest σ-algebra $\sigma(\mathcal{F})$ generated by the system $\mathcal{F}$ is the Borel algebra of subsets of $\mathbb{R}^n$ and is denoted by $\mathcal{B}(\mathbb{R}^n)$.

Example $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$

$\mathbb{R}^\infty$ is the space of ordered sequences of numbers $x = (x_1, x_2, ...)$, $-\infty < x_k < \infty$, $k = 1, 2, ...$ Consider the cylinder set
$\mathcal{F}(I_1 \times ... \times I_n) = \{x : x = (x_1, x_2, ...), x_1 \in I_1, ..., x_n \in I_n\}$. Then $\mathcal{B}(\mathbb{R}^\infty)$ is the smallest σ-algebras containing the above set.

Basic Concepts

Example $(\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T))$

Where T is an arbitrary set. The space $\mathbb{R}^T$ is the collection of real functions $x = (x_t)$ defined for $t \in T$. The cylinder set considered is

$$\mathcal{F}_{t_1,...,t_n}(I_1 \times ... \times I_n) = \{x : x_{t_1} \in I_1, ..., x_{t_n} \in I_n$$

and $\mathcal{B}(\mathbb{R}^T)$ is the smallest σ-algebra corresponding to this cylinder set.
Note that $\mathcal{F}_{t_1,...,t_n}(I_1 \times ... \times I_n)$ is just the set of functions that, at times $t_1, ..., t_n$ "get through the windows" $I_1 \times ... \times I_n$ and at other times have arbitrary values.

Example $(\mathcal{C}, \mathcal{B}(\mathcal{C}))$

Let $T = [0, 1]$ and let $\mathcal{C}$ be the space of continuous functions $x = (x_t)$, $0 \leq t \leq 1$. This is a metric space with the metric $f(x, y) = \underset{t \in T}{sup}|x_t - y_t|$. We introduce this metric because with $f(x, y)$ the σ-algebra generated by the standard cylinders and by open sets coincide (remember what you learnt in the Math courses, separability, complete space, etc.).

Example $(\mathcal{D}, \mathcal{B}(\mathcal{D}))$

Where $\mathcal{D}$ is the space of functions $x = (x_t)$, $t \in [0,1]$, that are continuous on the right ($x_t = x_{t_+}$ for all $t < 1$) and have limits from the left (at every $t > 0$).
Just as for $\mathcal{C}$, we can introduce a metric $d(x,y)$ on $\mathcal{D}$ such that the σ-algebra generated by the standard cylinders and by open sets coincide. In order to define open sets we need a metric $d(x,y)$ in this case is the one defined by sicorohod

$$d(x,y) = inf\{\epsilon > 0 : \exists \lambda \in \Lambda : \underset{t}{sup}|x_t - y_{\lambda(t)}| + \underset{t}{sup}|t - \lambda(t)| < \epsilon\}$$

where Λ is the set of strictly increasing functions $\lambda = \lambda(t)$ that are continuous on $[0,1]$ and have $\lambda(0) = 0, \lambda(1) = 1$.

Probability Measures on Measurable Spaces

Example $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

Let $P = P(A)$ be a probability measure defined on the Borel subsets A of the real line. $F(x) = P(-\infty, x]$, $x \in \mathbb{R}$. This function is called distribution function. There is a one-to-one correspondence between probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and distribution functions on $\mathbb{R}$.

Example $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$

Consider the n-dimensional distributional function
$F_n(x_1, ..., x_n) = P((-\infty, x_1] \times ... \times (-\infty, x_n]) = P(\infty, \mathbf{x}]$ where $\mathbf{x} = (x_1, ..., x_n)$.
Define
$\triangle_{a_i, b_i} F_n(x_1, ..., x_n) = F_n(x_1, ..., x_{i-1}, b_i, x_{i+1}, ...) - F_n(x_1, ..., x_{i-1}, a_i, x_{i+1}, ...)$
where $a_i \leq b_i$.
A simple calculation shows that $\triangle_{a_1, b_1} \times ... \times \triangle_{a_n, b_n} F_n(x_1, ..., x_n) = P(a, b]$ where $(a, b) = (a_1, b_1] \times ... \times (a_n, b_n]$. In general, $P(a, b] \neq F_n(b) - F_n(a)$. Represent this graphically for $n = 2$.

Probability Measures on Measurable Spaces

Example $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$

Take $P_n(B) = P(\mathcal{F}_n(B)), B \in \mathcal{B}(\mathbb{R}^n)$ with

$$\mathcal{F}_n(B) = \{x \in \mathbb{R}^\infty : (x_1, ..., x_n) \in B\}, B \in \mathcal{B}(\mathbb{R}^n)$$

The sequence of probability measures $P_1, P_2, ...$ defined respectively on $(\mathbb{R}, \mathcal{B}(\mathbb{R})), (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2)), ...$ and $B \in \mathcal{B}(\mathbb{R}^n)$, for $n = 1, 2, ...$, have the following "consistency" property

$$P_{n+1}(B \times \mathbb{R}) = P_n(B)$$

Kolmogorov's Theorem: Let $P_1, P_2, ...$ be a sequence of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R})), (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2)), ...$ possessing the above consistency property.
Then there is a unique probability measure P on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ such that $P(\mathcal{F}_n(B)) = P_n(B)$, $B \in \mathcal{B}(\mathbb{R}^n)$ for $n = 1, 2, ...$

Random Variables

Definition (3.1)

A real function $\xi = \xi(\omega)$ defined on $(\Omega, \mathcal{F})$ is a $\mathcal{F}$-measurable function, or a <u>random variable</u> if

$$\{\omega : \xi(\omega) \in B\} \in \mathcal{F}$$

for every $B \in \mathcal{B}(\mathbb{R})$; or, equivalently, if the inverse image

$$\xi^{-1}(B) = \{\omega : \xi(\omega) \in B\}$$

is a measurable set in Ω.

Example

Consider the probability space $(\Omega, \mathcal{F}, P)$ where

$$\Omega = \{1, 2, 3\} \quad and \quad F = (\Omega, \emptyset, \{1, 2\}, \{3\})$$

Then $\{1\}$ is not a measurable set, and therefore the identity function, with $f(1) = 1, f(2) = 2$ and $f(3) = 3$, which maps from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B})$ is not a measurable function. Therefore f is not a random variable.

Random Variables

Definition (3.2)

A probability measure P_ξ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with $P_\xi(B) = P(\omega : \xi(\omega) \in B)$, $B \in \mathcal{B}(\mathbb{R})$ is called the probability distribution of ξ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Definition (3.3)

The function $F_\xi(x) = P(\xi(\omega) \leq x), x \in \mathbb{R}$ is called the distribution function of ξ.

Definition (3.4)

Let $(\Omega, \mathcal{F})$ and $(E, \mathcal{E})$ be measurable spaces. We say that a function $X = x(\omega)$ defined on Ω and taking values in E, is $\mathcal{F}/\mathcal{E}$-measurable, or is a random element (with values in E), if $\{\omega : x(\omega) \in B\} \in \mathcal{F}$ for every $B \in \mathcal{E}$.

Some special cases:

- If $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R})) \implies$ random element=random variable
- Let $(E, \mathcal{E}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Then a random element $X(\omega)$ is a "random point in $\mathbb{R}^n$. If R_k is the projection of $\mathbb{R}^n$ on the k-th coordinate axis, $X(\omega)$ can be represented in the form $X(\omega) = (\xi_1(\omega), ..., \xi_n(\omega))$ where $\xi_k = R_k o X$.

Definition (3.5)

An ordered set $(\eta_1(\omega), ..., \eta_n(\omega))$ of random variables is called an n-dimensional <u>random vector</u>.

Random Elements, Stochastic Processes,... Time Series

Definition (Continued)

Let $(E, \mathcal{E}) = (\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T))$, where T is a subset of the real line. In this case every random element $X = X(\omega)$ can evidently be represented as

$$X = (\xi_t)_{t \in T}$$

with $\xi_t = \pi_t \circ X$, and is called a random function with time domain T.

Definition (3.6)

Let T be a subset of the real line. A set of random variables

$$X = (\xi_t)_{t \in T}$$

is called a random process with time domain T.
If $T = \{1, 2, ...\}$ we call $X = (\xi_1, \xi_2, ...)$ a random process with discrete time or a random sequence.
If $T = [0, 1]$, $(-\infty, \infty)$, $[0, \infty)$,..., we call $X = (\xi_t)_{t \in T}$ a random process with continuous time.

It is easy to check that every random process $X = (\xi_t)_{t \in T}$ is also a random function on the space $(\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T))$.

Definition (3.7)

Let $X = (\xi_t)_{t \in T}$ be a random process. For each given $\omega \in \Omega$ the function $(\xi_t(\omega))_{t \in T}$ is said to be a realization or a trajectory of the process, corresponding to the outcome ω, or a TIME SERIES.

Definition (3.8)

Let $X = (\xi_t)_{t \in T}$ be a random process. The probability measure P_X on $(\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T))$ defined by

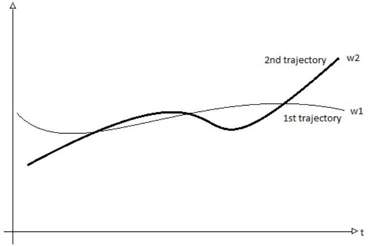
$$P_X(B) = P\{\omega : X(\omega) \in B\}, \; B \in \mathcal{B}(\mathbb{R}^T)$$

is called the probability distribution of X.
The probabilities

$$P_{t_1,\ldots,t_n}(B) \equiv P\{\omega : (\xi_{t_1}, \ldots, \xi_{t_n}) \in B\}$$

with $t_1 < t_2 < \ldots < t_n$, $t_i \in T$, are called finite-dimensional distribution functions.

A TIME SERIES is a realization of a stochastic process.



$X(\omega, t) = X_t(\omega)$
For fixed t, the function $X(., t)$ is a random variable.
For fixed ω, the function $X(\omega, .)$ is a sample path of the stochastic process.
We have

$$(\Omega \times T) \to \mathbb{R}$$
$$(\Omega, \mathcal{F}) \to (\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T))$$
$$(\mathbb{R}^\infty, \mathcal{B}_\infty, P) \to (X_1(..., x_{-1}, x_0, x_1, ...), X_2(..., x_{-1}, x_0, x_1, ...), ...) = (..., x_1, x_2, ...)$$

An Example of a Stochastic Process

Example

Let the index set be $T = \{1, 2, 3\}$ and let the space of outcomes (Ω) be the possible outcomes associated with tossing one dice:

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

define

$$X(t, \omega) = t + [\text{value on dice}]^2 \times t.$$

Therefore, for a particular ω, say $\omega_3 = \{3\}$, the realization or path would be $(10, 20, 30)$.
In this case Ω and T are finite. There are 6 realizations. Draw them.

More Examples of Stochastic Processes

Example

Let $X = \{X_t\}_{t \in T}$.

(a) Discrete stochastic processes: $T = \{0, 1, 2, ...\}$

(i) $\{X_t\}_{t=0}^{\infty} \equiv$ i.i.d. $(0, \sigma)$

(ii) $\{X_t\}_{t=0}^{\infty} \equiv$ AR(1), i.e., $X_t = AX_{t-1} + u_t$, $\{u_t\} \equiv i.i.d.(0, \sigma)$ and $X_0 \equiv$ intitial condition

(iii) $\{X_t\}_{t=0}^{\infty} \equiv$ Random Walk (σ), i.e., $X_t = X_{t-1} + u_t$, $\{u_t\} \equiv i.i.d.(0, \sigma)$.

In all these cases, $(\Omega, \mathcal{F}, P) = (\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T), P)$.

More Examples of Stochastic Processes

Example

(b) Continuous stochastic processes: $T = [0, 1]$, $T = [0, \infty)$, etc.

(i) $X(t) \equiv W(t) \equiv BM(1)$ (Standard Brownian Motion), where $W(t)$ is a Gaussian random element in $C[0, 1]$, that is defined by the following properties:

(1) $W(0) = 0$
(2) $W(t) = N(0, t)$
(3) $W(s)$ independent of $W(t) - W(s)$ for $0 \leq s \leq t \leq 1$
(4) $W(t)$ has continuous sample paths

(ii) $X(t) = B(t) \equiv BM(\sigma)$, i.e., $B(t) = \sigma^{1/2} W(t)$

(iii) $X(t) \equiv BM(\mu, \sigma)$, i.e., $X(t) = \mu t + B(t)$, where $B(t) = BM(\sigma)$.

What is New?

- We want to make inference about P_X, the probability law that governs the process. The evidence that we have is a single realization $x_1, x_2, ..., x_n, ...$ We are not in good shape. BUT the situation is even worse, we only have a finite realization $x_1, x_2, ..., x_n$.
- Compare this situations with the classical case of i.i.d. samples where we have i.i.d. draws each of which is like an entire history of the time series (like in standard Econometrics).
- We require two important assumptions:

($\star$) **Stationarity** (substituting identically distributed)
($\star$) **Weak dependence** (substituting independence)

Stationarity and Ergodicity

In order to define and properly understand these two issues it is helpful to identify our abstract probability space $(\Omega, \mathcal{F}, \underline{P})$ and random sequence $\{X_n\}_{-\infty}^{\infty}$ defined on it by their coordinate representations:

(i) Define $h : \Omega \to \mathbb{R}^{\infty}$ by

$$\begin{aligned} h(\omega) &= (..., \underline{X}_{-1}(\omega), \underline{X}_0(\omega), \underline{X}_1(\omega), ...) \\ &= (..., x_{-1}, x_0, x_1, ...) \\ &= x \end{aligned}$$

(ii) $\mathcal{B}^{\infty} = \mathcal{B}(\mathcal{R}^{\infty})$: Borel σ-field on $\mathbb{R}^{\infty}$ field generated by cylinder sets $\left(\times_{-\infty}^{r-1}\mathbb{R}\right)\left(\times_{r}^{s}\mathbb{I}\right)\left(\times_{s+1}^{\infty}\mathbb{R}\right)$

(iii) $X_n : \mathbb{R}^{\infty} \to \mathbb{R}$ (coordinate functions), where $X_n(x) = x_n$. Then $\{x | ..., x_{n-1} \leq \infty, x_n \leq a, x_{n+1} \leq \infty\}$

(iv) $P = \underline{P}h^{-1}$ s.t. $PB = \underline{P}h^{-1}B$ for $B \in \mathcal{B}^{\infty}$

Thus, $(\Omega, \mathcal{F}, \underline{P}) \sim (\mathbb{R}^{\infty}, \mathcal{B}^{\infty}, P)$.

Definition (Strict Stationarity)

A random sequence $\{X_n\}_1^\infty$ is stationary (in the strict sense) if

$$P\left((X_1, X_2, ...) \in B\right) = P\left((X_{k+1}, X_{k+2}, ...) \in B\right), \ B \in \mathcal{B}(\mathbb{R}^\infty), \text{for all } k \in \mathbb{Z}$$

Example

(i) $\{X_n\}_{-\infty}^\infty$ i.i.d.

(ii) $\{Y_n\}_{-\infty}^\infty$, $Y_n = Y_n(X)$, $X = \{X_n\}_{-\infty}^\infty$.

All the measurable functions of i.i.d. sequences are strictly stationary, e.g.,

$$Y_n = \sum_{j=-\infty}^{\infty} a_j X_{n-j}, \ \sum_{-\infty}^{\infty} a_j^2 < \infty, \ \{X_n\}_{-\infty}^\infty \ i.i.d. \text{ (linear processes)}$$

Note that strict stationarity is stronger than the identical distribution assumption (i.i.d.), since the latter ONLY requires all marginals to be identical.

($\star$) Measurable functions of stationary sequences are stationary sequences. (check references)

Stationarity and Ergodicity

Definition (Shift Operator)

Consider the space $(\mathbb{R}^\infty, \mathcal{B}^\infty, P)$. The backshift operator S to the series or sequence $x = (..., x_{-1}, x_0, x_1, ...)$ can be represented as

$$
\begin{aligned}
Sx &= (..., x_0, x_1, x_2, ...) \\
S^2 x &= (..., x_1, x_2, x_3, ...) \\
&\vdots
\end{aligned}
$$

The definition above induces an operator u_s. Observe that if $\{X_n\}$ is a sequence on $(\mathbb{R}^\infty, \mathcal{B}^\infty, P)$ then

$$
\begin{aligned}
X_1(x) &= (..., x_0, x_1, x_2, ...) = x_1 \\
X_1(x) &= X_1(Sx) = x_2 \\
&\vdots \\
X_n(x) &= X_1(S^{n-1}x) = x_n
\end{aligned}
$$

Stationarity and Ergodicity

inducing

$$u_s : L_0 \left(\mathbb{R}^\infty, \mathcal{B}^\infty, P\right) \to L_0 \left(\mathbb{R}^\infty, \mathcal{B}^\infty, P\right),$$

where L_0 is the space of all real r.v.'s defined on $(\mathbb{R}^\infty, \mathcal{B}^\infty, P)$, and u_s is defined by

$$u_s X(x) = X(Sx), \text{ for any } X \in L_0(P).$$

Then

$$u_s X_n = X_{n+1}, \text{ or } X_n = u_s^{n-1} X_1,$$

giving

$$E = \{x \mid (x_n, x_{n+1}, ...) \in B\}, \; S^{-h}E = E = \{x \mid (x_{n+h}, x_{n+h+1}, ...) \in B\}$$

Definition (4.1)

A transformation $T : \Omega \to \Omega$ is measurable if for any $A \in \mathcal{F}$,

$$T^{-1}A = \{\omega : T\omega \in A\} \in \mathcal{F}.$$

Definition (4.2)

A measurable transformation T is a measure-preserving transformation if for every $A \in \mathcal{F}$,

$$P(T^{-1}A) = P(A)$$

Stationarity and Ergodicity

Examples

(1) Consider the space $(\mathbb{R}^\infty, \mathcal{B}^\infty, P)$, a stationary sequence $\{X_n\}_{-\infty}^{\infty}$ and $S : \mathbb{R}^\infty \to \mathbb{R}^\infty$ the backshift operator. Then S is measurable $(S^{-1}B^\infty \subset B^\infty)$ and measure preserving $(P(E) = P(S^{-1}E))$ if $\{X_n\}_{-\infty}^{\infty}$ is strictly stationary.

(2) Consider the space $(\Omega, \mathcal{F}, P)$ and the transformation $S = I$ $(S\omega = \omega)$. Then S is measurable and measure-preserving.

Stationarity and Ergodicity

Theorem (Kolmogorov SLLN)

Let $\{X_j\}_{-\infty}^{\infty}$ be an i.i.d. sequence of r.v.'s with $E|X_1| < \infty$. Then

$$\frac{1}{n}\sum_{j=1}^{\infty} X_j \to E(X_1) \text{ a.s.}$$

Problem:

- How to extend this to temporally dependent data?
- Is it enough to require strict stationarity?

Stationarity and Ergodicity

Example

$X_t = u_t + z$, where $\{u_t\} \equiv$ i.i.d. $U[0,1]$ and $Z \equiv N(0,1)$ independent of $\{u_t\}$.
By Kolmogorov's SLLN,

$$\bar{X} = \bar{u} + Z \overset{a.s.}{\to} \frac{1}{2} + Z \text{ random variable}$$

We want to reduce dependence:

$$Cov(X_t, X_{t+h}) = Cov(u_t + Z, u_{t+h} + Z) = Var(Z) = 1.$$

Practical Ergodicity

Idea: when do temporal averages such as

$$\frac{1}{n}\sum_{i=1}^{\infty} X_i$$

converges to ensemble (spatial) averages $E(X)$?

Stationarity and Ergodicity

Given $(\Omega, \mathcal{F}, P)$, $S : \Omega \to \Omega$ a m.p. (measure preserving):

1. An event F in $\mathcal{F}$ is invariant if $F = S^{-1}F$
2. S is ergodic if for all invariant events F

$$P(F) = \{0, 1\} \quad \text{ignorable or certain}$$

3. Strict stationary process $\{x_t\}$ $(x_t = u_S^{t-1} x_1)$ is ergodic if S is ergodic

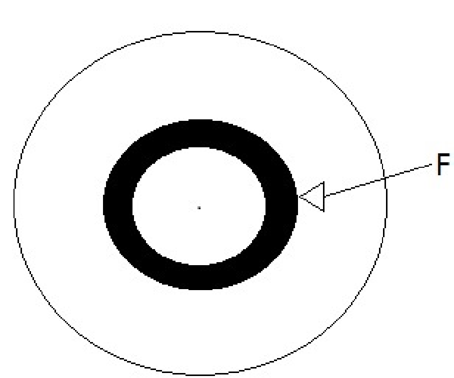
Remarks

- Absence of ergodicity means that $\exists$ invariant events F for which $0 < P(F) < 1$
- Hence it is impossible to fully sample Ω if we start off in F
- S doesn't properly mix the points of Ω

Non Ergodic Examples

Example

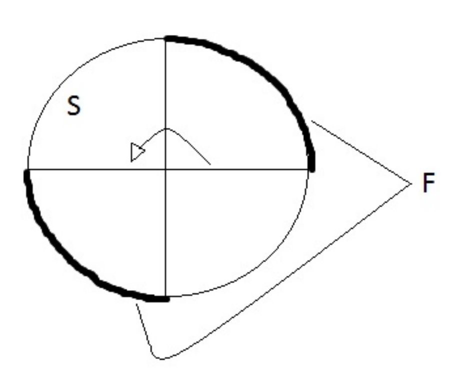
- $\Omega = \{z \mid \ |z| \leq 1\}$ Unit disk
- $F = annulus$
- S = rotation by Θ $^\circ$ $(Sz = az; a = e^{i\Theta})$
- F is invariant under S; $S^{-1}F = F$ but $0 < P(F) < 1$, so S is not ergodic.



Non Ergodic Examples

Example

- $\Omega = \{z \mid \; |z| = 1\}$ Unit circle
- $P = \textit{length}$ (normalized so P(Ω)=1)
- $S =$ rotation by 180° $(Sz = az; a = e^{i\pi})$
- F is invariant under S; $S^{-1}F = F$ but $0 < P(F) < 1$



Non Ergodic Examples: A Previous Example

Example

$\{X_t\}_{-\infty}^{\infty}$ $\quad X_t = u_t + Z \quad$ where:

u_t and Z are independent.

$u_t \equiv$ iid uniform [0,1]

$Z \equiv N(0,1)$

$$F = \cap_{t=-\infty}^{\infty} \{X_t(x) < 0\} = \{x| ...X_{t-1}(x) < 0, X_t(x) < 0, X_{t+1}(x) < 0, ...\}$$

clearly $S^{-1}F = \{x| ...X_{t-1}(SX) < 0, X_t(SX) < 0, X_{t+1}(SX) < 0, ...\}$

$$= \{X| ...X_t(x) < 0, X_{t+1}(x) < 0, X_{t+2}(x) < 0, ...\}$$

and F is invariant. But $P(F) = P(Z < -1)$ and $0 < P(F) < 1$

Some Ergodic Theorems

Theorem (Birhoff and Khinchin)

Let S be a m.p. transformation and $X = X(w)$ a random variable with $E|X| < \infty$. Then (P a.s.)

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} X(S^k w) = E(X|J)$$

where J is an invariant σ-field of $\mathcal{F}$ (σ-field of all events invariant under S).
If also S is ergodic then (P a.s.)

$$\lim_{n} \frac{1}{n} \sum_{k=0}^{n-1} X(S^k w) = E(X)$$

Some Ergodic Theorems

Example (previously discussed):

$X_t = u_t + Z = S^{t-1}u + Z$ $\qquad$ Z is an invariant r.v. under shift operator S

$\overline{X} \to_{a.s.} E[X|J] = E[u|J] + E[Z|J] = E[u] + Z = \frac{1}{2} + Z$

Prove Kolmogorov SLLN by using the above ergodic theorem and 0 - 1 law that says that if $\{X_t\}$ iid then $P(\text{tail events}) = 0$ or 1.

Theorem (Necessary & sufficient condition for ergodicity)

A measure-preserving transformation S is ergodic if and only if, for all A and $B \in \mathcal{F}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} P(A \cap S^{-k}B) = P(A)P(B)$$

Main Idea: Ergodicity of $\{X_n = S^{n-1}X\}^{\infty}$, is related to the capacity of S to thoroughly mix the points of Ω. MIXING attempts to measure this property directly.

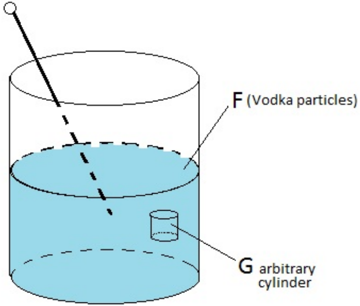
Definition (5.1)

A measure preserving transformation S: $\Omega \to \Omega$ on $(\Omega, \mathcal{F}, P)$ is mixing if $\forall F, G \in \mathcal{F}$,

$$\lim_{n\to\infty} P(F \cap S^{-n}G) = P(F)P(G)$$

Mixing and Weak Dependence: Halmos Martini Example

Ω = 90% vermouth, 10% vodka
S = action of swizzle stick on particles of Ω
P: Volume as % of vol (Ω)
F: Borel sets of R^3 in Ω



Observe proportion of vodka in an arbitrary cylinder inside Ω. If it tends to 10% the S is mixing.

Mixing and Weak Dependence: Halmos Martini Example

Set of vodka particles in G at $t = n = \{w | w \in F, S^n w \in G\} = F \cap S^{-n} G$
mixing requires

$$\frac{P(F \cap S^{-n} G)}{P(G)} \to P(F) \qquad \text{as n} \to \infty$$

i.e.

$$\frac{vol\ (vodka\ particles \cap particles\ in\ G\ after\ n^{th}\ swizzle)}{vol\ (G)} \to \qquad vol(F) = \text{\% of vodka particles}$$

Mixing and Weak Dependence

In a time series context, we say $\{X_n\}$ is mixing if

$$\begin{aligned} P(X_n \in G, X_0 \in F) &= P(X_0 \in S^{-n}G, X_0 \in F) \\ &= P(X_0 \in F \cap S^{-n}G) \\ &\to P(F)P(G) \end{aligned}$$

in other words, X_0 & X_n are independent as $n \to \infty$ (weak dependence).

Theorem

If S is mixing on $(\Omega, \mathcal{F}, P)$ then S is ergodic. The converse is not true.

Examples: $(K, \mathcal{F}, P) =$ unit circle
$K = \{z \in \mathbb{C} | \; |z| = 1\}$
$S\colon\; K \to K \qquad Sz = az; a = e^{i\Theta}$

1. S is ergodic iff $e^{i\theta}$ is not a root of unity (iff $\theta \neq \frac{2\pi}{n}$ $\forall n \in z$).
2. S is ergodic (with $a = e^{i\theta}$, $\theta \neq \frac{2\pi}{n}$ $\forall$ $n \in z$) but not mixing.

And Now We GO DOWN

- Concepts like strict stationarity and ergodicity are difficult to play with and to check.
- We will go from <u>strict stationarity</u> to <u>weak stationarity</u>.
- From <u>ergodicity at a general level</u> to <u>ergodicity for different moments</u>.

Weak Stationarity

The stochastic process $\{X_t\}_{t \in T}$ with $T = \{0, \pm 1, \pm 2, ...\}$ is said to be weak stationary if:

1. $\mathbb{E}|X_t^2| < \infty \qquad \forall t \in T$
2. $\mathbb{E}(X_t) = \mu \qquad \forall t \in T$
3. $\gamma_x(r,s) = \gamma_x(r+t, s+t) \qquad \forall r, s, t \in T$

   where

   $$\gamma_x(r,s) = Cov(X_r, X_s) = \mathbb{E}[(X_r - \mathbb{E}(X_r))(X_s - \mathbb{E}(X_s))]$$

Remember that:

Definition

The stochastic process $\{X_t, t \in T\}$ is said to be strictly stationary if the joint distribution of $(X_{t1}, ...X_{tk})'$ and $(X_{t1+h}, ...X_{tk+h})'$ are the same for all positive integers k and for all $t_1, t_2, ...t_k, h \in T$.

Strict Stationarity $\longrightarrow$ Weak Stationarity

(+ Existence of moments)

$\longleftarrow$ (if Gaussian)

Weak Stationarity: Examples

Example (1)

$$X_t = \begin{cases} Y_t \text{ if } t \text{ is even} \\ Y_t + 1 \text{ if } t \text{ is odd} \end{cases}$$

where $\{Y_t\}$ is a stationary time series.
Although $Cov(X_{t+h}, X_t) = \gamma_x(h)$, $\{X_t\}$ is not stationary because it does not have a constant mean.

Example (2)

$S_t = X_1 + X_2 + ... + X_t$ $\qquad$ $X_1, X_2 ... \sim$ are *iid* $(0, \sigma^2)$

For h>0

$$Cov(S_{t+h}, S_r) = Cov(\sum_{i=1}^{t+h} X_i, \sum_{j=1}^{t} X_j)$$

$$= Cov(\sum_{i=1}^{t} X_i, \sum_{j=1}^{t} X_j) = t\sigma^2$$

Ergodicity For The Mean

Let $(\bar{X})$ the sample mean

$$\boxed{\bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t}$$

$\{X_t\}_{t=1}^{\infty}$ is w. stationary, i.e., $\mathbb{E}(X_t) = \mu$ and $V(X_t) = \gamma_0$

$\mathbb{E}(\bar{X}) = \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}(X_t) = \mu$

$$V(\bar{X}) = \frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\text{Cov}(X_t, X_s) = \frac{\gamma_0}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\rho(t-s)$$

$$= \frac{\gamma_0}{n^2}\sum_{K=-(n-1)}^{n-1}(n - |K|)\rho_K$$

$$= \frac{\gamma_0}{n^2}\sum_{K=-(n-1)}^{n-1}(1 - \frac{|K|}{n})\rho_K \qquad (1)$$

where we let K = (t-s).

Ergodicity For The Mean

Thus if

$$\lim_{n\to\infty}[\sum_{K=-(n-1)}^{n-1}(1-\frac{|K|}{n})\rho_K]$$

is finite, then $V(\bar{X}) \to 0$ as $n \to \infty$, and $\bar{X}$ is a consistent (in mean square sense) estimator for μ.
In this case we say that the process $\{X_t\}_{t=1}^{\infty}$ is ergodic for the mean.

A sufficient condition for this result to hold is that $\rho_K \to 0$ as $K \to \infty$. This is so because $\rho_K \to 0$ as $K \to \infty$ implies that for any $\epsilon > 0$, we can choose an N such that

$$|\rho_K| < \frac{1}{4}\epsilon \text{ for all } K > N$$

Ergodicity For The Mean

Hence, for $n > (N+1)$, we have

$$
\begin{aligned}
\left|\frac{1}{n}\sum_{K=-(n-1)}^{n-1}\rho_K\right| &\le \frac{2}{n}\sum_{K=0}^{n-1}|\rho_K| \\
&\le \frac{2}{n}\sum_{K=0}^{N}|\rho_K| + \frac{2}{n}\sum_{K=N+1}^{n-1}|\rho_K| \\
&\le \frac{2}{n}\sum_{K=0}^{N}|\rho_K| + \frac{1}{2}\epsilon \\
&\le \epsilon
\end{aligned}
$$

where we choose an n large enough so that the first term in the next two last inequalities above is also less than $\frac{1}{2}\epsilon$.

Ergodicity For The Mean

This shows that when $\rho_K \to 0$ as $K \to \infty$ we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{K=-(n-1)}^{n-1} \rho_K = 0$$

which implies that in equation (1)

$$\lim_{n\to\infty} Var(\bar{X}) = 0$$

$$\boxed{\text{ERGODICITY FOR THE MEAN} : \rho_K \to 0}$$

Ergodicity For The Autocovariances

$$\boxed{\hat{\gamma}_K = \frac{1}{n}\sum_{t=1}^{n-K}(X_t - \bar{X})(X_{t+K} - \bar{X})}$$

When

$$\lim_{n\to\infty} \hat{\gamma}_K = \gamma_K \quad ??$$

A sufficient condition for $\hat{\gamma}_K$ to be mean square consistent and the process to be ergodic for the autocovariances is that

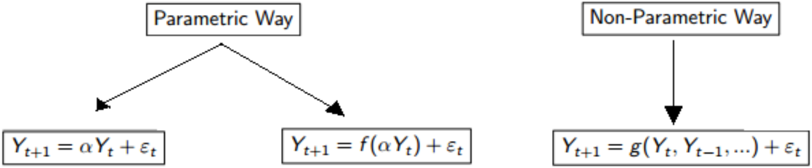
$$\boxed{\sum_{-\infty}^{\infty} |\gamma_i| < \infty}$$

WHERE ARE WE?

Problem: Forecast Y_{t+1} given some information set I_t at time t.
Solution:

$$\min_{\{f(X_{t-i}),i\geq 0\}} E\left[Y_{t+1} - f(X_{t-i})\right]^2,$$

where we obtain $f(X_t, X_{t-1}, ...) = E\left[Y_{t+1}|X_t, X_{t-1}, ...\right]$. We can model this conditional expectation in two ways:



Where $g(Y_t, Y_{t-1}, ...) = \hat{Y}_{t+1} = \frac{1}{n}\sum_{t=1}^{n} W_{n,t}(Y_t, Y_{t-1}, ...)$ and $W_{n,t}$ is a weight sequence depending on the past.
We are going to study PARAMETRIC LINEAR model in the time domain.