$(\Omega, \mathcal{F}, P)$

Ω (sample space): collection of all possible states of universe.

sample point (ω): leads to one particular outcome.

$\mathcal{F}$: collection of subsets of Ω (events) to which we want to assign probability. $\to$ (collection of events)

$\mathcal{F}$, a collection of subsets of Ω is a σ-field if

1. $\phi, \Omega \in \mathcal{F}$
2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
3. $A_i \in \mathcal{F}, i \geq 1 \Rightarrow \cup_i A_i \in \mathcal{F}$

A set $A \in \mathcal{F}$ is called an event, an $\mathcal{F}$-measurable set or simply a measurable set. The pair $(\Omega, \mathcal{F})$ is called a measurable space.

$P(\cdot) : \mathcal{F} \mapsto [0,1]$ such that

① $P(\Omega) = 1$

② if $\{A_i, i \geq 1\} \subset \mathcal{F}$ is a collection of disjoint sets, then

$$P(\cup_i A_i) = \sum_i P(A_i)$$

$\to$ Countable additivity (and not finite addivity)

Countable Set: can be put in 1-1 correspondence with positive integers.

also Cantor's diagonalisation argument.

Immediate Consequences

① $P(A^c) = 1 - P(A)$

② $A \subset B \Rightarrow P(A) \leq P(B)$ monotonicity.

③ $A_i \in \mathcal{F}$ $i \geq 1$

$P(\cup_i A_i) \leq \sum_i P(A_i)$ (Subadditivity)

④ $\lim_{n\to\infty} P(\bigcup_{i=1}^{n} A_i) = P(\bigcup_{i=1}^{\infty} A_i)$

⑤ $\lim_{n\to\infty} P(\bigcap_{i=1}^{n} A_i) = P(\bigcap_{i=1}^{\infty} A_i)$

P thus defined is an instance of measure, i.e. countable additive map from a σ field of subsets of set Ω to Real number which assigns the value 0 to the empty set.

Lemma Intersection of arbitrary families of σ-fields of subsets of Ω are σ-fields.

For any collection of $\mathcal{A}$ of subsets of Ω, the σ-field generated by $\mathcal{A}$ is intersection of all σ-fields of subsets of Ω that contain $\mathcal{A}$

Cauchy Sequence: A sequence $\{x_n\}$ is Cauchy if $\forall \varepsilon > 0, \exists M \in \mathbb{N}$ such that $\forall n, k \geq M$ $|x_n - x_k| < \varepsilon$

Unions of σ-fields need not be σ-fields.

Definition is not constructive i.e. you cannot generate all sets in the σ-field generated by $\mathcal{A}$ by beginning with $\mathcal{A}$ & applying countable intersection, or complement, unions.

Borel σ-field need not contain all subsets of ambient Ω

Given events $A_n, n \geq 1$ $\{A_n \text{ i.o.}\} = \{\omega \in \Omega : \omega \in A_n \text{ for } \infty\ n\}$

$A_i \in \mathcal{F}$

$\{A_n \text{ eventually}\} = \{\omega \in \Omega : \omega \in A_n\ \forall n \text{ sufficiently large}\}$

limsup A_n

$= \{A_n \text{ i.o.}\} = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$

Note $\bigcap_{m=n}^{\infty} \Leftrightarrow \forall m \geq n$

$\bigcup_{m=n}^{\infty} \Leftrightarrow \exists m \geq n$

liminf A_n

$= \{A_n \text{ eventually}\} = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m$

/almost always

$\{A_n \text{ i.o.}\}^c = \{A_n^c \text{ eventually}\} = \{A_n \text{ eventually not}\}$

Random Variable $X : (\Omega, \mathcal{F}, P) \to (E, \mathcal{E})$

- The image μ of P under X is probability measure $(E, \mathcal{E})$, called law of X. Denoted by $\mathcal{L}(X)$.
- The events $\{\omega \mid X(\omega) \in A\}$ for $A \in \mathcal{E}$ form a sub-σ-field of $\mathcal{F}$ called the σ-field generated by X & denoted by $\sigma(X)$.
- Given a family $X_\alpha, \alpha \in I$, of R.V. on $(\Omega, \mathcal{F}, P)$ taking values in measurable spaces $(E_\alpha, \mathcal{E}_\alpha), \alpha \in I$, the σ-field generated by $X_\alpha, \alpha \in I$ denoted by $\sigma(X_\alpha, \alpha \in I)$ is the smallest sub-σ-field wrt which they are all measurable.
- $(\Omega, \mathcal{F}, P) \xrightarrow{X} (S, \mathcal{S})$ is RV if $X^{-1}(A) = \{\omega : X(\omega) \in A\} \in \mathcal{F}\ \forall A \in \mathcal{S}$

also then X is a measurable map from $(\Omega, \mathcal{F})$ to $(S, \mathcal{S})$

- Corollary For a REAL R.V. X on $(\Omega, \mathcal{F}, P)$, $\exists$ Simple R.V. $\{X_n\}$ and elementary R.V. $\{X_n'\}$ on $(\Omega, \mathcal{F}, P)$ such that i) $X_n \to X$ and ii) $X_n' \to X$ uniformly, pointwise in both cases

$I_A(\omega) := \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise.} \end{cases}$

An elementary R.V. $X = \sum_i x_i I_{A_i}$ for countably ∞ collections $\{x_i\} \subset S$ & disjoint $\{A_i\} \subset \mathcal{F}$

A simple R.V. X take finitely many distinct values, say $x_1, \cdots, x_n \in S$. Letting $A_i = X^{-1}(\{x_i\}) \in \mathcal{F}$, A_i's disjoint wlog $X = \sum_{i=1}^{n} x_i I_{A_i}$

Proof: $X = X^+ - X^-$ $X^+ = \max(X, 0)$ $X^- = -\min(X, 0)$

Construct $X_n(\omega) = \sum_{k=0}^{n2^n - 1} \frac{k}{2^n} I\left\{\frac{k}{2^n} \leq X(\omega) < \frac{k+1}{2^n}\right\}$

Let $\sigma(X_\alpha, \alpha \in I)$ denote the smallest sub-σ-field of $\mathcal{F}$ wrt which all X_α are measurable. Let C denote the collection of countable subsets of I.

Lemma $\sigma(X_\alpha, \alpha \in I) = \bigcup_{J \in C} \sigma(X_\alpha, \alpha \in J)$

Theorem: X, Y are random variables taking values in $\mathbb{R}^d$, $\mathbb{R}^m$ respectively defined on $(\Omega, \mathcal{F}, P)$. If $\sigma(Y) \subset \sigma(X)$, then $Y = h(X)$ for some measurable $h : \mathbb{R}^d \mapsto \mathbb{R}^m$

Notational Abuse $P(X \in A) := P(\{\omega \in \Omega : X(\omega) \in A\})$

$\mathcal{F}$-measurable functions The f^n $X : \Omega \to \mathbb{R}$ defined on $(\Omega, \mathcal{F}, P)$ is called $\mathcal{F}$ measurable if $X^{-1} : \mathcal{B} \to \mathcal{F}$

$X^{-1}(A) = \{\omega : X(\omega) \in A\} \in \mathcal{F}$ for all $A \in \mathcal{B}(\mathbb{R})$

Consider of $\mathbb{R}^d$-valued R.V on probability space $(\Omega, \mathcal{F}, P)$. We then have $X^{-1}(A) \in \mathcal{F}\ \forall A \in \mathcal{B}$. Define probability measure μ on $(\mathbb{R}^d, \mathcal{B})$ by $\mu(A) = P(X^{-1}(A)) := P(X \in A)$, $A \in \mathcal{B}$

The probability law of a R.V.: Let $(\Omega, \mathcal{F}, P)$ be a probability space, & let $X : \Omega \to \mathbb{R}$ be a R.V.

a) For every Borel subset B of the real line ($B \in \mathcal{B}$) define $P_X(B) = P(X \in B)$

b) The resulting f^n $P_X : \mathcal{B} \to [0,1]$ is called the probability law of X.

c) The law P_X of X is a probability measure on $(\mathbb{R}, \mathcal{B})$

Definition is not dependent of choice of approximations $\{x^n\}$: Suppose

Measurable Space

A pair $(\Omega, \mathcal{F})$ where the former is a set and the latter a sigma algebra of subsets of Ω is called a Measurable space

Sub Sigma Algebra

A sub-sigma-algebra of $\mathcal{F}$ is a subset of $\mathcal{F}$ that is also a sigma algebra

$P(X^{-1}(B)) = P(\omega \mid X(\omega) \in B) = P(X \in B)$

If X is a non-negative R.V.

define $X^n(\omega)=\sum_{k=0}^{n2^n-1}\frac{k}{2^n} I\{\frac{k}{2^n}\le X(\omega)<\frac{k+1}{2^n}\}, \ n\ge 1$

define $E[X] := \lim_{n\uparrow\infty} E[X^n]$

If $E[X]$ is defined & finite, so is $E[|X|]=E[X^+]+E[X^-]$

Such R.V. are said to be integrable.

Properties of Expectation

① Linearity: $X,Y\in\mathbb{R}^d$ $a,b\in\mathbb{R}$ then $E[aX+bY]=aE[X]+bE[Y]$

② Positivity: $X\ge 0 \to E[X]\ge 0$

definition is not dependent of choice of approximations

Suppose $\{Y_n\}$ is another sequence of simple R.V with $0\le Y^n\uparrow X$ a.s.

Theorem $E[Y^n]\uparrow X$

Lemma: If $Z_n, n=1,2,\dots,\infty$ are simple R.V with $0\le Z_n\uparrow Z_\infty$ a.s. then $E[Z_n]\uparrow E[Z_\infty]$, need not be simple.

$E[X]=\sup\{E[\tilde X]: 0\le\tilde X\le X, \tilde X \text{ is simple}\}$

Useful Properties

① Change of variables formula: Let X be $\mathbb{R}^d$ valued integrable R.V. with law μ & $f:\mathbb{R}^d\to\mathbb{R}$ be a measurable f^n such that $f(X)$ is integrable.

$E[f(X)] := \int f(X(\omega))\,dP(\omega) = \int f(x)\,\mu(dx)$

a $f=I_A$ $A\in\mathcal{B}(\mathbb{R})$ $P(X^{-1}(A))=\mu(A)$

② R.V $X\ge 0$ $0<x\in\mathbb{R}$ $P(X\ge x)\le \frac{E[X]}{x}$

③ Jensen's Inequality: If $f:\mathbb{R}^d\to\mathbb{R}$ is convex then $E[f(X)]\ge f(E[X])$

Theorem If $X\ge 0$ a.s. $1\le p<\infty$

$E[X^p]=p\int_0^\infty t^{p-1}P(X\ge t)\,dt$

$\|X\|_p=(E[|X|^p])^{1/p}$, $\|X\|_\infty := \text{ess sup}\,\|X\|$

Minkowski Inequality. $\|X+Y\|_p\le\|X\|_p+\|Y\|_p$ $1\le p\le\infty$

Holder's Inequality. If $\frac{1}{p}+\frac{1}{q}=1$

$|E[XY]|\le\|X\|_p\|Y\|_q$

A normed space is a Banach space If it is complete wrt the metric induced by the norm

(all Cauchy sequence always converges)

Continuous Mapping Theorem

$X_n \xrightarrow{a.s./d} X \Rightarrow g(X_n)\xrightarrow{a.s./d} g(X)$ $\forall$ continuous g

Bounded Convergence Theorem

Let $X_n, n\ge 1$ be real R.V. satisfying $|X_n|\le K$ a.s. for $K<\infty$ and $X_n\to X$ a.s. Then $E[X_n]\to E[X]$

$\liminf_{n\uparrow\infty} x_n = \lim_{n\uparrow\infty}\inf_{m\ge n} x_m = \sup_{n\ge1}\inf_{m\ge n} x_m$

$\limsup_{n\uparrow\infty} x_n = \lim_{n\uparrow\infty}\sup_{m\ge n} x_m = \inf_{n\ge1}\sup_{m\ge n} x_m$

Clearly $\limsup_{n\uparrow\infty} x_n \ge \liminf_{n\uparrow\infty} x_n$

Fatou's Lemma

Let $X_n\ge 0$ be integrable real R.V. with $X_n\to X$ a.s.

Then $\liminf_{n\uparrow\infty} E[X_n]\ge E[X]$

Lemma Let $X\ge 0$ be integrable, Then $E[X\wedge N]\uparrow E[X]$ as $N\uparrow\infty$

Monotone Convergence Theorem

Let X_n $n\ge 1$ be integrable real R.V. with $X_n\uparrow X$ a.s.

The $E[X_n]\uparrow E[X]$

Dominated Convergence Theorem

Let $\{X_n\}, X, Y$ be Real R.V with $X_n\to X$ a.s. and $|X_n|\le Y$ with $E[Y]<\infty$ Then $E[X_n]\to E[X]$

Uniform Integrability

A family $\{X_\alpha,\alpha\in I\}$ of integrable R.V said to be uniformly integrable if $\lim_{a\uparrow\infty}\sup_\alpha E[|X_\alpha|\, I\{|X_\alpha|>a\}]=0$

1. $\{X_\alpha,\alpha\in I\}$ u.i $Y_1,\dots,Y_k$ integrable R.V., then $\{X_\alpha,\alpha\in I; Y_1,\dots,Y_k\}$ is u.i.

2. $\{X_\alpha,\alpha\in I\}$ u.i $\Rightarrow$ $\sup_\alpha E[|X_\alpha|]<\infty$

$\Leftrightarrow$ $\sup_\alpha E[|X_\alpha|]<\infty$ & $\lim_{P(A)\to 0}\sup_{\alpha\in I}\int_A|X_\alpha|\,dP=0$

Theorem

Let $\{X_n, n\ge 1\}$ be u.i and $X_n\to X_\infty$ a.s.

then $E[|X_n-X_\infty|]\to 0$

Lemma

If $E[|X_n-X|]\to 0$, then $\{X_n\}$ (or $X_n, n\ge 1; X$) are u.i.

Theorem

Let $\{X_n, 1\le n\le\infty\}$ be non-ve integrable R.V. with $X_n\to X_\infty$ a.s. Then

$E[|X_n-X_\infty|]\to 0 \Leftrightarrow E[X_n]\to E[X_\infty]$

Theorem (de La Vallee Poussin)

$\{X_\alpha,\alpha\in I\}$ is u.i. iff $\exists$ a $G:[0,\infty)\to[0,\infty)$ satisfying $\lim_{t\uparrow\infty}\frac{G(t)}{t}=\infty$ and $M:=\sup_{\alpha\in I}E[G(|X_\alpha|)]<\infty$

ex. $\sup_\alpha E[|X_\alpha|^a]<\infty$ for some $a>1$ suffices to ensure that $\{X_\alpha,\alpha\in I\}$ are u.i.

Metric Space: A metric space is a set X with metric $d: X\times X \to [0,\infty)$ such that $\forall x,y,z \in X$, d satisfies

1. Positive definite $d(x,y)\geq 0$, $d(x,y)=0 \Leftrightarrow x=y$
2. Symmetric $d(x,y)=d(y,x)$
3. Δ inequality $d(x,z)\leq d(x,y)+d(y,z)$

Pointwise Convergence: $n\in\mathbb{N}$, Let $f_n: S\to\mathbb{R}$, Let $f: S\to \mathbb{R}$.
$\{f_n\}$ converges pointwise to f if $\forall x\in S$ $\lim_{n\to\infty} f_n(x) = f(x)$

→ a sequence of continuous functions may not converge pointwise to a continuous f^n e.g. $f_n(x)=x^n$ on $[0,1]$

Uniform Convergence: f_n converges uniformly to f if $\forall \varepsilon>0$ $\exists M\in\mathbb{N}$ such that $\forall n\geq M, \forall x\in S$ $|f_n(x)-f(x)|<\varepsilon$

~~If $f_n: S\to\mathbb{R}$~~ $f_n\to f$ uniformly $\Rightarrow f_n\to f$ pointwise.

Bolzano Weirstrass Theorem: A bounded sequence always has a convergent subsequence.

Countable union of countable sets is countable

Limsup/Liminf: Let $\{x_n\}$ be a bounded subsequence. define if the limit exists $\limsup_{n\to\infty} x_n := \lim_{n\to\infty}(\sup\{x_k \mid k\geq n\})$

$\liminf_{n\to\infty} x_n = \lim_{n\to\infty}(\inf\{x_k \mid k\leq n\})$

Inverse images are preserved under set operations

$\mathbb{P}$ thus defined is an instance of the measure, a countably additive map from a σ-field of subsets of some set Ω to Real numbers which assign value '0' to the empty set

① Stochastic Convergence.

1. almost surely wp 1 $(X_n \xrightarrow{a.s.} X)$ if for each $\varepsilon>0$ $\mathbb{P}(|X_n - X|\geq\varepsilon\ i.o.)=0$
 $\Leftrightarrow \mathbb{P}(\lim_{n\to\infty} X_n(\omega)=X(\omega))=1$ $= 1-P(\exists\varepsilon>0, |Z_n - Z|\geq \varepsilon\ i.o.)$
2. in probability $(X_n \xrightarrow{P} X)$
 $\Leftrightarrow \lim_{n\to\infty}\mathbb{P}(\|X_n - X\|\geq\varepsilon)=0\ \forall\varepsilon>0$
3. in q^{th} mean
 $\Leftrightarrow E[\|X_n - X\|^q]\to 0$
 $q=1$ convergence in mean
 $q=2$ quadratic / mean square convergence.
4. in distribution/law (weak convergence)
 $\Leftrightarrow E[f(X_n)]\to E[f(X)]$ / $P(X_n\leq x)\to P(X\leq x)$
 3 examples to demonstrate this

$X_n \xrightarrow{a.s.} X \Rightarrow$, $X_n \xrightarrow{q^{th}} X \nearrow$ $X_n \xrightarrow{i.p} X \Rightarrow X_n \xrightarrow{D} X \Leftrightarrow [\phi_{X_n}(t)\to\phi_X(t), \forall t]$

1^{st} 2 implications assume that all R.V. be defined on same probability space.

power

Implies (hypothesis, Conclusion, Proposition)

P	Q	P ⇒ Q
T	T	T
T	F	F
F	T	T
F	F	T

An implication is true exactly when the IF part is false or the then-part is true.

If and only if

P	Q	P IFF Q
T	T	T
T	F	F
F	T	F
F	F	T

→ Convergence in law (or distribution) is not really convergence of R.V., but of their laws i.e. of probability measures. In particular $\{X_n\}$ need not be even define on same probability space.

② Borel Cantelli Lemma

Let $A_n\in\mathcal{F}$, $n\geq 1$, satisfy $\sum_n P(A_n)<\infty$. Then $P(A_n\ i.o.) = P(\bigcap_{n\geq 1}\bigcup_{m\geq n} A_m) = 0$

③ Theorem

$X_n\to X$ in probability iff any subsequence of $\{X_n\}$ has a further subsequence that converges a.s. to X.

→ For a.s to imply convergence in q^{th} mean, need precisely the uniform integrability of $|X_n|^q$

→ Convergence in law (or distribution) is NOT really a convergence of R.V., but of their laws i.e. of probability measures

→ In particular $\{X_n\}$ need not be even defined on the same probability space.

(ii) if $\sum P(A_n)=\infty$ and $\{A_n\}$ are independent, then $P(A_n\ i.o.)=1$

Convergence in law (or distribution) is NOT really a convergence of R.V., but of their laws, i.e. of probability measures. In definition $\{X_n\}$ need not even be defined on same probability space.

Metric on $\mathcal{P}(\mathbb{R}^d)$

$d(\mu,\nu) := \inf E[\|X-Y\|\wedge 1]$ where the infimum is over all pairs of R.V. (X,Y) such that $\mathcal{L}(X)=\mu$ & $\mathcal{L}(Y)=\nu$.

$\mu_n\to\mu$ in $P(\mathbb{R}^d)$ iff $d(\mu_n,\mu)\to 0$

Slutsky's Theorem

1. If c is constant $X_n\xrightarrow{d} c \Leftrightarrow X_n \xrightarrow{P} c$
2. If $X_n \xrightarrow{d} X$, $d(X_n, Y_n)\xrightarrow{P} 0$, then $Y_n \xrightarrow{d} X$
 $d(\nu_n,\mu)\leq d(\nu_n,\mu_n)+d(\mu_n,\mu) \leq 0+0$
3. If $X_n\xrightarrow{d} X$, $Y_n\xrightarrow{P} c$, then $\begin{pmatrix}X_n\\ Y_n\end{pmatrix}\xrightarrow{d}\begin{pmatrix}X\\ c\end{pmatrix}$

→ Consider space of all real R.V. on $(\Omega,\mathcal{F},P)$
$d(X,Y) := E[|X-Y|\wedge 1]$

Theorem: The metric $d(\cdot,\cdot)$ is complete & convergence wrt d coincides with convergence in probability.

4) $G \subset \mathbb{R}^d$ is an open set if $\forall x \in G, \exists r > 0$ such that $\{y : \|y - x\| < r\} \subset G$.

Finite intersections & arbitrary unions of open sets are open.

$F \subset \mathbb{R}^d$ is a closed set if F^c is open.

Finite unions & arbitrary intersection of closed sets are closed.

Interior ($\mathrm{int}(A)$)

Union of all open sets contained in A / Largest open set contained in A.

Closure ($\bar{A}$)

Smallest closed set containing A / intersection of all closed sets containing A.

$\mathrm{int}(A) \subset A \subset \bar{A}$

Boundary of A : ∂A $\quad \bar{A} \setminus \mathrm{int}(A)$

$C_b(\mathbb{R}^d)$: space of bounded continuous functions $\mathbb{R}^d \mapsto \mathbb{R}$

$\mathcal{P}(\mathbb{R}^d)$: space of probability measures on $\mathbb{R}^d$.

5) Portmanteau Theorem.

Let $\mu_n, n \geq 1$; $\mu \in \mathcal{P}(\mathbb{R}^d)$ Then following are equivalent.

0. $\mu_n \to \mu$ in $\mathcal{P}(\mathbb{R}^d)$
1. $\int f d\mu_n \to \int f d\mu \quad \forall f \in C_b(\mathbb{R}^d)$
2. $\int f d\mu_n \to \int f d\mu \quad \forall f \in C_b(\mathbb{R}^d)$ that are uniformly continuous wrt the Euclidean Norm.
3. $\liminf_{n \uparrow \infty} \mu_n(G) \geq \mu(G) \quad \forall$ open sets $G \subset \mathbb{R}^d$
4. $\limsup_{n \uparrow \infty} \mu_n(F) \leq \mu(F) \quad \forall$ closed sets $F \subset \mathbb{R}^d$
5. $\lim_{n \uparrow \infty} \mu_n(A) = \mu(A) \quad \forall$ sets $A \subset \mathcal{B}(\mathbb{R}^d)$ for which $\mu(\partial A) = 0$

Corollary In (3) above it suffices to consider bounded open G.

Corollary $\exists$ a countable family $\{f_i\} \subset C_b(\mathbb{R}^d)$ s.t.

$\int f_i d\mu_n \to \int f_i d\mu \quad \forall i \Rightarrow \mu_n \to \mu$

Any countable set $\{f_i, i \geq 1\} \subset C_b(\mathbb{R}^d)$ with the above property is called a Convergence Determining Class.

Fact: $d(\cdot,\cdot)$ is a metric, then $d(\cdot,\cdot) \wedge 1$, $\dfrac{d(\cdot,\cdot)}{1+d(\cdot,\cdot)}$ are equivalent bounded metric.

One such convenient metric $\rho(\mu,\nu) := \inf_{X \approx \mu, Y \approx \nu} E[\|X-Y\| \wedge 1]$

For $p \geq 1$ $\quad W_p(\mu,\nu) := \inf_{X \approx \mu, Y \approx \nu} \mathbb{E}[\|X-Y\|^p]^{1/p}$ defines a metric on $\{\mu \in \mathcal{P}(\mathbb{R}^d) : \int \|x\|^p d\mu(x) < \infty\}$

These are called Wasserstein-p metrics.

(these are complete metrics)

Consider Real R.V. X, whose distribution f^n F is continuous & strictly increasing. If $U \sim \mathrm{uni}[0,1]$

$\therefore X = F^{-1}(U)$

6) Skorohod's Theorem If $\mu_n \to \mu_\infty$ in $\mathcal{P}(\mathbb{R}^d)$, then $\exists$ $\mathbb{R}^d$-valued R.V. X_n $1 \leq n \leq \infty$ s.t. law of X_n is μ_n for $1 \leq n \leq \infty$ and $X_n \to X_\infty$ a.s.

7) Scheffe's Theorem Total Variation Convergence.

Suppose $\mu_n \in \mathcal{P}(\mathbb{R}^d)$ $1 \leq n \leq \infty$, have densities $p_n(\cdot)$ resp. (i.e. $\int f(x) d\mu_n(x) = \int f(x) p_n(x) dx \; \forall x$) and $p_n(x) \to p_\infty(x)$ $\forall x$ outside a set of zero Lebesgue measure, then

$\int |p_n(x) - p_\infty(x)| dx \to 0$

8) Characteristic Function of $\mu \in \mathcal{P}(\mathbb{R}^d)$ is

$t \in \mathbb{R}^d \mapsto \varphi_\mu(t) := \int e^{i\langle t, x\rangle} d\mu(x)$ i.e. F.T. of μ

$= E[e^{i\langle t, X\rangle}]$

1. $\varphi_\mu(0) = 1$, $|\varphi_\mu(t)| \leq 1$, $\varphi_\mu(-t) = \overline{\varphi_\mu(t)}$
2. φ_μ is uniformly continuous
3. φ_μ is positive definite. $\sum_{i,j} c_i \varphi(t_i - t_j) \bar{c}_j \geq 0$
4. $\nu :=$ law of $aX + b$

$\varphi_\nu(t) = e^{i\langle t, b\rangle} \varphi_\mu(at)$

Bochner's Theorem: Any $\varphi : \mathbb{R}^d \mapsto \mathbb{C}$ satisfying first 3 conditions is necessarily a characteristic f^n of some $\mu \in \mathcal{P}(\mathbb{R}^d)$

9) Lemma

If $\varphi_\mu = \varphi_\nu$ for $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, then $\mu = \nu$

10) Levy Continuity Theorem (standard tool for establishing convergence in law)

If $\mu_n \to \mu$ in $\mathcal{P}(\mathbb{R}^d)$ then $\varphi_{\mu_n} \to \varphi_\mu$.

Conversely if $\varphi_{\mu_n} \to$ some $\varphi : \mathbb{R}^d \mapsto \mathbb{C}$ and φ is continuous at the origin, then $\varphi = \varphi_\mu$ for some $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\mu_n \to \mu$ in $\mathcal{P}(\mathbb{R}^d)$

Skorohod's Theorem: Suppose $X_n \xrightarrow{d} X$. Then, there exists a probability space & R.V Y, Y_n defined on that space with following properties

a) For every n X_n, Y_n have same CDF ; similarly X and Y have same CDF

b) $Y_n \xrightarrow{a.s.} Y$

Elicit

Independence of arbitrary collection of events:
Any finite subcollection thereof be independent.

→ A finite collection $A_i \in \mathcal{F}$ is independent if

$P(\cap_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i)$

→ Pairwise independence

→ $\mathbb{R}^d$-valued R.V. are independent if $\forall$measurable $f_i : \mathbb{R}^d \to \mathbb{R}$

$E[\prod_{i=1}^{n} f_i(X_i)] = \prod_{i=1}^{n} \mathbb{E}[f_i(X_i)]$,

for $f_i = I_{A_i}$: $P(X_i \in A_i, 1 \le i \le n) = \prod_{i=1}^{n} P(X_i \in A_i) \ \forall A_i \in B(\mathbb{R}^D)$

1 Measure Theoretic Definition: Product Measures

$\mu \in P(\mathbb{R}^d)^n$: joint law of $(X_1, X_2, \cdots, X_n)$ and $\mu_i \in P(\mathbb{R}^d)$ law of X_i (called the i^{th} marginal)

$\mu(dx_1 \cdots dx_n) = \prod_{i=1}^{n} \mu_i(dx_i)$

$\mu(\prod_{i=1}^{n} A_i) = \prod_{i=1}^{n} \mu_i(A_i) \ \forall$choices of $A_i \in B(\mathbb{R})$

2 Kolmogorov's Zero-One Law

Let $\{X_n\}$ be ind. $\mathbb{R}^d$-valued R.V.

Tail Sigma Field: $\sigma^* := \cap_{n \ge 1} \sigma(X_n, X_{n+1}, \cdots)$. If $A \in \sigma^*$ it is called a Tail Event.

$A \in \sigma^* \Rightarrow P(A) = 0$ or 1

3 Borel-Cantelli Lemma (Continued) If $\{A_n, n \ge 1\}$ are independent, then $\sum_n P(A_n) = \infty \Rightarrow \boxed{P(A_n \text{ i.o.}) = 1}$

4 Levy's Result $\quad S_0 = 0$

Let $\{X_n\}$ be independent R.V. $S_n = \sum_{m=1}^{n} X_m$

S_n converges in law $\Leftrightarrow$ converges in probability $\Leftrightarrow$ converges a.s.

→ Let $\{X_n, n \ge 1\}$ be Real integrable R.V. with mean Zero.

WLLN Let $\{X_n\}$ be zero mean, pairwise independent and u.i. Then $\frac{S_n}{n} \to 0$ in probability.

SLLN Let $\{X_n\}$ be zero mean, pairwise independent with $\sup_n \mathbb{E}[|X_n|^2] \le K < \infty$ for some K.
Then $\frac{S_n}{n} \to 0$ a.s.

For pairwise independent and identically distributed R.V., a result of Etemadi establishes SLLN under weaker requirement that X_n be integrable.

Corollary If $\{X_n\}$ are pairwise independent & identically distributed with $E[X_1] = \infty$ then $\frac{S_n}{n} \to \infty$ a.s.

Glivenko-Cantelli

Let $\{X_n\}$ be i.i.d. with common distribution $f^n F$.

Empirical Distribution F $\quad F_n(x) = \frac{1}{n} \sum_{m=1}^{n} I\{X_n \le x\}$

Then $\sup_x |F_n(x) - F(x)| \to 0$ a.s.

Uniform LLN $\quad S_n := \sum_{m=1}^{n} (f(X_m) - \mathbb{E}[f(X_m)])$,

$\{X_m\}$ i.i.d. zero mean & f bounded continuous.

Uniform LLN gives conditions on families of $\mathcal{A}$ of f^n for which $\lim_{n\uparrow\infty} \sup_{f \in \mathcal{A}} |\frac{S_n}{n}| \to 0$ a.s.
justifies minimization of empirical loss $\frac{1}{n}\sum_{m=1}^{n} L(X_m, Y_m, \theta)$ in place of $\mathbb{E}[L(X_n, Y_n)]$ where $\{(X_n, Y_n)\}$ iid i/p o/p pairs

Central Limit Theorem $\quad \frac{S_n - na}{\sigma\sqrt{n}} \to \mathcal{N}(0,1)$ in law

Lindberg & Feller:
Let $\{X_n\}$ be independent zero mean with $E[X_n^2] = \sigma_n^2 \ \forall n$

$s_n^2 = \sum_{m=1}^{n} \sigma_m^2$. Lindberg Condition:

$\sum_{m=1}^{n} \int_{\{|X_m| > \varepsilon s_n\}} X_m^2 \, dP = o(s_n^2) \ \forall \varepsilon > 0$

Lindberg's Condition $\Leftrightarrow \frac{S_n}{s_n} \to \mathcal{N}(0,1)$ in law & $\frac{\sigma_n}{s_n} \to 0$

Law of Iterated Logarithms

Suppose for some $\varepsilon \in (0,1)$, $C > 0$,

$\sum_{m=1}^{n} \frac{\mathbb{E}[|X_m|^3]}{s_n^3} \le \frac{C}{s_n^{1+\varepsilon}}$. Then

$\limsup_{n\uparrow\infty} \frac{S_n}{\sqrt{2 s_n^2 \log\log s_n}} = 1$ a.s. $\quad \liminf_{n\uparrow\infty} \frac{S_n}{}$

Cramer's Theorem

Let $\{X_n\}$ be independent and identically distributed with mean a. define $I : \mathbb{R} \mapsto \mathbb{R} \cup \{\infty\}$

$I(x) := \sup_\theta (\theta x - \log \mathbb{E}[e^{\theta X}])$

1. $I(\theta) \ge 0$
2. Pointwise supremum $I(x) \ni$ convex.
3. $I(a) = 0 = \min_\theta I(\theta)$
4. $I(\cdot)$ is decreasing on $(-\infty, a]$ & increasing on $[a, \infty)$

Theorem Suppose $I(\theta)<\infty, \forall \theta \in R$. Then for any $B \subset B(R)$

$$-\inf_{B^\circ} I \le \liminf_{n\uparrow\infty} \frac{1}{n}\log P\left(\frac{S_n}{n}\in B\right) \le \limsup_{n\uparrow\infty} \frac{1}{n}\log P\left(\frac{S_n}{n}\in B\right) \le -\inf_{\bar{B}} I$$

If $a\notin \bar{B}$, the $P(\frac{S_n}{n}\in B)$ decays atleast at rate $\inf_{\bar{B}} I$.

The Large Deviation Principle (S.R.S Vardhan)

A sequence of probability measures $\mu_\varepsilon \in P(S)$, $\varepsilon>0$, satisfies the large deviations principle with rate f^n I if for all Borel sets in S,

LDP with rate $1/\varepsilon$

$$-\inf_{B^\circ} I \le \liminf_{\varepsilon\downarrow 0} \varepsilon \log \mu_\varepsilon(B) \le \limsup_{\varepsilon\downarrow 0} \varepsilon \log \mu_\varepsilon(B) \le -\inf_{\bar{B}} I$$

Laplace Vardhan Principle

$$\lim_{\varepsilon\to 0} \varepsilon \log E\left[e^{f(X_\varepsilon)/\varepsilon}\right] = \sup_x f(x) - I(x)$$

X_ε are R.V. obeying LDP with rate $\frac{1}{\varepsilon}$ & rate f^n J

We can define a R.V. $P(A|G)$ where $G :=$ the σ-field generated by $\{B_1,\dots,B_m\}$ as

$$P(A|G) = \sum_{i=1}^{m} \frac{P(A\cap B_i)}{P(B_i)} I_{B_i}$$

Conditional Expectation

Partition $D := \{B_1,\dots,B_m\} \subset F$ of Ω with $P(B_i)>0$

$G := \sigma$ field generated by $\{B_1,\dots,B_m\}$

Definition: Cond^n expectation of X given G is defined as the G-measurable R.V. denoted $E[X|G]$ that satisfies

$$\int_C X\,dP = \int_C E[X|G]\,dP \quad \forall C\in G$$

Projection Theorem

Let H be a Hilbert space over reals i.e. a vector space with an inner product $\langle\cdot,\cdot\rangle$ such that the associated norm $x\in H \mapsto \|x\| := \sqrt{\langle x,x\rangle}$ is complete

Theorem

Given a closed subspace G of H and $x\in H$ $\exists$ an unique $x^*\in G$ such that

$$\|x-x^*\| = \min_{y\in G} \|x-y\|$$

Corollary

For x, x^* as above $\langle x-x^*, y-x^*\rangle = 0 \ \forall y\in G$.

$H = L_2(\Omega, F, P)$: space of RV X on (Ω,F,P) that are square integrable i.e. $E[X^2]<\infty$

Theorem For an integrable R.V. X on (Ω,F,P) and a sub-σ-field $G\in F$, $\exists$ an a.s. unique integrable G-measurable R.V. $E[X|G]$ such that

$$\int_C X\,dP = \int_C E[X|G]\,dP \quad \forall C\in G$$

Radon-Nikodym Theorem Suppose μ is a positive measure on (Ω,F) and ν a signed measure on (Ω,F) such that ν is absolutely continuous wrt μ ($\nu \ll \mu$) $\mu(A)=0, A\in F \Rightarrow \nu(A)=0$. Then $\exists$ an F-measurable f^n $Z:\Omega\mapsto R$ such that

$$\int f\,d\nu = \int f Z\,d\mu \quad \forall F\text{-measurable } f$$

The measure Q is σ-finite if $\exists$ a countable covering $\{A_i\}$ of Ω with F-measurable sets such that $Q(A_i)<\infty$ for each i

Radon-Nikodym derivative of P wrt Q is an $\mathcal{F}$-measurable R.V. $\frac{dP}{dQ}: \Omega \to [0,\infty)$ such that

$\int I_A dP = \int I_A \frac{dP}{dQ} dQ$ $\quad P(A) = \int_A \frac{dP}{dQ} dQ \quad \forall A \in \mathcal{F}$ 2.11

For any X-~~measurable~~ P integrable R.V. $\int X dP = \int X \frac{dP}{dQ} dQ$.

Theorem 2.13: Let P, Q be measures on a common measurable space $(\Omega, \mathcal{F})$ and assume Q is σ-finite. Then dP/dQ exists iff P is absolutely continuous wrt Q $(P \ll Q)$. dP/dQ is uniquely defined upto a Q-null set so that for any f_1, f_2 satisfying 2.11 $f_1 = f_2$ holds Q-almost surely.

Properties of Conditional Expectation.

Fix $(\Omega, \mathcal{F}, P)$ & let $\mathcal{G}' \subseteq \mathcal{G} \subset \mathcal{F}$ be sub σ fields

X, Y, Z, W: integrable R.V. on $(\Omega, \mathcal{F}, P)$.

1. Monotonicity $X \geq Y$ a.s. $\Rightarrow E[X|\mathcal{G}] \geq E[Y|\mathcal{G}]$ a.s.
2. Linearity If X, Y are $\mathcal{G}$-measurable
$E[XW + ZY|\mathcal{G}] = X E[W|\mathcal{G}] + Y E[Z|\mathcal{G}]$
3. Filtering (Tower Property)
$E[E[X|\mathcal{G}]|\mathcal{G}'] = E[X|\mathcal{G}']$

Conditional Expectation is like a random integral. $\mathcal{G}$-measurable Random Variables are treated as const.

Theorem

Given an integrable random variable X on $(\Omega, \mathcal{F}, P)$ and a sub σ field $\mathcal{G} \subset \mathcal{F}$, $\exists$ a $P(\mathbb{R}^d)$-valued random variable μ such that

$E[f(X)|\mathcal{G}] = \int f d\mu$ $\forall$ bounded measurable $f: \mathbb{R}^d \mapsto \mathbb{R}$

To make the dependence of μ on $\mathcal{G}$ explicit use $\mu(\cdot|\mathcal{G})$

$\mu(A|\mathcal{G}) = P(X \in A|\mathcal{G})$ a.s. and $\int f d\mu(x|\mathcal{G}) = E[f(X)|\mathcal{G}]$

$\mathcal{F}_1 \vee \mathcal{F}_2$: smallest σ field containing $\mathcal{F}_1, \mathcal{F}_2$

Conditional Independence.

We say that r.v. X_i taking values in $\mathbb{R}^{d_i}$ are conditionally independent given a sub-σ-field $\mathcal{G}$ of $\mathcal{F}$ if any of following 3 hol

1. $P(X_i \in A_i, 1 \leq i \leq n|\mathcal{G}) = \prod_{i=1}^n P(X_i \in A_i|\mathcal{G})$
2. $E[\prod_{i=1}^n f_i(X_i)|\mathcal{G}] = \prod_{i=1}^n E[f_i(X_i)|\mathcal{G}]$
$\forall$ bounded measurable $f_i: \mathbb{R}^{d_i} \mapsto \mathbb{R}$.

Theorem

Let $\mathcal{F}_i$, $i = 1, 2, 3$ be sub-σ-fields of $\mathcal{F}$. Then following are equivalent

1. $\mathcal{F}_1, \mathcal{F}_3$ are conditionally independent given $\mathcal{F}_2$
2. For every integrable $\mathcal{F}_3$-measurable Y
$E[Y|\mathcal{F}_1 \vee \mathcal{F}_2] = E[Y|\mathcal{F}_2]$
3. For every integrable $\mathcal{F}_1$-measurable Z
$E[Z|\mathcal{F}_2 \vee \mathcal{F}_3] = E[Z|\mathcal{F}_2]$

classmate

Date
Page

Martingales

1. Markov Chains: Replace independent by conditionally independent

2. Martingales: " " " 'uncorrelated'

'Family of Increasing sub-σ-fields' Filtered σ-fields/ Filtration

$\{\mathcal{F}_n ; n \geq 0\}$ of $\mathcal{F}$', i.e. $\mathcal{F}_n \subset \mathcal{F}_{n+1}$

Typically, $\mathcal{F}_n := \sigma(X_m, m \leq n)$

A sequence of R.V. (DTSP) X_n is adapted to $\{\mathcal{F}_n\}$ if for each n, X_n is $\mathcal{F}_n$ measurable.

A sequence of integrable R.V. $\Delta M_n, n \geq 0$ adapted to filtration $\{\mathcal{F}_n\}$ is called MDS wrt $\{\mathcal{F}_n\}$ if

$E[\Delta M_{n+1} | \mathcal{F}_n] = 0 \ \forall n$

$\Leftrightarrow E[\Delta M_{n+1} Y] = 0 \ \forall$ bounded $\mathcal{F}_n$ measurable R.V. Y.

- If M_n (integrable R.V) adapted to $\{\mathcal{F}_n\}$.

$(M_n, \mathcal{F}_n)$ is Submartingale if $E[M_{n+1} | \mathcal{F}_n] \geq M_n \ \forall n$

$\Leftrightarrow E[M_n | \mathcal{F}_m] \geq M_m \ \forall n \geq m$

Supermartingale if $E[M_{n+1} | \mathcal{F}_n] \leq M_n$

Martingale. $E[M_{n+1} | \mathcal{F}_n] = M_n$

A martingale is both a submartingale and supermartingale $\Leftrightarrow \Delta M_n := M_n - M_{n-1}$ is MDS

Ex 2. Let Y be an Integrable RV. Then for any filtration $\{\mathcal{F}_n\}$, $M_n = E[Y | \mathcal{F}_n]$ is martingale wrt $\{\mathcal{F}_n\}$

3. Q: probability measure on $(\Omega, \mathcal{F})$ defined by

$Q(A) = \int_A \Lambda_n \, dP$, $A \in \mathcal{F}_n$, for some $\Lambda_n \geq 0$ a.s.

with $E[\Lambda_n] = 1$, $n \geq 0$. For $n > m$, $A \in \mathcal{F}_m \Rightarrow A \in \mathcal{F}_n$

so $\int_A \Lambda_m \, dP = \int_A \Lambda_n \, dP = \int_A E[\Lambda_n | \mathcal{F}_m] \, dP$

From a.s. uniqueness of conditional expectation $E[\Lambda_n | \mathcal{F}_m] = \Lambda_m$

Thus $(\Lambda_n, \mathcal{F}_n)$ is a martingale.

A process $\{Y_n\}$ is said to be Predictable if Y_n is adapted to $\mathcal{F}_{n-1} \ \forall n$.

ex. $Z_n := \sum_{m=0}^{n} E[Y_m | \mathcal{F}_{m-1}]$, $n \geq 0$ is predictable.

$Y_n = Z_n + M_n$ splits $\{Y_n\}$ into a predictable part ($\approx$ trend or signal) & Martingale ($\approx$ fluctuation or noise)

For convex $f : \mathbb{R} \to \mathbb{R}$

Conditional Jensen's Inequality $E[f(X) | \mathcal{G}] \geq f(E[X | \mathcal{G}])$

$\mathcal{G} \subset \mathcal{F}$ a sub σ field

6. $(M_n, \mathcal{F}_n)$ is a Martingale $\Rightarrow \begin{cases} (f(M_n), \mathcal{F}_n) & \text{is submartingale if } f \text{ convex} \\ (f(M_n), \mathcal{F}_n) & \text{is supermartingale if } f \text{ concave.} \end{cases}$

assuming $f(M_m)$ are integrable

7. Sum of a martingale & a non-decreasing predictable process is a submartingale.

Martingale + Non-increasing predictable process = Supermartingale.

Converse is also true.

8. If $\{Z_n\}$ is a bounded $\{\mathcal{F}_n\}$ adapted process & $(M_n, \mathcal{F}_n)$ a martingale, then $X_n = \sum_{m=0}^{n-1} Z_m \Delta M_{m+1}$, is $\{\mathcal{F}_n\}$ martingale

Stopping Time

Let $\mathcal{F}_n \subset \mathcal{F}_\infty$. A R.V. $\tau \in \{0, 1, 2, \cdots, \infty\}$ is a stopping time if for $0 \leq n \leq \infty$ $\{\tau \leq n\} \in \mathcal{F}_n$

$\equiv \{\tau = n\} \in \mathcal{F}_n$, $\{\tau = n\} = \{\tau \leq n\} \cap \{\tau \leq n-1\}^c$

Stopped σ-field $\mathcal{F}_\tau := \{A \in \mathcal{F} : A \cap \{\tau \leq n\} \in \mathcal{F}_n \ \forall n\}$

Let S, T be stopping times wrt $\{\mathcal{F}_n\}$

5. If $S \leq T$ then $\mathcal{F}_S \subset \mathcal{F}_T$

8. If $Y_n, 1 \leq n \leq \infty$ is adapted to $\{\mathcal{F}_n\}$, then Y_T is $\mathcal{F}_T$-measurable.

9. For $\{Y_n\}$ as above $E[Y_T | \mathcal{F}_S]$ is $\mathcal{F}_{S \wedge T}$ measurable.

Doob's Optional Sampling Theorem.

$\{M_n\}$ I.R.V. adapted to Filtration $\{\mathcal{F}_n\}$, Following equivalent.

1. $(M_n, \mathcal{F}_n)$ $n \geq 0$ is a submartingale

2. If T, S are bounded stopping times with $T \geq S$ a.s. then $E[M_T] \geq E[M_S]$ a.s.

3. If T, S are stopping times with T bounded a.s. then $E[M_T | \mathcal{F}_S] \geq M_{S \wedge T}$ a.s.

Corollary Let $(M_n, \mathcal{F}_n)$ be a sub-martingale and $\{T_n\}$ increasing bounded stopping times. Then $(M_{T_n}, \mathcal{F}_{T_n})$ is a Sub-Martingale.

In practice, we need to extend these results to possibly unbounded stopping times. This has to be justified by using Truncated Stopping time, followed by a limiting argument, justified using conditional versions of MCT, DCT etc

→ **Doob's Maximal and Minimal Inequalities**

Let $(X_n, \mathcal{F}_n), n \geq 0$ be a SubMartingale, $\forall \lambda > 0$

$$\lambda P(\max_{n \leq N} X_n \geq \lambda) \leq \int_{\{\max_{n \leq N} X_n \geq \lambda\}} X_N \, dP \leq E[X_N^+]$$

$$\lambda P(\min_{n \leq N} X_n \leq -\lambda) \leq -E[X_0] + \int_{\{\min_{n \leq N}, X_N \geq -\lambda\}} X_N \, dP$$

$$\leq -E[X_0] + E[X_N^+]$$

Corollary Let $(X_n, \mathcal{F}_n)$ be a martingale or a non-negative sub-martingale. Let $E[|X_n|^p] < \infty$ for some $p \in (1, \infty)$. Then

$$E[|\max_{n \leq N} X_n|^p] \leq \left(\frac{p}{p-1}\right)^p E[|X_N|^p]$$

for $p = 1$, one has

$$E[|\max_{n \leq N} X_n|] \leq \frac{e}{e-1}\left(1 + E[|X_N| \ln^+ |X_N|]\right)$$

→ **Doob Decomposition**

A submartingale $(X_n, \mathcal{F}_n), n \geq 0$ can be written as

$$Z_n \cancel{X_n} = M_n + A_n \quad , n \geq 0$$

1. $(M_n, \mathcal{F}_n)$ is a Zero-mean Martingale.
2. $\{A_n\}$ is an increasing process ($A_i \leq A_{i+1}$ a.s. $\forall i$) & $\{A_n\}$ is adapted to $\{\mathcal{F}_{n-1}\}$; where $\mathcal{F}_{-1} = \{\phi, \Omega\}$

This decomposition is a.s. unique.

$$M_n := \sum_{i=0}^{n} (Z_i - E[Z_i | \mathcal{F}_{i-1}])$$

$$A_n := E[Z_0] + \sum_{i=1}^{n} (E[Z_i | \mathcal{F}_{i-1}] - Z_{i-1}), \ n \geq 1$$

with $A_0 := E[Z_0]$

$N \geq 1$

Let $(X_n, \mathcal{F}_n), n \geq 0$ be a SubMartingale and $-\infty < a < b < \infty$

Define stopping times $T_0 = 0$

$T_1 = \min\{0 \leq n \leq N | X_n \leq a\}$ $\quad T_2 = \min\{T_1 < n \leq N | X_n \geq b\}$

$T_{2m-1} = \min\{T_{2m-2} < n \leq N | X_n \leq a\}$

$T_{2m} = \min\{T_{2m-1} < n \leq N | X_n \geq b\}$

If $\min_{n \leq N} X_n > a$, then $T_1 = N$ & $T_2, T_3, \cdots$ undefined

The number of times till N that $\{M_n\}$ crossed (a,b) from below a to above b/ no. of upcrossings of (a,b) till N is R.V. $\beta_N(a,b) := \max\{m \geq 0 : T_{2m} \text{ is defined}\}$

Upcrossing Inequality.

$$E[\beta_N(a,b)] \leq \frac{E[(X_N - a)^+]}{b-a} \leq \frac{E[X_N^+] + |a|}{b-a}$$

Convergence Theorems

" Submartingale Convergence

Let $(X_n, \mathcal{F}_n)$ $n \geq 0$ be a submartingale. $\sup_n E[X_n^+] < \infty$. Then $\lim_{n \to \infty} X_n = X_\infty$ a.s.

Corollary.

All non-negative (martingales and supermartingales converge) a.s.

Regular Martingales

let $(M_n, \mathcal{F}_n), n \geq 0$ is regular if there exists an integrable R.V. Y such that $M_n = E[Y | \mathcal{F}_n] \ \forall n$

Theorem

Let $(M_n, \mathcal{F}_n), n \geq 0$ be a Martingale. Then following equivalent.

1. $(M_n, \mathcal{F}_n), n \geq 0$ is regular
2. $\{M_n\}$ is Uniformly Integrable.
3. $\exists$ an integrable R.V. M_∞ s.t $E[|M_n - M_\infty|] \to 0$
4. $\sup_n E[|M_n|] < \infty$ and $M_\infty := \lim_{n \to \infty} M_n$, which exists by Martingale CT satisfies $M_n = E[M_\infty | \mathcal{F}_n] \ \forall n$

Square-Integrable Martingales

A martingale $(X_n, \mathcal{F}_n), n \geq 0$ is said to be square-integrable if $E[X_n^2] < \infty \ \forall n$. For such $\{X_n\}$ $(X_n^2, \mathcal{F}_n)$ is a Submartingale & has Doob Decomposition $X_n^2 = M_n + A_n$, where $(M_n, \mathcal{F}_n)$ is 0-mean martingale and $\{A_n\}$ is an $\{\mathcal{F}_{n-1}\}$ adapted increasing process called "Quadratic Variation Process" of $\{X_n\}$.

We denote A_n by $\langle X \rangle_n$

$$A_n = \sum_{m=1}^{n} \left(E[X_m^2 | \mathcal{F}_{m-1}] - X_{m-1}^2 \right) + E[X_0^2]$$
$$= \sum_{m=0}^{n-1} E[(X_{m+1} - X_m)^2 | \mathcal{F}_m] + E[X_0^2], \ n \geq 0$$

Lemma

In the above, let T be an $\{\mathcal{F}_n\}$ stopping time. Then $(X_{T \wedge n}, \mathcal{F}_{T \wedge n}), n \geq 0$, is a square integrable martingale whose Quadratic Variation Process is $\{A_{T \wedge n}\}$.

Lemma Kronecker's Lemma

Let $\{x_k\} \subset \mathbb{R}$, $\{a_k\} \subset (0, \infty)$ be such that $a_k \uparrow \infty$ and $|\sum_n (x_n / a_n)| < \infty$.

Then $\left(\sum_{m=1}^{n} x_m\right) / a_n \to 0$

Theorem sq-integrable

For $\{X_n\}, \{\mathcal{F}_n\}, \{A_n\}$ as above, let $A_\infty = \lim_{n \to \infty} A_n$
$\langle M \rangle_\infty := \lim_{n \uparrow \infty} \langle M \rangle_n$. Then

1. $\{X_n\}$ converges a.s. on $\{A_\infty < \infty\}$
2. $X_n = o(f(A_n))$ on $\{A_\infty = \infty\}$ for every increasing $f : [0, \infty) \to [0, \infty)$ satisfying
$$\int_0^\infty \frac{1}{(1 + f(t))^2} dt < \infty$$

Corollary (SLLN for martingales)

If in the above, $\langle M \rangle_n = O(n)$, then $\frac{M_n}{n} \to 0$ a.s.

Theorem For a martingale $(M_n, \mathcal{F}_n), n \geq 0$ satisfying $E[\sup_n |M_{n+1} - M_n|] < \infty$, let $A := \{M_n \text{ converges}\}$ & $B := \{\limsup_{n \uparrow \infty} M_n = \infty = -\liminf_{n \uparrow \infty} M_n\}$, then we must have $P(A \cup B) = 1$

Corollary 1 If $\{X_n\}$ is a non-negative process adapted to $\{\mathcal{F}_n\}$ with $|X_{n+1} - X_n| \leq K$ for some finite constant K, then $\sum_{m=1}^{n} X_m$ and $\sum_{m=1}^{n} E[X_m | \mathcal{F}_{m-1}]$ converges or diverges together.

Corollary Conditional Borel-Cantelli Lemma

For $A_n \in \mathcal{F}_n$, $n \geq 1$, $\{A_n \text{ i.o.}\} = \left\{\sum_{n=1}^{\infty} P(A_n | \mathcal{F}_{n-1}) = \infty\right\}$ a.s.

(a) $E[I\{\cdots\}|\mathcal{F}_n] = 1\times P(\cdots|\mathcal{F}_n) + 0\times P(\text{not}\cdots|\mathcal{F}_n) = P(\cdots|\mathcal{F}_n)$ Also $E[I_A]=P(A)$

(b) if Y is f^n of $X_m, m\le n$, then $E[ZY|\mathcal{F}_n] = Y E[Z|\mathcal{F}_n]$ in particular, for $Y\equiv$ a constant

(c) For $n\ge m\ge 0$, $E[E[X|\mathcal{F}_n]|\mathcal{F}_m] = E[X|\mathcal{F}_m]$ ('Tower' property)

A sequence $\{X_n, n=0,1,2,\cdots\}$ of S-valued R.V. is called a Markov Chain with state space S, initial distribution λ and transition matrix P if

a) $P(X_0\in A) = \lambda(A)$, $\forall A\subset S$

b) $P(X_{n+1}=j\,|\,X_n=i) = P_n(j|i) = p(i,j)$

Markov Property c) $P(X_{n+1}\in A\,|\,\mathcal{F}_n) = P(X_{n+1}\in A\,|\,X_n)$ $\forall A\subset S, n\ge 0$

where $\mathcal{F}_n = \sigma(X_0, X_1, \cdots, X_n)$

Equivalently $P(X_{n-1}=i\,|\,X_m, m\ge n) = P(X_{n-1}=i\,|\,X_n), n\ge 0$ $i\in S$

Let $\lambda(i) = P(X_0=i)$

$P(\cdots) = P_\lambda(\cdots)$, $P(\cdots|X_0=i) = P_i(\cdots)$

Strong Markov Property (SMP)

$\{X_n\}$ has SMP if for every $\{\mathcal{F}_n\}$ stopping time τ, $P(X_{\tau+\cdot}\in A|\mathcal{F}_\tau) = P_{X_\tau}(X_\cdot\in A)$ on $\{\tau<\infty\}$ a.s.

Theorem Markov Chain $\{X_n\}$ satisfies SMP

It suffices to show that for any $A\in\mathcal{F}_\tau$ and $B\in\mathcal{B}(S^\infty)$

$$\int_A I\{\tau<\infty\} P([X_\tau, X_{\tau+1},\cdots]\in B|\mathcal{F}_\tau)\,dP$$
$$= \int_A I\{\tau<\infty\} P_{X_\tau}([X_0, X_1,\cdots]\in B)\,dP$$

5.2 Classification of States

$\tau_j = \min\{n>0\,|\,X_n=j\}$ for $j\in S$ "first hitting time of j"

$\rho_{ij} = P_i(\tau_j<\infty)$ Probability of hitting j starting from i after time 0.

$N_j = \sum_{n=1}^{\infty} I\{X_n=j\}$ Total visits to j after time 0 possibly infinity.

$G(i,j) = E_i[N_j]$ Expected number of visits to j after time 0 starting from i possibly ∞

$$= E_i\Big[\sum_{n=1}^{\infty} I\{X_n=j\}\Big] = \sum_{n=1}^{\infty} p^n(j|i)$$

Let $\sigma_1 = \tau_j$, $\sigma_n = \min\{m>\sigma_{n-1} : X_m=j\}$

Times of successive visits to j.

Equivalent Def $n>m$ $A_i\subset S$, $m<i<n$

$P(X_{m+1}\in A_{m+1},\cdots, X_{n-1}\in A_{n-1}\,|\,X_k, k\le m \text{ or } k\ge n)$
$= P(X_{m+1}\in A_{m+1},\cdots, X_{n-1}\in A_{n-1}\,|\,X_m, X_n)$

Lemma $P_i(N_j=m) = \rho_{ij}\,\rho_{jj}^{m-1}(1-\rho_{jj})$

Theorem

i) $\rho_{jj}<1 \Rightarrow P_i(N_j=\infty)=0$ and $G(i,j)<\infty$

ii) $\rho_{jj}=1 \Rightarrow P_j(N_j=\infty)=1$ and $G(j,j)=\infty$

$\Rightarrow P_i(N_j=\infty) = \rho_{ij} = 1-P_i(N_j=0)$,

& $G(i,j) = \begin{cases}\infty & \rho_{ij}>0\\ 0 & \rho_{ij}=0\end{cases}$

Transient States $S_T = \{j\in S\,|\,\rho_{jj}<1\}$

Recurrent States $S_R = \{j\in S\,|\,\rho_{jj}=1\}$

Write $i\to j$ whenever $\rho_{ij}>0$ $\quad\rho_{ij}>0 \Leftrightarrow i\to j$

Theorem

If $i\in S_R$ and $i\to j$, then $j\in S_R$ and $\rho_{ij}=\rho_{ji}=1$

$N_n(j) = \sum_{m=1}^{n} I\{X_m=j\}$ ($:=$ number of visits to j till time n)

$G_n(i,j) = E_i[N_n(j)] = \sum_{m=1}^{n} P_i\{X_m=j\}$

(expected no. of visits to j till time n)

$T_j = E_j[\tau_j]$ (expected hitting/return time of j)

$R_j^0 = \tau_j$, $R_j^n = \min\{n>R_j^{n-1} : X_n=j\}$

(successive hitting times of j)

$\tau_j^n = R_j^n - R_j^{n-1}$

Theorem

(a) $j\in S_T \Rightarrow \dfrac{N_n(j)}{n}\to 0$, P_i-a.s., $\dfrac{G_n(i,j)}{n}\to 0$ $\forall i\in S$

(b) $j\in S_R \Rightarrow \dfrac{N_n(j)}{n}\to\dfrac{I\{\tau_j<\infty\}}{T_j}$ P_i a.s. & $\dfrac{G_n(i,j)}{n}\to\dfrac{\rho_{ij}}{T_j}$

Define

$S_P = \{j\in S_R\,|\,T_j<\infty\}$ Positive Recurrent

$S_N = \{j\in S_R\,|\,T_j=\infty\}$ Null Recurrent

Theorem If $i\in S_P$ and $i\to j$, then $j\in S_P$

CLOSED & CONVEX HULL LEFT

MARKOV CHAIN TREE THEOREM LEFT

Summarizing

1. S is a disjoint union of S_P, S_N, S_T
2. S_P, S_N are disjoint unions of communicating classes, any finite communicating class is in S_P
3. S_T may or may not have a communicating class and may have bounded paths leading to S_P or S_N or unbounded paths. If S_T is finite, it does not have a communicating class & if it is infinite, the communicating class is also infinite.
4. If S_T is finite, $X_n \in S_P \cup S_N$ from some n on.
5. If S is a single communicating class, $S=S_P$ or $S=S_N$ or $S=S_T$ & we say that chain is positive recurrent, null recurrent or transient.

Stationary Distributions

A probability measure π on S is said to be stationary if $\sum_{j\in S} \pi(j)p(i|j) = \pi(i) \quad \forall i \in S.$

$P(X_n=i)=\pi(i), \forall i,n. \Rightarrow$ Joint law of $[X_n, X_{n+1}, \cdots$ is independent of n. i.e. $\{X_n\}$ is a stationary process

iterating $\pi P^m = \pi \quad m=0$ to $n-1$

$\pi\left(\frac{G_n}{n}\right) = \pi \qquad G_n := [[G_n(i,j)]]$

Theorem $\pi(i)=0 \ \forall i \notin S_P$

Theorem Let $\{X_n\}$ be irreducible and positive recurrent i.e. $S=S_P$. Then there exists a unique stationary distribution π given by $\pi(i)=\frac{1}{T_i}, i\in S$. Furthermore, for any initial law and any bounded $f: S \mapsto R$

$$\frac{1}{n}\sum_{m=0}^{n-1} f(X_m) \to \sum_i \pi(i)f(i) \quad a.s.$$

- $\leftrightarrow$ is an equivalence relation on S_P, S_N.

Say that $A \subset S$ is closed if $p(A|i)=1 \ \forall i \in A$

- It is irreducible if in addition, $i \to j \ \forall i,j \in A$.

The Markov Chain itself is said to be irreducible if S is

Lemma

$d_i := g.c.d.\{n \geq 1 : p^n(i|i) > 0\}$

If $i \to j$ and $j \to i$ then $d_i = d_j$

Gap Lemma Let $I \subset N_0 := \{0,1,2,\cdots\}$ satisfy (i) $I+I \subset I$ (ii) $g.c.d.(I)=1$. Then $\exists$ an $n_0 \in N_0$ such that $n \in I \ \forall n \geq n_0$

Theorem

Let Markov Chain be irreducible aperiodic positive recurrent with stationary distribution π, then $p^n(j|i) \to \pi(j) \quad \forall i,j \in S.$

Theorem

If $\{X_n\}$ is irreducible positive recurrent and periodic with period $d \geq 2$, then for each $i,j \in S$, $\exists$ a $0 \leq r < d$ such that $p^n(j|i)=0$ unless $n=md+r$ for some $m \geq 0$ and $p^{md+r}(j|i) \to d\pi(j)$

More on Periodicity

If P has period d, then all the d^{th} roots of 1 $\{e^{\frac{2n\pi i}{d}}, 0 \leq n < d\}$ are Eigenvalues of P.

In particular, for d even, -1 is an EigenValue and graph is bipartite.

Note: A periodic reversible Markov Chain must have period 2 corresponding to a bipartite graph.

classmate

Date ______
Page ______

Reversible Markov Chain

Consider irreducible positive recurrent Markov Chain on state space S with transition matrix $P=[[p(j|i)]]_{i,j\in S}$ and unique stationary distribution π

Global Balance $\sum_j \pi(j)p(i|j)=\pi(i),\ i\in S$

Local Balance (Time Reversibility) $\pi(i)p(j|i)=\pi(j)p(i|j),\ i,j\in S.$

$q_n(j|i):=P(X_{n-1}=j|X_n=i)$ time-reversed MC.

$\pi_n:=$ law of X_n,

$$q_n(j|i)=\frac{p(i|j)\pi_{n-1}(j)}{\pi_n(i)},\ i,j\in S.$$

→ Time inhomogeneous Markov Chain

If $\{X_n\}$ initiated with initial distribution π then $\pi_n=\pi\ \forall n$, so $q(j|i)=q_n(j|i)=p(j|i)$

$\Pi=$ diagonal matrix $(\pi(1),\pi(2),\ldots)$

For Reversible MC
$$\frac{\sqrt{\pi(i)}\,p(j|i)}{\sqrt{\pi(j)}}=\frac{\sqrt{\pi(j)}\,p(i|j)}{\sqrt{\pi(i)}}$$

i.e. $\Pi^{\frac{1}{2}}P\,\Pi^{-\frac{1}{2}}$ is SYMMETRIC.

$\therefore$ P is similar to a Symmetric matrix and must have only real eigenvalues

$\{\lambda_i\}$: Eigen Values of P

$\lambda_1=1\geq\lambda^*(:=\max_{i\neq 1}|\lambda_i|)$

Spectral Gap $=1-\lambda^*$

$\{l_i\},\{r_i\}$ left and right $\vec{EV}$ s.t. $l_i r_i=1$

$l_1=\pi \quad r_1=[1,1,\ldots]^T$

$\pi_n=\pi_0P^n=\sum_i\lambda_i^n(\pi_0 r_i)l_i$

Reversible

$p(j|i)=\frac{w_{ij}}{w_i}\quad \pi(i)=\frac{w_i}{w}$ satisfies local balance

$w_{ij}:=\pi(i)p(j|i)$

Explicit expressions for ~~Markov Chain Tree Theorem~~ stationary distribution.

Theorem: For any $i\in S$

$$\pi(j)=\frac{E_i[\sum_{m=1}^{\tau_i}I\{X_m=j\}]}{E_i[\tau_i]}$$

$$\pi(j)=\lim_{n\uparrow\infty}\frac{\frac{1}{n}\sum_{k=0}^{n-1}\sum_{m=\sigma_k+1}^{\sigma_{k+1}}I\{X_m=j\}}{\frac{1}{n}\sum_{k=0}^{n-1}(\sigma_{k+1}-\sigma_k)}$$

Corollary:

For any $f:S\mapsto R$ with $\sum_i\pi(i)|f(i)|<\infty$

$$\sum_i\pi(i)f(i)=\frac{E_i[\sum_{m=1}^{\tau_i}f(X_m)]}{E_i[\tau_i]}$$

Poisson Equation

Consider for a finite state irreducible Markov Chain $\{X_n\}$ on the state space $S=\{1,2,\ldots,s\}$

$V(i)=f(i)-\beta+\sum_j p(j|i)V(j),\ i\in S$

$\downarrow \times\pi(i)$ sum

$\therefore\ \beta=\sum_i\pi(i)f(i)$

$$(P-I)\vec{V}=\vec{f}\qquad \vec{f}=\begin{bmatrix}\beta-f(1)\\ \beta-f(2)\\ \vdots\end{bmatrix}$$

Markov Chain Tree Theorem

Let $\{X_n\}$ be an irreducible Markov Chain on a finite state space S with transition probabilities $p(j|i)$, $i,j \in S$ and G the associated directed graph

An arborescence is a subgraph of G such that:

1. There is atmost one edge out of each node,
2. There are no cycles, and,
3. It is maximal w.r.t. 1 and 2

Fact: There is exactly one node with no outgoing edge, called the root of the arborescence.

Lemma: If $q(j|i)$, $i,j \in S$, is another transition probability f" on S and π a probability measure on S such that

$$\pi(i)\, p(j|i) = \pi(j)\, q(i|j) \quad , \; i,j \in S$$

Then π is stationary under $p(\cdot|\cdot)$

Define.

$H_j :=$ set of all arborescences with root j, $j \in S$

$H :=$ set of all arborescences $= \bigcup_j H_j$

$w(a) :=$ 'weight of an arborescence' a

$:=$ product of transition probabilities associated with edges of a

$\|H_j\| :=$ sum of weights of arborescences in H_j, $j \in S$

$\|H\| := \sum_j \|H_j\| =$ sum of weights of all arborescences.

Theorem $\quad \pi(j) = \dfrac{\|H_j\|}{\|H\|}$, $j \in S$