Some Basic Concepts

Lecture Notes prepared making partial use of:

- A.N. Shiryayev "Probability"
- P.C.B Phillips "Lecture Notes"
- K. Saxe "Beginning Functional Analysis"

Some Basic Concepts

Take $\Omega = [0, 1)$ and consider the problem of choosing points at random from this set.

For reasons of symmetry all the points are equiprobable.

The set $[0, 1)$ is uncountable, and if we suppose its probability is 1, then it follows that

$$p(w) = 0 \quad \forall w \in \Omega.$$

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

$$P(A) = \sum_{w \in A} p(w),$$

the above assignment of probabilities ($p(w)=0, w\in\Omega$) doesn't let us to define the probability that a point chosen at random from $[0,1)$ belongs to the set $A=[0,\frac{1}{2})$. At the same time it is intuitively clear that the probability should be $\frac{1}{2}$.

These remarks should 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>.

Def 1: Let Ω be a set of points w. A system $\mathcal{A}$ of subsets of Ω is called an algebra if

(a) $\Omega \in \mathcal{A}$

(b) $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$,

$$A \cap B \in \mathcal{A}$$

(c) $A \in \mathcal{A} \Rightarrow \bar{A} \in \mathcal{A}$

Def 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 disjoints sets A and B in $\mathcal{A}$.

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

Def 3: An ordered triple $(\Omega, \mathcal{A}, P)$, where

(a) Ω is a set of point w

(b) $\mathcal{A}$ is an algebra of subsets of Ω

(c) P is a finitely additive probability on $\mathcal{A}$.

is a PROBABILISTIC MODEL, in the extended sense.

It turns out, however, that this model is too broad to lead to a fruitful mathematical theory.

Def 4. A system $\mathcal{F}$ of subsets of Ω is a σ-algebra if it is an algebra and satisfies

(b*) If $A_n \in \mathcal{F}$, $n = 1, 2, \ldots$, then

$$\bigcup A_n \in \mathcal{F}, \quad \bigcap A_n \in \mathcal{F}.$$

Def 5. The space Ω together with a σ-algebra $\mathcal{F}$ of its subsets is a measurable space, and is denoted by $(\Omega, \mathcal{F})$.

Def 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 $\mathcal{A}$,

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

If a countably additive measure P on the algebra $\mathcal{A}$ satisfies $P(\Omega) = 1$, it is called a PROBABILITY MEASURE

Probability measures have the following properties:

- If ϕ is the empty set then

$$P(\phi) = 0$$

- 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 $\bigcup A_n \in \mathcal{A}$, then

$$P(A_1 \cup A_2 \cup \ldots) \leq P(A_1) + P(A_2) + \ldots$$

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

(a) Ω is a set of points w,
(b) $\mathcal{F}$ is a σ-algebra of subsets of Ω
(c) P is a probability on $\mathcal{F}$.

is called a PROBABILITY MODEL or

PROBABILITY SPACE

Examples of measurable spaces $(\Omega, \mathcal{F})$ which are extremely important

Example 1 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

Let $R = (-\infty, \infty)$ the real line, and

$$(a, b] = \{ x \in R : a < x \leq b \}.$$

Let $\mathcal{A}$ be the system of subsets of R which are finite sums of disjoint intervals of the form $(a, b]$:

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

This system is an algebra (easy to prove it) but not a σ-algebra:

$$\text{if } A_n = (0, 1 - \tfrac{1}{n}] \in \mathcal{A}, \text{ we have}$$

$$\bigcup_n A_n = (0, 1) \notin \mathcal{A}.$$

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 called Borel sets.

Example 2 $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$

$$\mathbb{R}^n = \underbrace{R \times \ldots \times R}_{n}$$

$$I = I_1 \times \ldots \times I_n \quad \text{where } I_k = (a_k, b_k].$$

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 3 $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$

$\mathbb{R}^\infty$ is the space of ordered sequences of numbers

$$x = (x_1, x_2, \ldots), \quad -\infty < x_k < \infty, \ k = 1, 2, \ldots$$

Consider the cylinder set

$$\mathcal{F}(I_1 \times \ldots \times I_n) = \{ x : x = (x_1, x_2, \ldots) \\ x_1 \in I_1, \ldots, x_n \in I_n \}$$

$\mathcal{B}(\mathbb{R}^\infty)$ is the smallest σ-algebras containing all the above sets.

Example 4 $(\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T))$ where T is an arbitrary set.

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

In general we shall be interested in the case when T is an uncountable subset of the real line, for instance $T = [0, \infty)$

The cylinder set considered is

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

and $B(\mathbb{R}^T)$ is the smallest σ-algebra corresponding to this cylinder set.

Note that $\mathcal{F}_{t_1, \dots, t_n}(I_1 \times \dots \times I_n)$ is just the set of functions that, at times $t_1, \dots, t_n$ "get through the windows" $I_1, \dots, I_n$ and at other times have arbitrary values.

Example 5 $(C, B(C))$. Let $T = [0, 1]$ and let C be the space of continuous functions $x = (x_t)$, $0 \le t \le 1$. This is a metric space with the metric

$$\rho(x, y) = \sup_{t \in T} |x_t - y_t|$$

We introduce this metric because with $\rho(x, y)$ the σ-algebra generated by the standard cylinders and by open sets coincide.

<u>Example 6</u> : $(D, \mathcal{B}(D))$ where 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 C, we can introduce a metric $d(x,y)$ on 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 Skorohod:

$$d(x,y) = \inf \{ \varepsilon > 0 : \exists \lambda \in \Lambda : \sup_t |x_t - y_{\lambda(t)}| + \sup_t |t - \lambda(t)| \le \varepsilon$$

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

Introducing of Probability Measures on Measurable Spaces

Example 1 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Let $P = P(A)$ be a probability measure defined on the Borel subsets A of the real line. Put

$$F(x) = P(-\infty, x] \quad x \in \mathbb{R}$$

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 2 $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$

$$F_n(x_1, \dots, x_n) = P((-\infty, x_1] \times \dots \times (-\infty, x_n])$$

$$= P(-\infty, \mathbf{x}]$$

where $\mathbf{x} = (x_1, \dots, x_n)$

$F_n(x_1, \dots, x_n)$ n-dimensional distribution function

Define

$$\Delta_{a_i, b_i} F_n(x_1, \dots, x_n) = F_n(x_1, \dots, x_{i-1}, b_i, x_{i+1}, \dots) - F_n(x_1, \dots, x_{i-1}, a_i, x_{i+1}, \dots)$$

where $a_i \leq b_i$

A simple calculation shows that

$$\Delta_{a_1 b_1} \cdots \Delta_{a_n b_n} F_n(x_1 \dots x_n) = P(a, b]$$

where $(a, b] = (a_1, b_1] \times \dots (a_n, b_n]$

In general $P(a, b] \neq F_n(b) - F_n(a)$.

Example 3 : $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$.

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

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

The sequence of probability measures $P_1, P_2, \dots$ defined respectively on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$, .. has the following evident consistency property:

for $n = 1, 2, \dots$ and $B \in \mathcal{B}(\mathbb{R}^n)$

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

Kolmogorov's Theorem on the Extension of Measures in $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$. Let $P_1, P_2, \dots$ 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

$$P(\mathcal{F}_n(B)) = P_n(B), \quad B \in \mathcal{B}(\mathbb{R}^n)$$
$$\text{for } n = 1, 2 \dots$$

Random Variables

Def 1: A real function $\xi = \xi(w)$ defined on $(\Omega, \mathcal{F})$ is an $\mathcal{F}$-measurable function, or a RANDOM VARIABLE, if

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

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

$$\xi^{-1}(B) \equiv \{w : \xi(w) \in B\}$$

is a measurable set in Ω.

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

$$\Omega = \{1, 2, 3\} \text{ and } F = \{\Omega, \emptyset, \{1,2\}, \{3\}\}$$

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

Def 2: A probability measure P_ξ on $(R, B(R))$ with $P_\xi(B) = P(w : \xi(w) \in B)$, $B \in \mathcal{B}(R)$ is called the probability distribution of ξ on $(R, \mathcal{B}(R))$

Def 3: The function

$$F_\xi(x) = P(w : \xi(w) \le x) \qquad x \in R$$

is called the distribution function of ξ.

Random Elements

Def 1: Let $(\Omega, \mathcal{F})$ and $(E, \mathcal{E})$ be measurable spaces. We say that a function $X = X(w)$, defined on Ω and taking values in E, is $\mathcal{F}/\mathcal{E}$-measurable, or is a random element (with values in E), if

$$\{w : X(w) \in B\} \in \mathcal{F}$$

for every $B \in \mathcal{E}$.

Some special cases:

(1) If $(E, \mathcal{E}) = (R, \mathcal{B}(R))$ random element = random variable

(2) Let $(E, \mathcal{E}) = (R^n, \mathcal{B}(R^n))$. Then a random element $X(w)$ is a "random point" in $\mathbb{R}^n$. If π_k is the projection of $\mathbb{R}^n$ on the kth coordinate axis, $X(w)$ can be represented in the form

$$X(w) = (\xi_1(w), \ldots, \xi_n(w)),$$

where

$$\xi_k = \pi_k \circ X.$$

Def 2: An ordered set $(\eta_1(w), \ldots, \eta_n(w))$ of random variables is called an n-dimensional <u>random vector</u>.

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 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.

Def 3 : 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 $T = \{1, 2, \ldots\}$ we call $X = (\xi_1, \xi_2, \ldots)$ a RANDOM PROCESS with discrete time or a random sequence

If $T = [0, 1], (-\infty, \infty), [0, \infty), \ldots$ 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 $(\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T))$

Def 4 : 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.

Def 5. 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

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

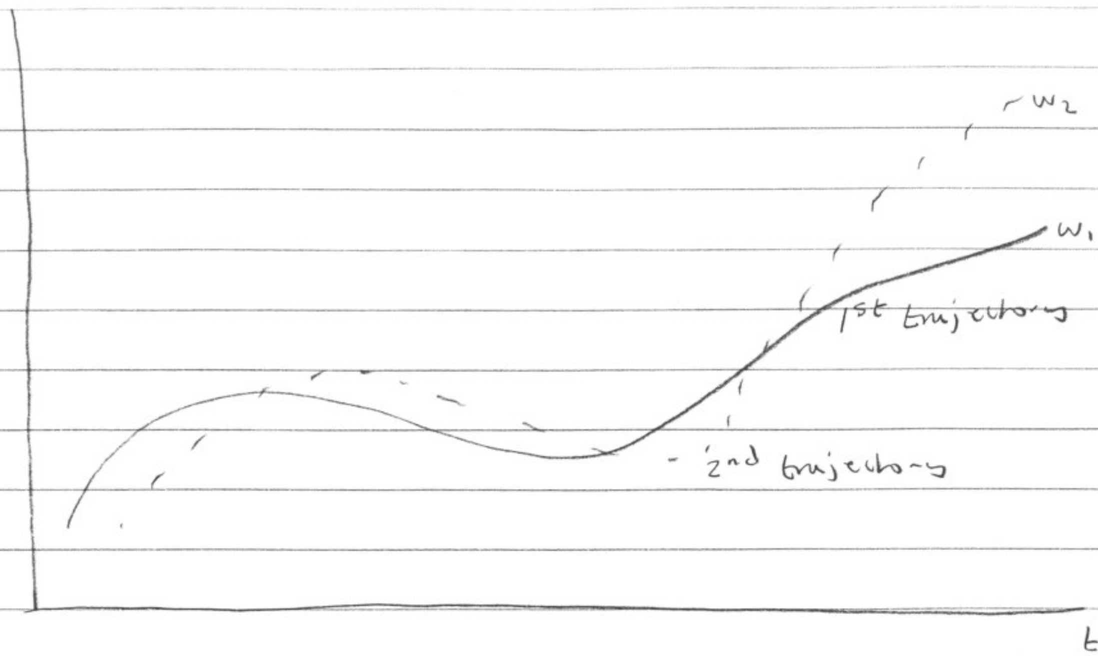
is called the probability distribution of X.

The probabilities

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

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

A TIME SERIES is a trajectory of an stochastic process



$X(w, t) = X_t(w)$

For fixed t, the function $X(\cdot, t)$ is a random variable

For fixed w, the function $X(w, \cdot)$ is a sample path of the stochastic process

$$(\underset{\underline{w}}{\Omega} \times \underset{\underline{t}}{T}) \longrightarrow R$$

$$(\underset{\underline{w}}{\Omega}, \mathcal{F}) \longrightarrow (\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T))$$

$X_1(w)$
$X_2(w)$
$\vdots$
$X_T(w)$

$(\mathbb{R}^\infty, \mathcal{B}_\infty, P) \longrightarrow X_1(\ldots X_{-1}, X_0, X_1 \ldots) = x_1$

$(\ldots x_{-1}, x_1, x_0, x_2, x_3) \times T \qquad X_2(\ldots \ldots) = x_2$

An Example of Stochastic Process

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

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

Therefore for a particular w, say $W_3 = \{3\}$, the realization or path would be $(10, 20, 30)$.

In this case Ω and T are finite. There are 6 realizations.

Think on dependence and stationarity

More examples of stochastic processes

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

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

(i) $\{X_t\}_0^\infty \equiv iid(0, \Omega)$

(ii) $\{X_t\}_0^\infty \equiv AR(1)$; ie $X_t = A X_{t-1} + u_t$

$\{u_t\} \equiv iid(0, \Omega)$

$X_0 \equiv$ initial condition

(iii) $\{X_t\}_0^\infty \equiv$ Random Walk (Ω); i.e $X_t = X_{t-1} + u_t$

$\{u_t\} \equiv iid(0, \Omega)$

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

(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) $W(t) \equiv N(0, t)$
(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) $X(t) = B(t) \equiv BM(\Omega)$

i.e

$$B(t) = \Omega^{\frac{1}{2}} W(t)$$

(iii) $X(t) \equiv BM(\mu, \Omega)$

i.e

$$X(t) = \mu t + B(t)$$

where

$$B(t) = BM(\Omega)$$

What is new?

- We want to make inference about $\mathbb{P}_x$, the probability law that governs the process. The evidence that we have is a single realization $x_1, x_2, \ldots, x_n \ldots$.

  The situation is even worse, we only have a $\boxed{\text{finite}}$ realization $\boxed{x_1, x_2, \ldots x_n}$

- Compare this situation with the classical case of iid samples where we have iid draws each of which is like an entire history of the time series

- We need to make assumptions

  (*) stationarity

  (*) weak dependence

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 $h: \Omega \longrightarrow \mathbb{R}^\infty \left(= \bigtimes_{-\infty}^{\infty} \mathbb{R}\right)$

$$h(\omega) = (\ldots, \underline{X}_{-1}(\omega), \underline{X}_0(\omega), \underline{X}_1(\omega), \ldots)$$
$$= (\ldots, x_{-1}, x_0, x_1, \ldots)$$
$$= x$$

(ii) $\mathcal{B}^\infty = \mathcal{B}(\mathbb{R}^\infty)$ Borel σ-field on $\mathbb{R}^\infty$ field generated by cylinder sets

$$\left(\bigtimes_{-\infty}^{r-1} \mathbb{R}\right)\left(\bigtimes_{r}^{s} \mathbb{R}\right)\left(\bigtimes_{s+1}^{\infty} \mathbb{R}\right)$$

(iii) $X_n : \mathbb{R}^\infty \longrightarrow \mathbb{R}$ <u>coordinate functions</u>

$$\boxed{X_n(x) = x_n}$$

Then $\{x \mid X_n(x) \leq a\} = \{x \mid \ldots x_{n-1} < \infty, x_n \leq a, x_{n+1} < \infty \ldots\}$

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

Thus

$$[\Omega, \mathcal{F}, \underline{P}] \sim [\mathbb{R}^\infty, \mathcal{B}^\infty, P]$$

Def Strict Stationarity

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

$$P((X_1, X_2, \ldots) \in B) = P((X_{k+1}, X_{k+2}, \ldots) \in B), \; B \in \mathcal{B}(\mathbb{R}^\infty)$$

Examples

(i) $\{X_n\} \equiv$ iid

(ii) $\{Y_n\}$, $\underbrace{Y_n = Y_n(X)}_{\text{measurable}}$; $X = \{X_n\}_{-\infty}^{\infty}$

All the measurable functions of iid sequences are strictly stationary (*)

e.g. $Y_n = \sum_{j=-\infty}^{\infty} a_j X_{n-j}$ $\quad \sum_{-\infty}^{\infty} a_j^2 < \infty$

Linear processes $\{X_n\} \equiv$ iid$(0,\sigma^2)$

Note that strict stationarity is STRONGER than the identical distribution assumption (of iid) since it requires all marginals to be identical.

(*) There is a Theorem that says that measurable functions of stationary sequence are stationary sequences.

Shift Operator

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

Backshift operator S

$$x = (\ldots, x_{-1}, x_0, x_1, \ldots)$$

$$Sx = (\ldots, x_0, x_1, x_2, \ldots)$$

$$S^2x = (\ldots, x_1, x_2, x_3 \ldots)$$

Induced operator U_s

Observe that if $\{X_n\}$ is a sequence on $(\mathbb{R}^\infty, \mathcal{B}^\infty, P)$ then

$$X_1(x) = x_1$$

$$X_2(x) = X_1(Sx) = x_2$$

$$\vdots$$

$$X_n(x) = X_1(S^{n-1}x) = x_n$$

inducing

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

defined by space of all real r.v.s defined on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$

$$U_s X(x) = X(Sx), \quad \text{any } X \in L_0(P)$$

and then

$$U_s X_n = X_{n+1} \quad \text{or} \quad X_n = U_s^{n-1} X_1$$

giving

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

<u>Definition</u> A transformation $T: \Omega \to \Omega$ is <u>measurable</u> if for any $A \in \mathcal{F}$,

$$T^{-1}A = \{w : Tw \in A\} \in \mathcal{F}$$

<u>Definition</u> A measurable transformation T is a measure-preserving transformation if for every $A \in \mathcal{F}$

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

<u>Examples</u>

① $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty), P)$ and stationary sequence $\{X_n\}_{-\infty}^{\infty}$

$S: \mathbb{R}^\infty \to \mathbb{R}^\infty$ backshift

Then S is measurable $(S^{-1}\mathcal{B}^\infty \subset \mathcal{B}^\infty)$
and
measure preserving $P(E) = P(S^{-1}E)$
if $\{X_n\}_{-\infty}^{\infty}$ is stationary

② $(\Omega, \mathcal{F}, P)$, $S = I$ $(Sw = w)$

Laws of Large numbers (LLN)

Kolmogorov SLLN $\{X_j\} \equiv iid$, $E|X_1| < \infty$

$$\frac{1}{n}\sum_1^n X_j \longrightarrow E(X_1) \quad a.s$$

Problem :- To extend this to temporally dependent data

- is it enough to require strict. stat

Example : $X_t = U_t + Z$

$\{U_t\} \equiv iid$ Uniform $[0,1]$

$Z \equiv N(0,1)$ indep of $\{U_t\}$

$$\bar{X} = \bar{U} + Z \xrightarrow{a.s} \frac{1}{2} + Z \quad \text{(random variable}$$

The problem is that there is too much dependence

$$Cov(X_t, X_{t+h}) = Cov(U_t + Z, U_{t+h} + Z$$

$$= V(Z) = 1$$

Ergodicity

Idea: When do temporal averages such as

$$\frac{1}{n}\sum X_i$$

Converge to ensemble (spatial) averages

$$E(X)$$

Some extra concepts :-

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

1) An event F in $\mathcal{F}$ is *invariant* if $F = S^{-1}F$

2) S is *ergodic* if for all invariant events F
$$P(F) = 0, 1 \quad \text{(ignorable or certain)}$$

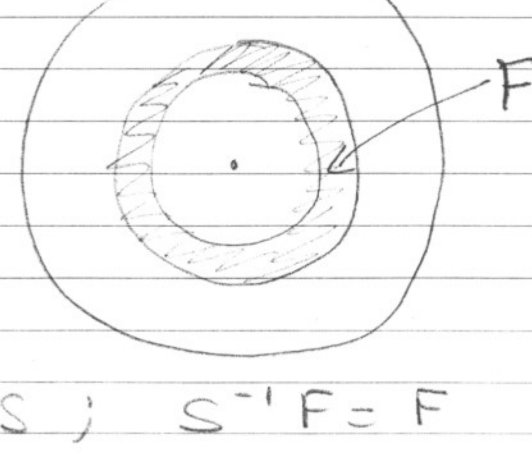
3) Strict stat process $\{X_t\}$ $(X_t = U_S^{t-1} X_1)$
is *ergodic* if S is ergodic

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

1.) $\Omega = \{ z \mid |z| \leq 1 \}$
unit disk



F = annulus
S = rotation by $\theta°$

$(Sz = az; \; a = e^{i\theta})$

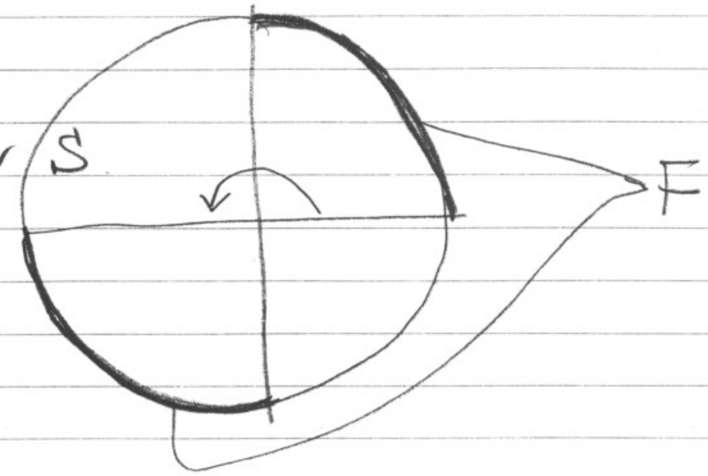
F is invariant under S; $S^{-1}F = F$

but $0 < P(F) < 1$, so S is not ergodic

2. $\Omega = \{ z \mid |z| = 1 \}$ unit circle

P = length (normalized so $P(\Omega) = 1$)

S = rotation by $180°$ ($Sz = az$ $a = e^{i\pi}$)



F is invariant under S

$S^{-1}F = F$

but $0 < P(F) < 1$

Problem: $a = e^{i\pi}$ is a root of unity $(e^{i\pi})^2 = 1$

S is m.p. but not ergodic

3. $\{X_t\}_{-\infty}^{\infty}$ $\quad X_t = \underbrace{U_t + Z}_{\text{independent}}$ $\quad U_t \equiv$ iid Uniform $[0,1]$

$Z \equiv N(0,1)$

$$F = \bigcap_{t=-\infty}^{\infty} \{ X_t(x) < 0 \}$$

$$= \{ x \mid \ldots X_{t-1}(x) < 0, X_t(x) < 0, X_{t+1}(x) < 0, \ldots \}$$

clearly

$$S^{-1}F = \{ x \mid \ldots X_{t-1}(Sx) < 0, X_t(Sx) < 0, X_{t+1}(Sx) < 0, \ldots$$

$$= \{ X \mid \ldots X_t(x) < 0, X_{t+1}(x) < 0, X_{t+1}(x) < 0, \ldots$$

and F is <u>invariant</u>. But $P(F) = P(Z < -1)$

and

$0 < P(F) < 1$

Ergodic Theorems

Theorem (Birhoff and Khinchin).

Let S be a m.p transformation and $X = X(w)$ a r. 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 invariant σ-field of $\mathcal{F}$ (σ-field of all events invariant under S).

If also S is ergodic then (P a.s)

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

Exaple (previously discussed)

$$X_t = U_t + Z = S^{t-1} U + Z$$

Z is an invariant r.v under shift operator S

$$\bar{X} \xrightarrow[a.s]{} E[X|J) = E(U|J) + E(Z|J)$$

$$= E(U) + Z$$

$$= \frac{1}{2} + Z$$

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

Thm (necessary & sufficient condition for ergodicity)

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

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

Mixing and Weak Dependence

Main Idea: Ergodicity of $\{X_n = S^{n-1}X\}_1^\infty$ is related to the capacity of S' to thoroughly mix the points of Ω. MIXING attempts to measure this property directly

Definition: A m.p. transformation $S': \Omega \to \Omega$ on $(\Omega, \mathcal{F}, P)$ is mixing if $\forall F, G \in \mathcal{F}$,

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

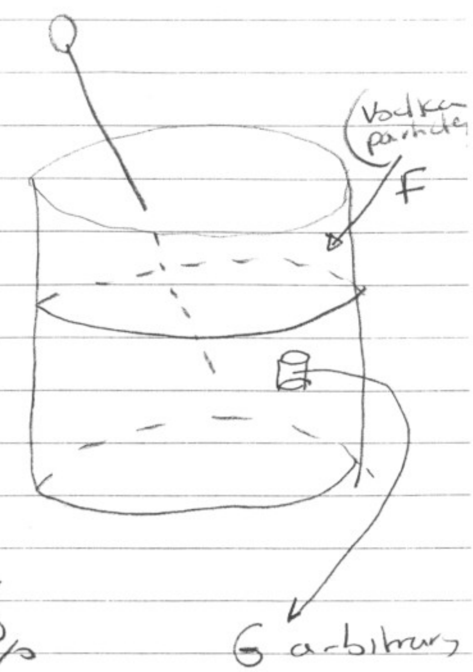
Example (Halmos martini example)

Ω = 90% vermouth, 10% vodka

S = action of swizzle stick on particles of Ω

P : Volume as % of Vol(Ω)

$\mathcal{F}$: 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

set of vodka particles in G at $t=n$ =

$$\{w \mid w \in F,\ S^n w \in G\} = F \cap S^{-n} G$$

mixing requires

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

i.e
$$\frac{\text{vol}\left[\text{vodka particles} \cap \text{particles in } G \text{ after } n\text{th swizzle}\right]}{\text{Vol}(G)}$$

$\longrightarrow$ Vol(F) = % of vodka particles.

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

$$P(X_n \in G, X_0 \in F) = P(X_0 \in S^{-n}G, X_0 \in F)$$
$$= P(X_0 \in F \cap S^{-n}G)$$
$$\longrightarrow P(F)\,P(G)$$

in other words X_0 & X_n are indep 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

$$K = \{ z \in \mathbb{C} \mid |z| = 1 \}$$
$$S : K \to K \qquad Sz = az \quad a = e^{i\theta}$$

(1) S is ergodic iff $e^{i\theta}$ is not a root of unity (iff $\theta \neq 2\pi/n$ for $n \in \mathbb{Z}$)

(2) S is ergodic (with $a = e^{i\theta}$, $\theta \neq 2\pi/n$ $\forall n \in \mathbb{Z}$) 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 strict stationarity	to	Weak stationarity
From ergodicity at a general level	to	Ergodicity for different moments

<u>Weak Stationarity</u>: The time series $\{X_t\}_{t\in T}$ with $T=\{0,\pm1,\pm2,\dots\}$ is said to be stationary if

$$\text{(i)}\quad E|X_t^2|<\infty \qquad \forall t\in T$$

$$\text{(ii)}\quad E(X_t)=\mu \qquad \forall t\in T$$

and

$$\text{(iii)}\quad \gamma_x(r,s)=\gamma_x(r+t,s+t) \qquad \forall r,s,t\in T$$

where

$$\gamma_x(r,s)=Cov(X_r,X_s)=E\left[(X_r-E(X_r))(X_s-E(X_s\right.$$

Remember that

<u>Def</u> (strict stationary): The time series $\{X_t, t\in T\}$ is said to be strictly stationary if the joint distribution of $(X_{t_1},\dots X_{t_k})'$ and $(X_{t_1+h},\dots X_{t_k+h})'$ are the same for all positive integers $\underline{k}$ and for all $t_1,t_2\dots t_k, h\in T$.

$$\begin{matrix}\text{Strict stationarity} \\ \left(\begin{matrix}\text{+ Existence of} \\ \text{moments}\end{matrix}\right)\end{matrix} \implies \begin{matrix}\text{Weak} \\ \text{stationarity}\end{matrix}$$

$$\Longleftarrow \text{(if gaussian)}$$

Examples

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 $\operatorname{cov}(X_{t+h}, X_t) = \gamma_x(h)$, $\{X_t\}$ is not stationary because it does not have a constant mean

② $S_t = X_1 + X_2 + \cdots + X_t$ $\quad X_1, X_2 \ldots \sim$ are iid $(0, \sigma^2)$

For $h > 0$

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

$$= \operatorname{Cov}\left(\sum_{i=1}^{t} X_i, \sum_{j=1}^{t} X_j\right)$$

$$= t\sigma^2$$

Ergodicity for the mean

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

$\{X_t\}_{t=1}^{\infty}$ is W. stationa
$E(X_t) = \mu$
$V(X_t) = \gamma_0$

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

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

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

$$= \frac{\gamma_0}{n^2} \sum_{k=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right) \rho_k \quad (*)$$

where we let $k = (t-s)$. Thus if

$$\lim_{n \to \infty} \left[\sum_{k=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right) \rho_k \right]$$

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 $\varepsilon > 0$, we can choose an N such that

$$|\rho_k| < \frac{1}{4} \varepsilon \text{ for all } k > N$$

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

$$\left| \frac{1}{n} \sum_{k=-(n-1)}^{n-1} \rho_k \right| \le \frac{2}{n} \sum_{k=0}^{n-1} |\rho_k|$$

$$\le \frac{2}{n} \sum_{k=0}^{N} |\rho_k| + \frac{2}{n} \sum_{k=N+1}^{n-1} |\rho_k|$$

$$\le \frac{2}{n} \sum_{k=0}^{N} |\rho_k| + \frac{1}{2} \varepsilon$$

$$\le \varepsilon$$

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

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 (*)

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

$$\boxed{\begin{array}{c}\text{ERGODICITY FOR THE MEAN} \\ \rho_k \to 0\end{array}}$$

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 $\lim_{n \to \infty} \hat{\gamma}_k = \gamma_k$??

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 $\underline{t}$.

Solution:
$$\underset{f(x_{t-i},\, i\geq 0)}{\text{Min}} \; E\left[y_{t+1} - f(x_{t-i})\right]^2$$

$$f(x_t, x_{t-1}, \ldots) = E\left[y_{t+1} \mid x_t, x_{t-1}, \ldots\right]$$

This conditional Expectation can be specified in a

$\boxed{\text{Parametric Way}}$ or in a $\boxed{\text{Non-Parametric way}}$

Linear

$\boxed{y_{t+1} = \alpha y_t + \varepsilon_t}$

Non-Linear

$\boxed{y_{t+1} = f(\alpha y_t) + \varepsilon_t}$

$$y_{t+1} = g(y_t, y_{t-1}, \ldots) + \varepsilon$$

$$\boxed{\hat{y}_{t+1} = \frac{1}{n} \sum_{t=1}^{n} W_{n,t}(y_t, y_{t-1}, \ldots)}$$

A weight sequence depending on the past

We are going to study PARAMETRIC LINEAR MODELS in the time domain