Table of Contents

Preliminary Definitions:

a. Sup and Inf

Sup is the least upper bound and inf is the greatest lower bound.

Note: By completeness of R, we know that Sup and Inf always exist for any subset of R.

Note: Sup is weakly increasing as sets get larger, and Infs are weakly decreasing as sets get larger.

b. Lim Inf and Lim Sup of a Sequence of reals

$\limsup_n X_n = \lim_{n\to\infty} \sup_{k>n} X_k$ *and* $\liminf_n X_n = \lim_{n\to\infty} \inf_{k>n} X_k$

Note: Again, these are always defined.

Note: If the limit of the sequence exist, then it can be shown that limsup and liminf must be the same (i.e. equal to the limit).

c. Lim Inf and Lim Sup of a Set

Let $\{A_i : i \in I\}$ be a sequence of events

Then, $(A_i, i.o.) \equiv (A_i$ occurs infinitely often)

$\equiv \limsup A_i$

$\equiv \bigcap_{m=1} \bigcup_{n\geq m} A_n$ (Intuition: For n>1, $\omega \in$ Limsup is in ONE of the tail A_n's. This is true for all n.)

$= \{\omega : \forall m, \exists n(\omega) \geq m \text{ s.t. } \omega \in A_n(\omega)\}$

$= \{\omega : \omega \in A_n \text{ for infinitely many n}\}$

Idea: If $\omega \in$ LimSup $A_i \Rightarrow \omega \in \bigcap_{m=1} \bigcup_{n\geq m} A_n \Rightarrow \omega \in \bigcup_{n\geq 1} A_n$ AND $\omega \in \bigcup_{n\geq 2} A_n$ AND ... $\omega \in \bigcup_{n\geq k} A_n$ $\forall$k.

That is, ω is in ∞-many A_k's. So, a collection of such a collection of ω is a collection of ω s.t. A_k's occur i.o.

$(A_i, e.v.) \equiv (A_i$ eventually)

$\equiv \liminf A_i$

$\equiv \bigcup_m \bigcap_{n\geq m} A_n$ (Intuition: For n>1, $\omega \in$ Liminf is in ALL of the A_n's. This is true for all n.)

$= \{\omega : \exists m \text{ s.t. } \forall n>m, \omega \in A_n(\omega)\}$

$= \{\omega : \omega \in A_n \text{ for all large n}\}$

Idea: If $\omega \in \lim\text{Inf}\, A_i \Rightarrow \omega \in \bigcup_m \bigcap_{n\geq m} A_n \Rightarrow \omega \in \bigcap_{n\geq 1} A_n$ AND $\omega \in \bigcup_{n\geq 2} A_n$ AND ... $\omega \in \bigcap_{n\geq k} A_n$ $\forall$k.

That is, ω is in all the A_k's eventually. So, a collection of such a collection of ω is a collection of ω s.t. all A_k's occur eventually.

(See notes)

d. Metric Spaces

e. Normed Spaces

f. Inner Product Spaces

g. Stochastic Order

O(.): As $x \to ?$, $f(x) = O(g(x))$ iff $\exists$ C such that eventually $|f(x)| \leq C|g(x)|$. Or, the fraction $\frac{f(x)}{g(x)}$ is eventually bounded as x approaches limit.

o(.): As $x \to ?$, $f(x) = o(g(x))$ iff $\lim_{x\to ?} \frac{f(x)}{g(x)} = 0$. (i.e. f(x) goes to 0 faster than g(x) as x approaches the limit.

Op(.):

op(.)

Manipulating this notation:

Most of what we have here is common sense. Some useful ideas:

$O(g_1(x)) + O(g_2(x)) = O(\max(g_1(x), g_2(x))$

$O(g_1(x)) \cdot O(g_2(x)) = O(g_1(x) \cdot g_2(x))$

We'd like to be able to say that for reasonable functions w, $w(O(g(x))) = O(w(g(x))$. There's no problem with saying $(O(h))^2 = O(h^2)$, but $e^{O(\ln x)}$ isn't well defined. You have to be careful there.

You should avoid dividing by big-O or little-o. However, one can make sense of something like $\frac{1}{2+O(x)}$ by long division: $\frac{1}{2+O(x)} = \frac{1}{2} + O(x)$ as $x \to 0$.

Chapter 1: Probability Measure and Integration

I. Definitions:

a. Probability Space

Probability space is a triple (Ω, F, P) where:

$\Omega =$ Sample space = the set of all possible outcomes of some rando experimnt or phenomenon.

F = Event space = a subset of 2^{Ω} (set of all subsets), consisting of all "allowed events". Or, events to which we shall assign prob's.

(It represents both the amount of information available and the events of possible interest to us.)

P= Probability Measure = a set function P: $A \subseteq F \mapsto [0,1]$

b. (D.1.1.1)Sigma-Algebra: (We want to impose some structure on F)

We say that $F \subseteq 2^{\Omega}$ is a σ-field/σ-algebra if

(a) $\Omega \in F$

(b) $A \in F \Rightarrow \Omega \setminus A \in F$ (Complement Rule)

(c) $A_i \in F \Rightarrow \bigcup_i A_i \in F$ (Countable Union Rule) $\Rightarrow_{by \text{ De Morgan's Law}}$ $[A_i \in F \Rightarrow \bigcap_i A_i \in F]$ (Countable Intersection Rule) ([1])

$\Rightarrow (\emptyset, \Omega)$ always in σ-field

Note on 2^{Ω}: We denote "set of subsets" as 2^{Ω} because for a set of finite elements, it can be show that the power set has $2^{|\Omega|}$ elements.

Pf: Let $|\Omega| = N$.

Count all subsets with 0,1, 2, 3, 4, ... ,N elements, we get, $\binom{N}{0} + \binom{N}{1} + \ldots + \binom{N}{N} = \sum_{i=0}^{N} \binom{N}{i} = (1+1)^N = 2^N$ by binomial th.

Recall, binomial th. says $(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i}$

c. (D.1.1.2)Measurable Space: A pair (Ω, F) with F a sigma-field is called a measurable space.

d. (D.1.1.2)Probability Measure:

Given a measurable space, a probability measure is a function $\mathbf{P}: F \mapsto [0,1]$ s.t.

(a)$0 \le P(A) \le 1$ (b)$P(\Omega)=1$ (c)$P(A) = \sum A_n$ whenever $A = \bigcup A_n$ is a countable union of disjoint sets $A_n \in F$

e. (D.1.1.5)Generated Sigma Fields:

Given a collection of subsets $A_\alpha \subseteq \Omega$, where $\alpha \in \Gamma$ not necessarily a countable index set,

we denote **smallest sigma – field s.t.** $A_\alpha \in F\ \forall\, \alpha$ by $\sigma(\{A_\alpha\})$ or $\sigma(\{A_\alpha, \alpha \in \Gamma\})$, and call $\sigma(\{A_\alpha\})$ the sigma-field generated

by the collection $\{A_\alpha\}: \sigma(\{A_\alpha\}) = \bigcap \{G : G \subseteq 2^{\Omega}$ is a sigma-field, $A_\alpha \in G\ \forall\ \alpha \in \Gamma\}$

(intersection of all possible sigma-fields that contain $\{A_\alpha\}$ yields the smallest one)

(Note: This definition works because a (possibly uncountable) intersection of sigma-fields is a sigma-field)

f. (D.1.1.7/E.1.1.8)Borel Sigma-Field on R:

$\beta = \sigma(\{(a,b): a < b \in \Re\}) = \sigma(\{[a,b]: a, b \in \Re\}) = \sigma(\{(-\infty, b): b \in \Re\}) = \sigma(\{(-\infty, b]: b \in Q\}) = \sigma(\{O \subseteq \Re \text{ open}\})$

Note on Borel and $\sigma(X)$: It turns out that this has all the sets/events that we could possibly ever want. Now, we generally assume this sigma-field on the output space of a random variable. So, we want to be able to measure the Borel events on the output space. To do so, we want to make sure that the sigma field in the input space corresponds to this (since our probability measure is defined on the input space, and the measure on the output space is induced by the random variable), and thus we require $F = \sigma(\{\omega : X(\omega) \le \alpha\ \forall \alpha \in R\})$ so that in the output space we have all the Borel sets (because then we have in the output space $\sigma(\{(-\infty, \alpha], \alpha \in R\})$ which is just Borel).

---

[1] De Morgan's Laws: $A \wedge B = \neg(\neg A \vee \neg B)$ *and* $A \vee B = \neg(\neg A \wedge \neg B)$ or in set notation: $A \cap B = (A^c \cup B^c)^c$ *and* $A \cup B = (A^c \cap B^c)^c$

Equivalently: $\neg(A \wedge B) = (\neg A \vee \neg B)$ *and* $\neg(A \vee B) = (\neg A \wedge \neg B)$ or in set notation: $(A \cap B)^c = (A^c \cup B^c)$ *and* $(A \cup B)^c = (A^c \cap B^c)$

Thus, $\bigcap_i A_i = (\bigcup_i A_i^c)^c$. $A_i \in F \Rightarrow A_i^c \in F$ by Complement Rule $\Rightarrow \bigcup_i A_i^c \in F$ by Countable Union $\Rightarrow \bigcap_i A_i = (\bigcup_i A_i^c)^c \in F$ by Complement.

g. (D.1.2.8)Sigma Field Generated by a RV:

Given RV X, we denote by σ(X) the smallest σ-field G $\subseteq$ F s.t. X(ω) is measurable on (Ω, G).

We call σ(X) the σ-field generated by X, sometimes denoting by F_X.

It can be shown that:

$$\boxed{\sigma(X) = \sigma(\{\omega : X(\omega) \le \alpha\})}$$

Note on σ(X) : Naturally, we may want to consider $\hat{\sigma}$(X)=$\sigma(\{\omega : X(\omega) \in B\})$ for all Borel sets B. But, it can be shown that $\hat{\sigma}(X) = \sigma(X)$.

Note on σ(X) and information:

σ(X) is used to produce a rigorous mathematical theory. But it also has the crucial role of quantifying the amount of information we have.

For example, σ(X) contains exactly those events A for which we can say whether $\omega \in$ A or not.

So, if we knew whether all the sets in σ(X) happened, then we'd know the value of X.

h. (D.1.2.8)Sigma Field Generated by a Sequence of R.V.s:

Given $X_1,...,X_n$ on the same probability space (Ω, F), denote by $\sigma(X_k, k \le n)$ the smallest σ-field F s.t. $X_k(\omega)$ are measurable on (Ω, F).

Thus, $\sigma(X_k, k \le n) = \bigcap_{k=1}^{n} H_k$ (where H_k are sigma fields that contain $\sigma(X_k)$ i.e. smallest σ-field containing $\sigma(X_k)$ $\forall$k=1,...,n)

i. (D.1.2.1)Random Variable:

A R.V. on (Ω, F) is a function X:Ω $\mapsto R$ s.t. $\forall \alpha \in R$, the set $\{\omega : X(\omega) \le \alpha\} \in F$ (called F-measurable function)

(Equivalently: $\{\omega : X(\omega) \in A\} \in F \quad \forall$A)

Note on F-Measurability: Why do we care about this? So that we can induce a probability measure Px on R.

i.e. $\forall$A $\subset$ R, $P_X(A) = P(\{\omega : X(\omega) \in A\})$ is well-defined because $\{\omega : X(\omega) \in A\} \in F$ and thus we can assign a probability measure!

Note on Choice of F: In defining RV, F is implicit (i.e. X on (W,F,P) is by def F-measurable).

Note on X as Limit of SF: Any RV X can be expressed as a limit of a sequence of SF's.

j. (E.1.2.2)Indicators:

We call a R.V. (it can be easily verified) $I_A(\omega) = \begin{cases} 1, \omega \in A \\ 0, \quad o.w. \end{cases}$ an indicator function

Note on Use: Indicators are very useful because it allows us to use simple functions, and all R.V.'s can be expressed as limits of SF's.

k. (E.1.2.3)Simple Functions:

We call the R.V. $X_N(\omega) = \sum_{n=1}^{N} c_n I_{A_n}(\omega)$ (for finite N, non-random $c_n \in R$, and sets $A_n \in F$) a simple function, denoted by $X_N \in SF$.

l. (D.1.2.7) X,Y Almost Surely the same: We say X,Y def on same (Ω, F, P) are almost surely the same if $P(\{\omega : X(\omega) \ne Y(\omega)\}) = 0$

(Can show by showing Var(X-Y) = 0)

m. (D.1.2.11)Borel Measurable Function:

A function g:R $\mapsto$ R is called Borel (measurable) function if g(.) is a R.V. on (R, B) so that $\{x \in R : g(x) \le \alpha\} \in B \quad \forall \alpha$

Note: **Every continuous function and every piecewise constant (at most countably many jump points between which it is constant) function is Borel measurable.**

n. (D.1.2.18)Mathematical Expectation of a R.V.:

The mathematical expectation of a RV X(ω) is denoted EX.

$$EX = E\left[\lim_{n\to\infty} nI_{x>n} + \sum_{k=0}^{n2^n-1} k2^{-n} I_{(k2^{-n},(k+1)2^{-n}]}(x)\right] = \lim E\left[\sum x_{k,n} P(\{\omega : X(\omega) \in I_{k,n}\})\right]$$ **(A Lebesgue Integral)**

Note: This above formula is derived using the fact that any random variable is a limit and using some convergence theorem to switch order of E.

o. (D.1.2.24) Positive (X+) and Negative (X-) Parts of a RV

For a general R.V. X, consider the non-negative R.V.'s $X_+ = \max(X,0)$ and $X_- = -\min(X,0)$.

Then,

$X = X_+ - X_-,\ |X| = X_+ + X_-,\ and\ EX = EX_+ - EX_-$ (provided either $EX_+ < \infty$ or $EX_- < \infty$)

Note on Expectation Being Undefined: If $EX_+ = \infty$ and $EX_- = \infty$, then the expectation is undefined. Where as if only 1 of them is infinity, then the expectation is defined, but it does not exist for that random variable. Furthermore, when in many theorems we require the integrability condition, $E|X| < \infty \Leftrightarrow EX_+ + EX_- < \infty$ precisely to avoid the un-defined expectation problem. E|X| is always defined (since inf + inf = inf), where as EX is not always defined (since inf – inf = undefined. E.g. lim 1/n – lim 1/n = 0 but lim 2/n – lim 1/n = inf)

p. (D.1.2.25)Integrable R.V.:

A R.V. X is integrable (or has finite expectation) if E|X|<∞ (i.e. both $EX_+ < \infty$ and $EX_- < \infty$)

$\Leftrightarrow E\left(|X| I_{|X|>M}\right) \to_{M\to\infty} 0$

(See HW2)

q. (D.1.4.21,L.1.4.25)Uniform Integrability of a Collection of R.V.s

A collection of RV's $\{X_\alpha, \alpha \in I\}$ is called Uniformly Integrable (U.I.) if $\lim_{M\to\infty} \sup_\alpha E\left(|X_\alpha| I_{|X_\alpha|>M}\right) = 0$

OR EQUIVALENTLY

A collection of RV's $\{X_\alpha, \alpha \in I\}$ is U.I. iff

(a) $\sup_\alpha E|X_\alpha| < \infty$

(b) $\forall\ \varepsilon>0, \exists \delta>0$ s.t. $E\left(|X_\alpha| I_A\right) < \varepsilon\ \ \forall \alpha \in$ I and $\forall$ events A s.t. P(A)<δ

Note on Intuition: This says that if the tails of the collection of RV's are uniformly bounded in some sense (i.e. the largest tail expectation of the collection goes to 0 as we go farther and farther out in the tail.

Note on Use: In our class, this becomes useful when we know that X(n)→pX, and $|X(n)|^q$ uniformly integrable, then $X(n)\to_{q.m.}X$.

r. Sufficient Conditions for Uniform Integrability of a Collection of R.V.'s

(Lecture 4 pg.8)

a) If $X \in L^1$ then X is U.I. (a single X that is integrable is trivially U.I.)

b) If $\exists$ C<∞ s.t. $|X_n| \le C$ a.s. ∀n, then $\{X_n\}$ U.I.

c) If $X_i \in L_1$ for i=1,...,n (a finite collection of R.V.), then $\{X_1,...,X_n\}$ U.I.

d) If $\mathrm{Sup}_n\ E|X_n|^{1+\varepsilon} < \infty$ for some ε>0, then $\{X_n\}$ is U.I.

s. (D.1.3.1,D.1.3.2,D.1.4.9,P.1.4.10)Convergences of R.V.s

(0) Pointwise: $X_n \to_{ptwise} X$ if $\forall \omega \in \Omega$, $X_n(\omega) \to_{n\to\infty} X(\omega)$

(1) Almost Sure: $X_n \to_{as} X$ if $\exists A \in$ F w/ P(A)=1 s.t. $X_n(\omega) \to_{n\to\infty} X(\omega)$ for each $\omega \in A$

$\Leftrightarrow\ P\left(\{\omega\in\Omega: \lim_n X_n(\omega) = X(\omega)\}\right) = 1 \Leftrightarrow \forall\varepsilon>0\ P\left(\{\omega\in\Omega: |\lim_n X_n(\omega) - X(\omega)| < \varepsilon\}\right) = 1 \Leftrightarrow \forall\delta>0\ P\left(\{\omega\in\Omega: |\lim_n X_n(\omega) - X(\omega)| > \delta\}\right) = 0$

$\Leftrightarrow\ \forall\varepsilon>0,\ P\left(\bigcup_n \bigcap_{k\ge n} \{\omega: |X_k(\omega) - X(\omega)| \le \varepsilon\}\right) = 1\ \Leftrightarrow \forall\varepsilon>0,\ \lim_n P\left(\bigcap_{k\ge n}\{\omega: |X_k(\omega) - X(\omega)| \le \varepsilon\}\right) = 1$

(i.e. eventually, it's within ε, which is just the def of convergence)

$\Leftrightarrow\ \forall\varepsilon>0,\ P\left(\bigcap_n \bigcup_{k\ge n} \{\omega: |X_k(\omega) - X(\omega)| > \varepsilon\}\right) = 0\ \Leftrightarrow \forall\varepsilon>0,\ \lim_n P\left(\bigcup_{k\ge n}\{\omega: |X_k(\omega) - X(\omega)| > \varepsilon\}\right) = 0$

(i.e. this is just an equivalent statement of the above. The event that we're outside ε neighborhood of the limit infinitely often must have measure 0)

(2) Probability: $X_n \to_P X$ if $P\left(\{\omega\in\Omega: |X_n(\omega) - X(\omega)| > \varepsilon\}\right) \to 0\ \ \forall\varepsilon>0$ fixed

(3) Lq/q-Mean: $X_n \to_{q.m.} X$ if $X_n, X \in L^q$ and $\|X_n - X\|_q \to_{n\to\infty} 0$ $\left(\text{i.e. } E\left(|X_n - X|^q\right) \to_{n\to\infty} 0\right)$

(4) Distribution/Law/Weakly: $X_n \to_D X$ if $F_{X_n}(\alpha) \to_{n\to\infty} F_X(\alpha)$ for each α that is a continuity point of F_X.

(or equivalently) $X_n \to_D X$ iff ∀h continuous and bounded (on the range of X), we have $E\left(h(X_n)\right) \to_{n\to\infty} E(h(X))$

(5) Expectation: $X_n \to_{\exp} X$ if $E(X_n) \to_{n\to\infty} E(X)$

Note: Convergence in expectations depends on monotone/dominated convergence theorems

Note complete probability space: In principle, when dealing with a.s. convergence, we need to also check that the limit X is also a RV. If we assume that the probability space is complete (i.e. contains all limits), which we always do, then we can ignore this technical point. (See below)

Note on 2nd definition of convergence in probability: This is an equivalent definition (P.1.4.10) that is an alternative definition which applies to more general R.V. whose range is not R (for example random vectors with values in R_n). This condition is harder to check.

t. (D.1.3.4)Complete Probability Space

We say that (Ω, F, P) is a complete probability space if ANY SUBSET N of B$\in$ F with $P(B)=0$ is also in F.

Note: Any probability space can be completed by adding to F all the subsets of sets of probability 0.
Note on why we care: Completeness guarantees that an a.s. limit of a RV is itself a RV.

It is possible to show that if X_n are RV's s.t. $X_n(\omega) \to_{n\to\infty} X(\omega)$ for each $\omega \in$ A and P(A)=1, then there exists a RV $\hat{X}$ s.t. N=$\{\omega : X(\omega) \neq \hat{X}(\omega)\}$ is a subset of B=$A^C \in F$ and $X_n \to_{as} \hat{X}$. By assuming that the probability space is complete, we guarantee that N is in F, and therefore $X =_{as} \hat{X}$ is necessarily a random variable.

u. (D.1.3.2) L^q space

Fixing $1 \le q < \infty$, we denote by $L^q(\Omega, F, P)$ the collection of RV's X on measurable space (Ω, F) for which $E(|X|^q) < \infty$

Note on convergence: Convergence in Lq is the convergence of RV's in the Lq space (defined above), with the $\|\,.\,\|_q$ metric.
Examples: L^1 denotes the space of all integrable random variables. L^2 denotes the space of all square-integrable random variables.

Important structural properties of Lq (P.1.3.16):

$L^q(\Omega, F, P)$ is a complete, normed (topological) vector space (i.e. $\alpha X + \beta Y \in L^q$ whenever X,Y $\in L^q$, $\alpha, \beta \in R$)

with the norm $\|\,.\,\|_q = E(|\,.\,|^q)^{1/q}$. If $\|X_n - X_m\|_q \to_{m,n\to\infty} 0$ then $X_n \to_{qm} X$ for some X in Lq. (i.e. complete)

v. (D.1.4.1)Law of a R.V.

The law of a RV X, denoted P_X, is the probability measure on (R, B) s.t. $P_X(U) = P(\{\omega : X(\omega) \in U\})$ $\forall U \in$ Borel

(In other words, P_X is the measure induced by X on R)

w. (D.1.4.4.)Distribution Function of a R.V.

The distribution function F_X of a real-valued R.V. X is $F_X(\alpha) = P(\{\omega : X(\omega) \le \alpha\}) = P_X((-\infty, \alpha])$ $\forall \alpha$

Note on Existence of F_x: Since by definition X must be measurable w.r.t. $\sigma(X) \equiv \sigma(\{\omega : X(\omega) \le \alpha \ \forall \alpha\})$ **,** therefore we know the measure of these sets is well defined always. Thus, the **distribution function always exists! But PDF does not always exist, because (as shown below) this is only true for continuous and differentiable (almost everywhere) CDF's**.

Note on Law and F_X: The distribution function uniquely determines the law P_x of X (just as the characteristic function).

x. (D.1.2.22,P.1.4.7)Probability Density Function (PDF):

A R.V. X(ω) has a PDF f_x if $P(a \le X \le b) = \int_a^b f_X(x)dx$ $\forall$a < b$\in$ R. Such a density must be a non-ive function with $\int_R f_X(x)dx = 1$.

OR Equivalently

A R.V. X has a PDF f_x iff its distribution function F_X can be expressed as

$$\boxed{F_X(\alpha) = \int_{-\infty}^{\alpha} f_X(x)dx \ \ \forall \alpha} \qquad \text{where } f_x \ge 0 \text{ and } \int f_x(x)dx = 1 \ .$$

Note on Non-Existence of PDF:

If PDF exists, then it must be that $F_X(b) = \int_{-\infty}^{b} f_X(x)dx$, and thus by Fundamental Th. Of Calculus, $F_X'(b) = f_X(x)$**.**

However, if F(x) is not differentiable, then PDF does not exist (by contrapositive)! PDF exists is to say that we can write the CDF as a Riemann integral over R as mentioned above. But, remember that the Lebesgue integral can still be written, since by definition,

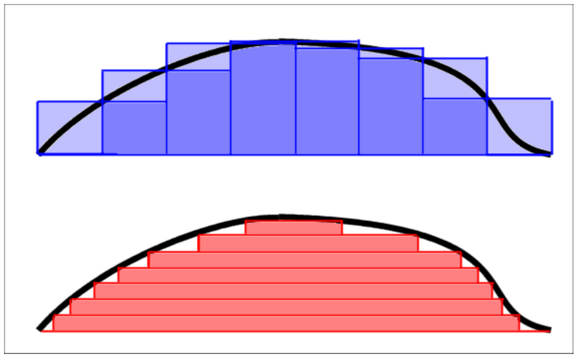
$F_X(b) = P_X(X \in [-\infty, b]) = P(\{\omega : X(\omega) \le b\}) = \int_{\{\omega : X(\omega) \le b\}} dP(\omega)$.

Note on Existence of PDF and Property of CDF: If PDF exists, then we know that F_X is continuous and almost everywhere differentable with $F_X'(b) = f_X(x)$ for almost every x**.**

Note on PDF's, Lebegue vs. Riemann Integral:

The Lebesgue integral of X wrt the probability P is denoted E(X)=$\int X(\omega)dP(\omega)$. It is based on splitting the range of X(ω) into finitely many small intervals and approximating X(ω) by a constant on the preimage of each of the intervals. This allows us to deal with rather general domain Ω.
In contrast, the Riemann integral splits the domain of integration into finitely many small intervals (hence we're limited to R^d). Even when Ω=[0,1] it allows us to deal with measures for which P for which $\omega \to P([0,\omega])$ is not smooth (and hence Riemann integral fails to exist). But if the Riemann integral exists, then it necessarily coincides with the Lebesgue Integral.

Thus, if PDF exists, then (Riemann) $\int x f_X(x)dx = \int_\omega X(\omega)dP(\omega)$ (Lebesgue)



y. (D.1.4.32,D.1.3.33,D.1.4.35) Independence of Events, Sigma-Fields, and Random Variables

(1)Independence of Events: Events $A_i \in F$ are P-mutually independent if for any L<∞ and distinct indices $i_1,...,i_n$,

$$\boxed{P\left(A_{i_1} \cap ... \cap A_{i_n}\right) = \prod_{j=1}^{n} P\left(A_{i_j}\right)}$$

(2)Independence of σ-Fields: Two σ-fields H,G ⊆ F are P-independent if

$$\boxed{P(g \cap h) = P(g)P(h) \quad \forall g \in G, h \in H}$$

n σ-fields $H_1,...,H_n \subseteq F$ are P-independent if

$$\boxed{P(h_1 \cap ... \cap h_n) = P(h_1)...P(h_n) \quad \forall h_i \in H_i \quad \forall i}$$

(3)Independence of R.V.'s: Two RV's X,Y defined on (Ω, F, P) are independent if $\sigma(X) \perp \sigma(Y)$ (where $\sigma(X), \sigma(Y) \subseteq F$)

OR EQVALENTLY

if $P(X = x, Y = y) = P(X = x)P(Y = y)$

(4)Independence of a Random Vector: For any finite n,m ≥ 1, two random vectors $(X_1,...,X_m)$ and $(Y_1,...,Y_n)$ with values in R^m and R^n, respectively, are independent iff for all bounded Borel measurable functions g: $R^m \to R$ and h:$R^n \to R$

$$\boxed{E\left(h(X_1,...,X_m)g(Y_1,...,Y_n)\right) = E\left(h(X_1,...,X_m)\right)E\left(g(Y_1,...,Y_n)\right)}$$

Note on Proving (4): From (3), we have that if h is the indicator variable, then (4) immediately follows. Now, we do the trick to show that if (4) is true for indicators, then it's true for simple functions. Then, if it's true for simple functions then it's true for all functions. However, we impose the bounded and measurable conditions on the function so that the expectation is defined.

Note on Checking Independence Condition: It's useful, when checking that all sets in sigma-field are mutual independent, to use the following

$$A \perp B \Leftrightarrow A^c \perp B \Leftrightarrow A \perp B^c \Leftrightarrow A^c \perp B^c$$

II. Theorems/Propositions:

a. (E.1.1.3)Properties of Probability Measures

Let $(\Omega, F, \mathbf{P})$ be a probability space and A, B, A_i events in F. Then, **P** satisfies:

(a) Monotonicity: If $A \subseteq B$ then $P(A) \le P(B)$

(b) Sub-additivity: If $A \subseteq \bigcup_i A_i$ then $P(A) \le \sum P(A_i)$

(c) Continuity from Below: If $A_i \uparrow A$ (i.e. $A_1 \subseteq A_2 \subseteq \ldots$ and $\bigcup_i A_i = A$), then $P(A_i) \uparrow A$

(d) Continuity from Above: If $A_i \downarrow A$ (i.e. $A_1 \supseteq A_2 \supseteq \ldots$ and $\bigcap_i A_i = A$), then $P(A_i) \downarrow A$

(e) Inclusion-Exlusion Rule: $P(\bigcup_i A_i) = \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n+1} P(A_1 \cap \ldots \cap A_n)$

b. (E.1.1.6) Properties of Sigma-Fields (Unions and Intersections)

Let A_α be a σ-field for each $\alpha \in \Gamma$, an arbitrary (possibly uncountable) index set. Then, $\bigcap_{\alpha \in \Gamma} A_\alpha$ is a σ-field, but $A_{\alpha_1} \cup A_{\alpha_2}$ not necessarily one.

c. (P.1.1.19) Borel set does not contain all subsets of R

There exists a subset of R that is not in Borel set. Thus, not all subsets of R are Borel sets. (But for all intents and purposes, this is a technical and unimportant detail)

d. (E.1.2.3) Properties of Indicator R.V.'s:

(a) $I_\emptyset(\omega) = 0$ and $I_\Omega(\omega) = 0$

(b) $I_{A^C}(\omega) = 1 - I_A(\omega)$

(c) $I_A(\omega) \le I_B(\omega)$ iff $A \subseteq B$

(d) $I_{\bigcap_i A_i}(\omega) = \prod_i I_{A_i}(\omega)$

(e) If A_i disjoint then $I_{\bigcup_i A_i}(\omega) = \sum_i I_{A_i}(\omega)$

e. (P.1.2.6)RV as (pt-wise) limit of SF

For every RV $X(\omega)$, there exists a sequence of SF's $X_n(\omega)$ s.t. $X_n(\omega) \to X(\omega)$ for each fixed $\omega \in \Omega$ as $n \to \infty$

Use $f_n(x) = nI_{x>n} + \sum_{k=0}^{n2^n-1} k2^{-n} I_{(k2^{-n},(k+1)2^{-n}]}(x)$

What does this look like? It partitions R into the following intervals of width $\frac{1}{2^n}$:

$$0, \frac{1}{2^n}, \quad \frac{1}{2^n}, \frac{2}{2^n}, \quad \frac{3}{2^n}, \frac{4}{2^n}, \ldots, \frac{n2^n - 1}{2^n}, \frac{n2^n}{2^n} = n - \frac{1}{2^n}, n, \quad (n, \infty$$

Then, whenever the function's values fall in one of these intervals, we approximate by the lower limit. (So, whenever x>n, we give it n)
(See notes for a picture)

f. (D.1.2.7)X=Y A.S. : **If Var(X-Y) = 0 then X = Y a.s.**

g. (E.1.2.10,P.1.2.13,P.1.2.14,T.1.2.16)Closure Properties of R.V.'s:

1. (E.1.2.10): Limit of R.V. on a measurable space is a R.V. on that space

   Let (Ω, F) be a measurable space and let X_n be a sequence of random variables on it. If $\forall \omega \in \Omega$, $X_\infty(\omega) = \lim_{n\to\infty} X_n(\omega)$ exists and is finite.

   Then, $X_\infty(\omega)$ is a RV on (Ω, F).

   Note on usefulness:

   For a typical measurable space with uncountable Ω, it is impractical to list all possible R.V.s that are defined on it.

   Using this closure property, we are given a tool that helps us show that a particular function $X(\omega)$ is inded a RV.

   (i.e. by showing that it's a limit to a sequence of RV's on (Ω, F))

2. (P.1.2.13) Borel function of R.V.'s on a measurable space is a R.V. on that space

   If $g: R^n \mapsto R$ is a Borel measurable function and $X_1, \ldots, X_n$ are R.V. on (Ω, F), then $g(X_1, \ldots, X_n)$ also a R.V. on (Ω, F)

3. **(P.1.2.14) Borel functions of R.V.'s generate a smaller sigma field (corollary from above)**

   For any $n < \infty$, $g: R^n \to R$ Borel measurable and any R.V. $Y_1, \ldots, Y_n$ on the same measurable space, then $F_{g(Y_1,\ldots,Y_n)} \subseteq \sigma(Y_1, \ldots, Y_n)$.

   If g invertible, then $F_{g(Y_1,\ldots,Y_n)} = \sigma(Y_1, \ldots, Y_n)$.

Note on Intuition: The idea here about information. If Y(n) creates a particular information partition, then some function of Y(n) can't give you any more information. G(Y(n)) gives you exactly the same information if G is invertible, because from G you can always back out Y.

h. (P.1.2.31)Properties of Expectation

(1) (Indicator) $EI_A = P(A)$ for any $A \in F$

(2) If $X(\omega)=\sum_{n=1}^{N} c_n I_{A_n}$ is a simple function, then $E(X)=\sum_{n=1}^{N} c_n P(A_n)$

(3) (Linearity) If X and Y are integrable R.V. then for any constants α, β the R.V. $\alpha X+\beta Y$ is integrable and $E(\alpha X+\beta Y) = \alpha E(X) + \beta E(Y)$

(4) (Constant) E(X)=c if $X(\omega)$ = c w.p. 1

(5) (Monotonicity) If X ≥ Y a.s., then E(X) ≥ E(Y). Further, if X ≥ Y a.s. and EX=EY, then X=Y a.s.

i. (P.1.2.34,T.1.2.36,P.1.2.38) Inequalities (Jensen's, Markov's, Chebychev's, Cauchy-Schwartz)

(1) Jensen's: Suppose g(.) is a convex function, i.e. $g(\lambda x+(1-\lambda)y) \le \lambda g(x) + (1-\lambda)g(y)$ $\forall x,y \in R$, $\lambda \in [0,1]$

If X is an integrable R.V. and g(X) is also integrable, then $\boxed{E(g(X)) \ge g(E(X))}$ w/ strict inequality if g(.) strictly convex.

(2) (General) Markov's: Suppose f is a non-decreasing, Borel measurable function with f(x) > 0 for x > 0.

Then for any R.V. X, $\boxed{\forall \varepsilon>0,\ P(|X(\omega)| \ge \varepsilon) \le \frac{1}{f(\varepsilon)} E(f(|X|))}$

(3) (Usual) Markov's: Take above with f(x) = x. Then, $\boxed{\forall \varepsilon>0,\ P(|X(\omega)| \ge \varepsilon) \le \frac{E|X|}{\varepsilon}}$

(4) Chebychev's: Take above with f(x) = x^2 and X = Y-E(Y). Then, $\boxed{\forall \varepsilon>0,\ P(\text{Y-E(Y)}^2 \ge \varepsilon) \le \frac{E(\text{Y-E(Y)}^2)}{\varepsilon^2} = \frac{Var(Y)}{\varepsilon^2}}$

(5) Cauchy-Schwartz: Suppose X,Y RV's on the same probability space w/ E Y^2 < ∞ and E X^2 < ∞. Then, $\boxed{E|YZ| \le \sqrt{E(Y^2)E(Z^2)}}$

Note: In particular, for X,Y with mean 0, $|EXY| \le_{Jensen's} E|XY| \le_{Cauchy-Schwartz} \sqrt{E(Y^2)E(X^2)}$ OR, $\boxed{|Cov(X,Y)| \le Var(X)Var(Y)}$

(6) Triangle: $d(x,y) \le d(x,z) + d(z,y)$

Using Euclidean d(x,y)=|x-y|: $|x+y| \le |x| + |y|$ and $|x-y| \ge ||x|-|y||$

j. (T.1.3.6,P.1.3.21,C.1.3.15) A.S., Q.M. P, D Convergence Relationships

a.s. and p convergence

(a) $X_n \to_{as} X \Rightarrow X_n \to_P X$ (and converse not true)

(b) (BUT) $X_n \to_P X \Rightarrow \exists$ subsequence n_k s.t. $X_{n_k} \to_{as} X$ for $n \to \infty$

a.s. and q.m. convergence

(c) $X_n \to_{qm} X$ does not imply $X_n \to_{as} Y$ and vice versa.

(d) If $X_n \to_{qm} X$ *and* $X_n \to_{as} Y$ then $X = Y$ *a.s.*

q.m. and r.m. convergence

(e) If $X_n \to_{q.m.} X \Rightarrow X_n \to_{r.m.} X \quad \forall r \le q$

q.m. and p convergence

(f) $X_n \to_{q.m.} X \Rightarrow X_n \to_p X$ (Converse not true except following)

(g) $X_n \to_p X$, $|X_n|^q$ U.I. $\Rightarrow X_n \to_{q.m.} X$

(Dominated convergence is a special case of (g), with q = 1)

other convergences and d convergence

(h) $X_n \to_P X \Rightarrow X_n \to_D X$ (Converse not true, except for special case when X is nonrandom constant a.s.)

$X_n \to_{qm} X \Rightarrow X_n \to_D X$

$X_n \to_{as} X \Rightarrow X_n \to_D X$

(i) If $X_n \to_D X$ and X is nonrandom constant a.s. $\Rightarrow X_n \to_P X$

k. (P.1.3.12)Property of the $\|\cdot\|_q$ Norm: $\|X\|_q$ is nondecreasing in q

l. (L.1.3.10,L.1.3.11) Borel Cantelli Lemmas

BCI: Let $A_k \in F$ and $\sum_{k=1}^{\infty} P(A_k) < \infty$. Then, $P(A^{\infty}) \equiv P\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k\right) = P\left(A_k \text{ i.o.}\right) = 0$

BCII (converse with independence): If A_k independent and $\sum_{k=1}^{\infty} P(A_k) = \infty$, then $P\left(A^{\infty}\right) \equiv P\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k\right) = P\left(A_k \text{ i.o.}\right) = 1$

Note on Intuition: BC1 states that a.s., A_k occurs for only finitely many values of k (i.e. A_k does not occur for infinitely many k) if the sequence $P(A_k)$ converges to 0 fast enough (so that the sum is < inf). BC2 is the converse of BCI when we know that A_k's are mutually independent. The converse is not true if the independence does not hold.

Note on Usefulness: BC lemmas allow us to prove many results (ABOUT ALMOST SURE CONVERGENCE). For example, j(b) above can be proven using this. See following for a typical application of BCI.

m. (P.1.3.22) Typical Application of BCI: $E(X_n^2) \le 1\ \forall n \Rightarrow \frac{1}{n} X_n(\omega) \to 0$ a.s.

Pf: WTS: $P\left(\lim \frac{1}{n} X_n(\omega) = 0\right) = 1$ or equivalently $\forall \delta > 0\ P\left(\omega \in \Omega: \left|\lim_n \frac{1}{n} X_n(\omega)\right| \ge \delta\right) = 0$

Fix $\delta > 0$. Let $A_k = \left\{\omega : \left|\frac{1}{k} X_k(\omega)\right| \ge \delta\right\}$. Then we want to show that the Prob that A_k occurs for infinitely many k is 0.

Using BCI, we know $P\left(A_k\right) = P\left(\omega : \left|\frac{1}{k} X_k(\omega)\right| \ge \delta\right) \le \frac{E\left(X_k(\omega)^2\right)}{k^2\delta^2} = \frac{1}{k^2\delta^2}$ by Chebychev

Then, $\sum_{k=1}^{\infty} P\left(A_k\right) \le \frac{1}{\delta^2}\sum_{k=1}^{\infty}\frac{1}{k^2} \le \infty$.

Thus, $P\left(A^{\infty}\right) = P\left(\omega \in \Omega: \left|\lim_n \frac{1}{n} X_n(\omega)\right| \ge \delta\right) = 0$. So, $\frac{1}{n} X_n(\omega) \to 0$ a.s.

n. (T.1.4.26,T.1.4.27,C.1.4.28,T.1.4.29) (Dominated/Bounded/Monotone) Convergence Theorems to Show $E\left(\lim X_n\right) = \lim E\left(X_n\right)$

(0) Fatou's Lemma: If $\exists\ X \in L^1$ and $X_n(\omega) \ge X(\omega)\ \forall n$, then $\lim \inf_{n\to\infty} E\left(X_k\right) \ge E\left(\liminf X_k\right)$

where $\liminf_{n\to\infty} X_n$ denotes $\lim_{n\to\infty} \inf_{k>n} X_k$

(1) Dominated Convergence: Suppose $\exists$ a RV Y s.t. $EY<\infty$, $\left|X_n\right| \le Y\ \forall n$. If $X_n \to_P X$, then $E\left(X_n\right) \to_{n\to\infty} E(X)$ [i.e. $E\left(\lim X_n\right) = \lim E\left(X_n\right)$]

(2) Bounded Convergence: Suppose $\forall n$, $\left|X_n\right| \le C$ for some constant C finite. If $X_n \to_P X$, then $E\left(X_n\right) \to_{n\to\infty} E(X)$ [i.e. $E\left(\lim X_n\right) = \lim E\left(X_n\right)$]

(3) Monotone Convergence: Suppose $X_n \ge 0$ and $X_n(\omega) \uparrow X(\omega)$ for a.e. ω (almost sure convergence), then $E\left(X_n\right) \uparrow_{n\to\infty} E(X)$

where $X_n(\omega) \uparrow X(\omega)$ means $X_n(\omega) \le X_{n+1}(\omega)\ \forall \omega \in A$ s.t. P(A)=1, and $X_n(\omega) \to_{n\to\infty} X(\omega)$

Note on Fatou's Lemma: Fatou's is useful because it tells us something about the expectations even if we don't know that X(n)→pX, or even if we don't know that the limit of the expectation exists. Since {E(Xn)} is just a sequence of reals, then liminf / limsup always exists by completeness of R. Now, if a limit to E(Xn) exists, then lim inf E(Xn) and lim sup E(Xn) must coincide. Note also that Fatou's does HALF THE WORK. If we can show the reverse inequality, then we can show that the limits are equal.

Note on Dominated vs. Bounded: Bounded convergence is just a special case of dominated where the RV Y is a constant.

Note on Dominated Convergence as a Special Case: Dominated convergence is a special case of $X_n \to_p X$, $|X_n|^q$ U.I. $\Rightarrow$ $X_n \to_{q.m.} X$

o. (T.1.4.22) Uniform Integrability + Convergence in Prob → q.m. Convergence

If $X_n \to_p X$ and $\left\{\left|X_n\right|^q\right\}$ are U.I., then $X_n \to_{q.m.} X$ $\left(\text{i.e. } E\left(X_n^q\right) \to_{n\to\infty} E\left(X^q\right)\right)$

Note on Dominated Convergence as a Special Case: Dominated convergence is a special case of $X_n \to_p X$, $|X_n|^q$ U.I. $\Rightarrow$ $X_n \to_{q.m.} X$

Dom. Convergence is a special case because if ∃ a RV Y s.t. EY<∞, $|X_n| \le Y \ \forall n \Rightarrow \{X_n\}$ U.I.
So, if we have this dominated convergence condition, we have U.I., and we also have convergence in prob, then we have L1 convergence by theorem, which gives us the result of dominated convergence.

Note on Versions of this Theorem: Some versions of this uses Xn→as X. But it does not matter because in this case they're equivalent, by Skohorod representation theorem. Remember for the theorem to hold, we only need to show that it holds for the distribution of X^q.

p. (P.1.4.3) Change of Variables (We can measure a RV either in the input space of in the output space!)
Let X be a RV on (Ω, F, P) and let g be a Borel function on R. Suppose either g is nonnegative or $E|g(X)| < \infty$. Then,

$$E(g(X)) = \int_{\omega:\Omega} g(X(\omega))\, dP(\omega) = \int_R g(x)\, dP_X(x)$$

Note on Borel-Measurability:
$g(x):(R,B) \to (R,B)$ Borel measurable $\Rightarrow \forall \alpha, \{x \in R : g(x) \le \alpha\} \in B$. But the RV X has induced a measure on (R,B), so we can use that measure on R, OR, we can map it back to the original probability space and measure via the original P.
(All functions we will care about are going to be Borel measurable).

Note on Usefulness: This allows us to have a consistent framework in thinking about taking expectations either by using the measure on the output space (i.e. cdf/pdf) or by using measure on the input space (i.e. P).

Examples: f(x) = x identity function. $E(g(X)) = \int_R x\, dP_X(x) = \int_{\omega:\Omega} X(\omega)\, dP(\omega)$

q. (P.1.4.5) Distribution Function F_X Uniquely Determines the Law P_X of X (Just as Characteristic Function)

r. (P.1.4.3) $X =_D Y$ → E[h(X)] = E[h(Y)] for all h bounded and Borel measurable.
We cal show this using the delta method.

Chapter 2: Conditional Expectation and Hilbert Spaces

I. Definitions

a. Expression for Conditional Expectation for Discrete R.V. Y:

$$\boxed{E(X \mid Y=y)=\frac{E\left(XI_{\{Y=y\}}\right)}{P\left(Y=y\right)}=\int_{\{\omega:Y(\omega)=y\}} X(\omega)\,\frac{dP(\omega)}{P\left(\left\{\omega:Y(\omega)=y\right\}\right)}}$$

II. Theorems/Propositions

a. (P.2.1.2)Projection Theorem:

There exists a unique (a.s.) optional $Z \in H_Y = L^2(\Omega, \sigma(Y), P)$ s.t. $E\left[(X-Z)^2\right] = \inf_{W \in H_Y} E\left[(X-W)^2\right]$.

Further, the optimality of Z is equivalent to the orthogonality property $\boxed{E\left[(X-Z)V\right] = 0 \ \ \forall V \in H_Y}$

b. (T.2.1.6)Consistency of L2 and General Definition of E(X|Y):

The C.E. of integrable R.V. X given any σ-field G exists and is a.s. unique. That is , there exists Z measurable on G satisfying partial averaging and if Z_1 and Z_2 are both measurable on G satisfying partial avraging, then $Z_1 = Z_2$ a.s.

Further, if in addition $E(X^2) < \infty$ then such Z also satisfies $E\left[(X-Z)V\right] = 0 \ \forall V \in L^2(\Omega, G, P)$.

(Hence for G = F_Y, R.V. Z coincides with the definition of E(X|Y) for X in L^2)

c. (E.2.3.1,E.2.3.3.,P.2.3.4,P.2.3.5) Properties of Conditional Expectations

(The following can be easily verified by checking measurability and partial averaging condition)

(1) (Independence) If σ(X) ind σ(Y), E(X|Y) = E(X)

(2) (Trivial σ-Field) If $F_0 = \{\Omega, \varnothing\}$, then $E(X|F_0) = E(X)$ for any X integrable

(3) (Linearity) Let $X, Y \in L^1(\Omega, F, P)$. Then, $E(\alpha X + \beta Y \mid G) = \alpha E(X \mid G) + \beta E(Y \mid G)$ for any $\alpha, \beta \in R$

(4) (Tower Property) Suppose $H \subseteq G \subseteq F$ (H contains less information than G) and $X \in L^1(\Omega, F, P)$.

Then, $\boxed{E(X \mid H) = E(E(X \mid G) \mid H) = E(E(X \mid H) \mid G)}$ (smaller set wins!)

(5) (Take Out What You Know) Suppose Y bdd and measurable on G, and that $X \in L^1(\Omega, F, P)$. Then, $E(XY \mid G) = YE(X \mid G)$

Note on Tower Property: **First equality can be checked via definitions. Second equality is trivial since G is richer than H, and knowing G gives us the value of H, and thus E(X|H) is a constant given G so pull it out. When we take H to be the trivial sigma-field, then we get the law of iterated expectations!**

d. (P.2.3.10) Jensen's Inequality for Conditional Expectations

Let f:R $\to$ R be a convex function. Suppose $X \in L^1(\Omega, F, P)$ is s.t. $E|f(X)| < \infty$. Then, $E(f(X) \mid G) \geq f(E(X \mid G))$

e. (T.2.3.13,T.2.3.14) Monotone and Dominated Convergence for Conditional Expectations
(see notes: same applies)

f. (E.2.3.8) Conditional Variance Identity

Let $Var(Y|G) = E\left(Y^2 \mid G\right) - E(Y \mid G)^2$ where $Y \in L^2(\Omega, F, P)$ and G a σ-field. Then, $Var Y = E(Var(Y \mid G)) + Var(E(Y \mid G))$.

Chapter 3: General Theory of Stochastic Processes

I. Definitions:

a. (D.3.1.1) Stochastic Process

Given (Ω, F, P), a stochastic process $\{X_t\}$ is a collection $\{X_t : t \in I\}$ of RVs where the index t belongs to the index set I.

If I is an interval in R, we say it's continuous time. If I a subset of Z_+ we say it's discrete time.

Fixing ω, we call $t \mapsto X_t(\omega)$ the sample function or sample path of the S.P.

Note on Sigma Field Generated by S.P.: Clearly, all RV's in this collection are defined on a common probability space.

b. (D.3.1.2) Random Walk

Given (Ω, F, P), a stochastic process $\{X_t\}$ is a collection $\{X_t : t \in I\}$ of RVs where the index t belongs to the index set I.

If I is an interval in R, we say it's continuous time. If I a subset of Z_+ we say it's discrete time.

Fixing ω, we call $t \mapsto X_t(\omega)$ the sample function or sample path of the S.P.

A random walk is the sequence $S_n = \sum_{i=1}^{n} \xi_i$ where ξ_i are iid RV defined on the same probability space (Ω, F, P).

Special Case: If $\xi_i \in \{-1,1\}$, then S_n is a simple random walk.

If $\xi_i \in Z$, then S_n is a random walk on the integers

c. (E.3.1.4) S.P. with Independent Increments

S.P. $\{X_t\}$ has independent increments if X_{t+h}-X_t is independent of $\sigma(X_s, 0 \le s \le t)$

d. (D.3.1.5)Finite Dimensional Distribution Function:

Given $N < \infty$ and a collection $t_1, t_2, ..., t_N$ in I, we denote the joint distribution of $\left(X_{t_1}, ..., X_{t_N}\right)$ by $F_{t_1,...,t_N}(.)$,

so $F_{t_1,...,t_N}(\alpha_1, ..., \alpha_N) = P\left(X_{t_1} \le \alpha_1, ..., X_{t_N} \le \alpha_N\right) \ \ \forall \alpha_1, ..., \alpha_N \in \Re.$

We call the collection of functions $F_{t_1,...,t_N}(.)$ the finite dimensional distributions of the S.P.

Note on the limitations of FDD: This is the S.P. analog of the distribution function, which always exist. However, FDD does not tell us about some important properties of the S.P. such as the continuity of its sample path. We can have S.P. with same FDD but one is continuous, the other is not.

EX: (See notes)

e. (D.1.3.12) Consistency of FDD

We say that a collection of finite dimensional distributions is consistent if

$$\lim_{\alpha_k \uparrow \infty} F_{t_1,...,t_N}(\alpha_1, ..., \alpha_N) = F_{t_1,...,t_{k-1},t_k,...t_N}(\alpha_1, ..., \alpha_{k-1}, \alpha_k, \alpha_N)$$

for any $1 \le k \le N$, $t_1 < ... < t_N \in I$ and $\alpha_i \in R$

Note on S.P.'s and Consistency of FDD: The FDD of any SP must be consistent (and converse also true). This is intuitive because FDD is a joint distribution function of a subset of the random variables, and thus if we integrated one out, it would give us the joint distribution function of the rest.

f. (D. 3.1.14) Sigma-Field of a S.P.

(See notes for construction)

g. (D.1.3.7) Versions:

S.P.'s $\{X_t\}, \{Y_t\}$ are versions of one another if they have the same f.d.d.

h. (D.1.3.8) Modifications:

A S.P. $\{Y_t\}$ is a modification of another S.P. $\{X_t\}$ if $P\left(\{\omega : X_t(\omega) = Y_t(\omega)\}\right) = 1 \ \forall t \in I$

Note on Version vs. Modification: Modification is stronger. It implies versions. Two S.P. are modifications implies also that they're on the same probability space. But two stochastic processes can be versions of each other and not necessarily defined on the same probability space.

i. (E.3.3.7) Indistinguishable sample path (a.s.):

$\{X_t\}, \{Y_t\}$ have indistinguishable sample paths (a.s.) if $P\left(\{\omega : X_t(\omega) = Y_t(\omega) \ \forall t \in I\}\right) = 1$

j. (D.3.2.1) Characteristic Function of a Random Vector

A random vector X=$\left(X_1, X_2, ..., X_n\right)$ with values in R^n has the characteristic function

$$\Phi_X\left(\underline{\theta}\right) = E\ \ e^{i\underline{\theta}'X} \ = E\ \ e^{i\left(\sum_j \theta_j X_j\right)} \qquad \text{where } \underline{\theta} = \left(\theta_1, ..., \theta_n\right) \in R^n \text{ and } i = \sqrt{-1}$$

$$= E\ \ e^{\cos\left(\sum_j \theta_j X_j\right) + i \sin\left(\sum_j \theta_j X_j\right)}$$

Note on Existence: Unlike moment generating functions, characteristic functions always exist.

$\Phi_{\mathrm{X}}:\mathrm{R}^n \mapsto \mathrm{C}$ exists for any $\underline{\mathrm{X}}$ because $\sum_j \theta_j X_j = \cos\ \sum_j \theta_j X_j\ + i \sin\ \sum_j \theta_j X_j$ has both real and imaginary parts that are bounded and hence integrable R.V.'s.

Note on Properties: $\Phi_X\left(\underline{0}\right)$=1 and $\left|\Phi_X\left(\underline{\theta}\right)\right| \le 1\ \forall \underline{\theta} \in R^n$ (by prop or sin and cos)

Note on Usefulness: Characteristic function uniquely determines the law of a random variable. (IFF!)

Note on Fourier Transforms:

When the R.V. X has a PDF f_X, then the characteristic function is $\Phi_X\left(\theta\right) = E\ \ e^{i\theta X} \ = \int_{-\infty}^{\infty} e^{i\theta x} f_X(x) dx$

which is just the Fourier transform of the density function f_X.

k. (D.3.2.4) PDF of a Random Vector

A random vector $X = \left(X_1, X_2, ..., X_n\right)$ has a PDF $f_{\underline{X}}$ if for every $a_i < b_i\ \forall i$

$$P\left(a_1 \le X_1 \le b_1, ..., a_n \le X_n \le b_n\right) = P\left(\left\{\omega : a_i \le X_i\left(\omega\right) \le b_i, i = 1, 2, ..., n\right\}\right) = \int_{a_1}^{b_1} ... \int_{a_n}^{b_n} f_{\underline{X}}\left(x_1, ..., x_n\right) dx_n ... dx_1$$

Note: Such a function (also called the joint density of X(1)…X(n)) must be a non-negative Borel measurable function that integrates out to 1.

l. (D.3.2.6) Positive Semi-definite Matrix

An nxn matrix **A** with entries A_{jk} is called p.s.d. if $\forall$ nonzero vector $\theta \in \mathrm{R}^n$, $\theta' \mathrm{A} \theta \ge 0$.

m. (D.3.2.7) Gaussian Distribution

We say that a random vector $X = \left(X_1, X_2, ..., X_n\right)$ has a Gaussian distribution if

$$\Phi_X\left(\underline{\theta}\right) = \ E\ \ e^{i\underline{\theta}'X} \ = e^{-\frac{1}{2}\theta'\Sigma\theta} e^{i\theta'\mu}$$

for some psd symmetric nxn matrix Σ (VarCov matrix), some μ=$(\mu_1, ..., \mu_n)$ (vector of means) and all $\theta \in \mathrm{R}^n$

Note on R:

A RV X is Gaussian if for some $\mu \in R, \sigma^2 \ge 0$, and all $\theta \in \mathrm{R}$, E $e^{i\theta X} \ = e^{-\frac{1}{2}\theta^2\sigma^2 + i\theta\mu}$

Note on Non-Degeneracy (D.3.2.8):

We say that X has a nondegenerate Gaussian distribution if the matrix Σ is invertible, or, when Σ is strictly positive definite.

What's the big fuss? The definition allows for a non-invertible VarCov matrix so to include non-random constants as (technically) Gaussian, even though it does not have a density. But this is done for technical reasons, so that the set of Gaussian distributions is closed in L2, i.e. contains all its limit points. (For example, a sequence of normal RV's s.t. X(n) = N(0,1/n) converges to the limit 0, a non-random constant. With this definition, 0 is Gaussian, and the limit point is contained in the space).

n. (P.3.2.5) Gaussian Density

A random vector X with a nondegenerate Gaussian distribution has the density

$$f_{\underline{X}}(\underline{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det\Sigma}} \exp\left(-\frac{1}{2}(\underline{x}-\underline{\mu})^T \Sigma^{-1}(\underline{x}-\underline{\mu})\right)$$

In particular, if $\sigma^2 > 0,$ then a Gaussian RV has density

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

o. (D.3.2.14) Gaussian Stochastic Process

A stochastic process $\{X_t, t \in I\}$ is Gaussian if for all n<∞, and all $t_1,...,t_n \in I$, the random vector $\left(X_{t_1},...,X_{t_n}\right)$ has a Gaussian distribution.

i.e. All FDD's of the process are Gaussian.

Note on Distributional Properties (C.3.2.15):

All distributional properties of a Gaussian processes are determined by the mean μ(t)=E(X_t)

of the process and its autocovariance function $\rho(t,s) = E\left[(X_t - \mu(t))(X_s - \mu(t))\right]$

This is not surprising because the distributional properties are characterized uniquely by the characteristic function, which in this case

Takes the form $\Phi_X(\underline{\theta}) = E\left[e^{i\underline{\theta}'X}\right] = e^{-\frac{1}{2}\theta'\Sigma\theta} e^{i\theta'\mu}$, which is parameterized by the mean vector and the Var-Cov matrix. Elements of the mean vector and var-cov function are exactly what μ(t) and $\rho(t,s)$ are.

p. **(D.3.2.19) Strong and Covariance/Weak Stationary Process**

Strong

A stochastic process $\{X_t, t \in R\}$ is called (strong) stationary if its FDD's satisfy: $\forall \tau \in R, \alpha_i \in R$ i=1,...,N, and $t_1 < ... < t_N \in R$

$$F_{t_1,...,t_N}(\alpha_1,...,\alpha_N) = P\left(X_{t_1} \le \alpha_1,..., X_{t_N} \le \alpha_N\right) = P\left(X_{t_{1+\tau}} \le \alpha_1,..., X_{t_{N+\tau}} \le \alpha_N\right) = F_{t_{1+\tau},...,t_{N+\tau}}(\alpha_1,...,\alpha_N)$$

Weak

A stochastic process $\{X_t, t \in R\}$ is called (weak/covariance) stationary if

μ(t)=μ (non-random constant) and $\rho(t,s) = r(|t-s|)$ (i.e. a function of the time difference).

Note on Discrete Time Definitions:

Strictly Stationary Processes: A stochastic process/sequence of RV's (e.g. a time series) $\{z_i\}$ $(i = 1,2...)$ is (strictly) stationary if

$$F(Z_t,...,Z_{t+K}) \text{ does not depend on } t \text{ for all } K = \{0,1,2,...\} \Leftrightarrow F(Z_1,...,Z_K) = F(Z_{t+1},...,Z_{t+K}) \text{ for all } t \text{ and } K$$

→ **F(Z_t) does not depend on t! (all observations come from same distribution. Though this says identical but not necessarily independent)**

Prop: If $\{z_i\}$ $(i = 1,2...)$ (strictly) stationary, then $\{g(z_i)\}$ $(i = 1,2...)$ (strictly) stationary for some cont. function g[2].

Weakly Stationary Process: A stochastic process/sequence of RV's $\{z_i\}$ $(i = 1,2...)$ is weakly (or covariance) stationary if:

E($\mathbf{z}_i$) does not depend on *i* and

Cov($\mathbf{z}_i$,$\mathbf{z}_{i-j}$) exists, is finite, and depends only on *j* but not on *i* (e.g. Cov($\mathbf{z}_1$,$\mathbf{z}_5$) =Cov($\mathbf{z}_{12}$,$\mathbf{z}_{16}$))

II. Theorems/Propositions

a. Modification →Versions

b. X(t) , Y(t) modifications of each other does not imply that they're indistinguishable.

c. If X(t), Y(t) modifications of each other and are either a) discrete time or b) right continuous continuous time, then they are also indistinguishable.

d. (P.3.2.2) Characteristic Function Uniquely Determines the Law P_X of a Random Vector

The characteristic function determines the law of a random vector. That is, if $\Phi_X(\underline{\theta}) = \Phi_Y(\underline{\theta})\ \forall \underline{\theta} \in R^n$, then X has the same law as Y.

(i.e. has the same probability measure on R^n, and thus, the same distribution.)

Note on MGF's: The law of a non-negative RV is uniquely determined by MGF. However, MGF is not defined for many RV's, and thus not very useful.

e. (P.3.2.5) Properties of Characteristic Functions

(1) (Indepdence) If $X = (X_1, X_2, ..., X_n)$ mutually independent RV's iff

$$\Phi_X(\underline{\theta}) = E\left[e^{i\underline{\theta}'X}\right] = E\left[\prod e^{i\theta_j X_j}\right] = \prod E\left(e^{i\theta_j X_j}\right)$$

(2) (Linearity) For a,b∈ R, $\Phi_{aX+b}(\theta) = e^{i\theta b}\Phi_X(a\theta)$

(3) (Gaussian) $\boxed{\Phi_X(\underline{\theta}) = E\left[e^{i\underline{\theta}'X}\right] = e^{-\frac{1}{2}\theta'\Sigma\theta}e^{i\theta'\mu}}$

f. (P.3.2.9,P.3.2.10,E.3.2.11,P.3.2.12,P.3.2.13) Properties of Gaussian Random Vectors

(1) Uncorrelated $\Rightarrow$ Independent: If X,Y jointly Gaussian, then $\text{Cov}(X,Y) = 0 \Rightarrow X \perp Y$

(Converse not true. See 3.2.11. i.e. you can have Gaussian RV's that are not jointly Gaussian, have cov 0, and not independent)

(2) $\underline{X}_{nx1}$ Gaussian $\Leftrightarrow \sum_{i=1}^{n}\alpha_{ji}X_i,\ j = 1,..,m$ is a m-dimensional Gausian random vector, for linear combinations $\left\{\underline{\alpha}_{j(nx1)}\right\}_{j=1}^{m}$.

(i.e. any sub-vector composed of linear combinations of the elements of a Gaussian vector is also Gaussian.)

(3) $\underline{X}$ Gaussian has PDF: $f_{\underline{X}}(\underline{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det\Sigma}}\exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu})^T\Sigma^{-1}(\underline{x}-\underline{\mu})\right]$

(4) Distribution of $\underline{X}$ Gaussian has parameters Σ and $\underline{\mu}$, the variance-covariance matrix and the mean vector, respectively.

(5) Closure: If a sequence of n-dimensional Gaussian random vectors $\left\{X^{(k)}\right\}_{k\geq 1}$ converges in L2 to a n-dimensional vector X,

then, X is a Gaussian random vector whose parameters μ and Σ are the limits of the corresponding parameters of

the sequence of random vectors, $\mu^{(k)}$ and $\Sigma^{(k)}$ of $X^{(k)}$.

A stochastic process $\{X_t, t \in I\}$ is Gaussian if for all n<∞, and all $t_1,...,$

g. (C.3.2.15,P.3.2.16)Properties of Gaussian Processes

(1) μ(t) and $\rho(t,s)$ of a Gaussian SP: All distributional properties of a Gaussian processes are determined by the mean $\mu(t)=E(X_t)$

of the process and its autocovariance function $\rho(t,s) = E\left[(X_t - \mu(t))(X_s - \mu(t))\right]$

(2) If the SP $\{X_t, t \in I\}$ and the Gaussian S.P. $\left\{X_t^{(k)}, t \in I\right\}$ (a sequence of collections of RVs, or sequence of functions from IxΩ → R)

are such that for each fixed t∈ I, $E\left[\left|X_t - X_t^{(k)}\right|^2\right] \to_{k\to\infty} 0$ (each function converges pointwise on t, based on $\|.\|_2$ metric),

then, X_t is a Gaussian S.P. with the same mean and auto-covariance functions as the pointwise limits of $\left\{X_t^{(k)}, t \in I\right\}$.

(3) Stationarity: A Gaussian SP is stationary iff it is covariance/weak stationary

i.e. $\mu(t)=\mu$ (non-random constant) and $\rho(t,s) = r(|t-s|)$ (i.e. a function of the time difference).

Note on (2): The proof here is not difficult to see, given what we know about the closure of a Gaussian random vector, and given that we know that a SP is Gaussian iff all FDD's are Gaussian. That is, take any arbitrary index set, and look at the sequence of functions over the sub-index (forming a vector since now countable). We know that the vector is Gaussian, that it converges to the corresponding sub-set of the SP X, which must be Gaussian by the closure of Gaussian random vectors in L2. Since we picked an arbitrary index set, then it must be true that all FDD's are Gaussian.

Note on Random "Process" and Random Vector: When we say "process", t in I may not be finite, or countable. Whereas when we say "vector" we are assuming countable (and often finite).

h. (T.3.3.16) Fubini's Theorem

If X_n is measurable SP, then for a.e. ω, the function t $\mapsto X_t(\omega)$ is measurable from R to R.

Moreover, for any interval I s.t. $\int_I E\,|\,X_t\,|\,dt < \infty$, we have that $\int_I X_t dt < \infty$, $\int_I EX_t dt = E\int_I X_t dt$ (can exchange the order of integration).

Further, if E$\left|X_t\right| < \infty$ for all t $\in$ I then the function t $\mapsto EX_t$ is Borel measurable (in I)

Chapter 4: Martingales and Stopping Times

(Martingales is a collection of S.P. – Random Walk and Brownian Motion are examples)

Definitions

a. **(D.4.1.1) Filtration: A (discrete time) filtration is a non-decreasing family of sub-sigma-fields of $\{F_n\}$ of our measurable space (Ω, F)**

That is, $F_0 \subseteq F_1 \subseteq ...F_n \subseteq ... \subseteq F$ where F_n is a sigma field for each n.

Note on Usefulness: A filtration represents any procedure of collecting more and more information as time goes on. Given a filtration, we're interested in S.P. s.t. for each n the information gathered by that time suffices for evaluating the value of the n-th element of the process.

b. **(D.4.1.2) S.P. Adapted to Filtration: A (discrete time) S.P. $\{X_n, n = 0,1,...\}$ is adapted to a filtration $\{F_n\}$ if $\omega \mapsto X_n(\omega)$ is a R.V. on (Ω, F_n) for each n. i.e.) $\sigma(X_n) \subseteq F_n$ for each n.**

Note: It can be shown that $\{X_n, n = 0,1,...\}$ is adapted to a filtration $\{F_n\}$ iff $\boxed{\sigma(X_0,..., X_n) \subseteq F_n \quad \forall n}$

c. **(D.4.1.3) Minimal Filtration / Canonical Filtration**

The filtration $\{G_n\} = \sigma(X_0,..., X_n)$ is the minimal filtration with respect to which $\{X_n\}$ is adapted.

We call it the canonical filtration for the S.P. $\{X_n\}$. i.e. $\forall \{A_n\}$ s.t. S.P.$\{X_n\}$ is adapted to $\{A_n\}$, $A_n \subseteq G_n$

d. **(D.4.1.4,D.4.2.1) Martingale**

(D.4.1.3) Discrete Time

A martingale (denoted MG) is a pair (X_n, F_n), where $\{F_n\}$ is a filtration and X_n an integrable (i.e. $E|X_n| < \infty$) S.P. adapted to this filtration s.t.

$$\boxed{E\left[X_{n+1} | F_n\right] = X_n \quad \forall \text{n, a.s.}}$$

Note: If you're MG wrt a filtration, then you're a MG w.r.t. a "slower" filtration.

The slower a filtration $n \mapsto F_n$ grows, the easier it is for an adapted S.P. to be a martingale. That is, if $H_n \subseteq F_n$ $\forall$n, and S.P. $\{X_n\}$ is adapted to filtration $\{H_n\}$ is s.t. (X_n, F_n) is MG, then (X_n, H_n) also a MG.

Pf: Use Tower Property: $E(X_{n+1} | H_n) = E\left[E(X_{n+1} | F_n) | H_n\right]$ by Tower $= E(X_n | H_n) = X_n$ since $E(X_{n+1} | F_n) = X_n$

Note on the Meaning of Filtration: $E\left[X_{n+1} | F_n\right] = E\left[X_{n+1} | X_n, X_{n-1},..., X_0\right]$

(D.4.2.1) Continuous Time

The pair (X_t, F_t), $t \geq 0$ real-valued, is called a continuous time martingale (in short MG) if:

(a) The σ-Fields $F_t \subseteq F$, $t \geq 0$ form a continuous time filtration.

That is, $\boxed{F_t \subseteq F_{t+h} \ \forall t \geq 0 \text{ and } h>0}$

(b) The continuous time S.P. $\{X_t\}$ is integrable and adapted to this filtrationa.

That is, $\boxed{E|X_t| < \infty \text{ and } \sigma(X_t) \subseteq F_t \quad \forall t \geq 0}$

(c) For any fixed $t \geq 0$ and h>0, the identity $\boxed{E(X_{t+h} | F_t) = X_t}$

e. **(D.3.1.14) L2 Martinglale**

A L^2-MG (or square-integrable MG) is a MG $\{X_n\}$ s.t. $E\left(X_n^2\right) < \infty$ $\forall$n.

f. **Martingale Differences**

$D_n = X_n - X_{n-1}$ for $n \geq 1$ and $D_0 = X_0$ are called the martingale differences associated with a martingale $\{X_n\}$.

g. **(D.4.1.11) Previsible / Predictable Process**

We call a sequence $\{V_n\}$ previsible (or predictable) for the filtration $\{F_n\}$ if V_n is measurable on F_{n-1} for all $n \geq 1$.

h. (D.4.1.16) Orthogonal Sequence of R.V.'s

We say that $D_n \in L^2(\Omega, F, P)$ is an orhogonal sequence of RV if

$$E\left[D_n h(D_0, D_1, ..., D_{n-1})\right] = E\left[D_n\right] E\left[h(D_0, D_1, ..., D_{n-1})\right]$$

for any $n \geq 1$ and every Borel function $h: R^n \mapsto R$ s.t. $E\left[h(D_0, D_1, ..., D_{n-1})^2\right] < \infty$

Equivalently,

$$E\left[D_n \mid D_0, ..., D_{n-1}\right] = E\left[D_n\right]$$

i. (D.4.1.18) Super-martingales and Sub-martingales

A submartingale (denoted subMG) is an integrable SP $\{X_n\}$, adapted to the filtration $\{F_n\}$, s.t.

$$E\left[X_{n+1} \mid F_n\right] \geq X_n \quad \forall n \ a.s. \quad \text{i.e. } P\left(\left\{\omega : E\left[X_{n+1} \mid F_n\right](\omega) - X_n(\omega) < 0\right\}\right) = 0$$

Alternatively: $\{X_n\}$ subMG iff $\{-X_n\}$ superMG

A super-martingale (denoted superMG) is an integrable $\{X_n\}$ adapted to the filtration $\{F_n\}$ s.t.

$$E\left[X_{n+1} \mid F_n\right] \leq X_n \quad \forall n \ a.s. \quad \text{i.e. } P\left(\left\{\omega : E\left[X_{n+1} \mid F_n\right](\omega) - X_n(\omega) > 0\right\}\right) = 0$$

Alternatively: $\{X_n\}$ subMG iff $\{-X_n\}$ superMG

Note on SubMG, SupMG, and MG: All the results about SubMG are dual stamnts for supMG's and vice versa. Further, every MG is both a SubMG and SupMG by definition, therefore all the results holding for either SubMG or SupMG hold for MG as well.

j. **(D.4.2.8) Last Element of a subMG or supMG**

We say that a subMG (X_t, F_t) has a last element (X_∞, F_∞) if $F_t \subset F_\infty$, the integrable RV X_∞ is measurable on F_∞ and for each $t \geq 0$, a.s. $E(X_\infty \mid F_t) \geq X_t$.

(similarly for supMG, but with $E(X_\infty \mid F_t) \leq X_t$)

Note: Last element preserves sub/supMG

k. **(D.4.2.9) Right-Continuous Filtration**

A filtration is called right-filtration if for any $t \geq 0$, $\bigcap_{h>0} F_{t+h} = F_t$

Note1: We usually assume without proof that continuous time filtrations are right-continuous.

Note2: Note that not all filtrations are right-continuous. Consider $F_0 = \{\varnothing, \Omega\}$ while $F_h = F$ (where $F \neq F_0$) is not right-continuous at t=0.

bc $\bigcap_{h>0} F_h = F \neq F_0$

l. (D.4.3.1, D.4.3.11) Stopping Time t for a filtration {F$_t$}
Discrete Time

A RV τ taking values in $\{0, 1, ..., n, ..., \infty\}$ is a stopping time for the filtration $\{F_n\}$ if the event $\{\omega : \tau(\omega) \leq n\} \in F_n$ for each finite $n \geq 0$

(set of w's s.t. the process stops before time n)
Note: This means that the filtration contains information about whether or not the process has stopped.

Continuous Time

A non-negative RV $\tau(\omega)$ is called stopping time wrt the continuous filtration $\{F_t\}$ if $\{\omega : \tau(\omega) \leq t\} \in F_t$ for all $t \geq 0$.

Note on Min (M,N), Max(M,N) and M+N as Stopping Time:
It can be shown (E.4.3.3) that if M,N stopping time, then M+N and Min(M,N) and Max(M,N) also is.

Note on expressing {w:t(w)<u><</u>n} to check stopping time (DISCRETE TIME):

Let $\tau(\omega) = \min(k \geq 0 : X_k(\omega) \in B)$ for some set $B \in$ Borel

$\{\omega : \tau(\omega) \leq n\} = \bigcup_{i=1}^{n} \{\omega : X_i(\omega) \in B\}$

Idea: LHS is the set of outcomes such that the process is stopped by n. RHS is the union of the set of all such outcomes.

Note on Showing Stopping Time for Continuous Time:
Use Prop 4.3.13 – I f t→X(t) is continuous a.s. and B is a closed set, then t(w) = inf(t≥0:X(t) in B) is a stopping time!

Note on Diff Bt Discrete & Cont Time First Hitting Time: For discrete , FHT is a stopping time. For continuous, not necessarily, esp if B is not closed or sample path not continuous (see bottom of pg. 83).

m. **(D.4.3.5) Stopped Process**

Using the notation $n \wedge \tau = \min(n, \tau(\omega))$, the stopped at τ stochastic process $\{X_{n\wedge\tau}\}$ is given by

$$X_{n\wedge\tau}(\omega) = \begin{cases} X_n(\omega), & n \le \tau(\omega) \\ X_{\tau(\omega)}(\omega), & n > \tau(\omega) \end{cases}$$

n. **(D.4.4.2) Innovation Process**

When using Doob's decomposition for the canonical filtration $\sigma(X_k, k \le n)$, the MG $\{Y_n\}$ is called the innovation process associated with $\{X_n\}$.

Note: The reason for this name is that $X_{n+1} = Y_{n+1} + A_{n+1} = (A_{n+1} + Y_n) + (Y_{n+1} - Y_n)$, where $A_{n+1} + Y_n$ is measurable on $\sigma(X_k, k \le n)$ while $Y_{n+1} - Y_n$ describes the "NEW" part of X_{n+1}.

o. **(D.4.4.8) Increasing Part of a M.G. $\{M_t\}$**

The S.P. $\{A_t\}$ in the Doob-Meyer decomposition of $\{M_t^2\}$ is called the increasing part or the increasing process associated with the MG $\{M_t\}$.

p. **(D.4.6.1) Branching Process** (We use MGs to study the extinction probabilities of branching processes)

The branching process is a discrete time S.P. $\{Z_n\}$ taking non-negative integer values, s.t.

$$Z_0 = 1 \text{ and } Z_n = \sum_{j=1}^{Z_{n-1}} N_j^{(n)} \quad \forall n \ge 1$$

where N and $N_j^{(n)}$ for j=1,2,... are iid non-negative integer valued R.V. with finite mean $m = E(N) < \infty$,

and where we use the convention that if $Z_{n-1} = 0$ then also $Z_n = 0$.

Note on Interpretation:

The S.P. $\{Z_n\}$ is interpreted as counting the size of an evolving population so that Z_n is the size of the n^{th} generation.

The S.P. $N_j^{(n)}$ is the number of offsprings of j^{th} individual of generation (n-1).

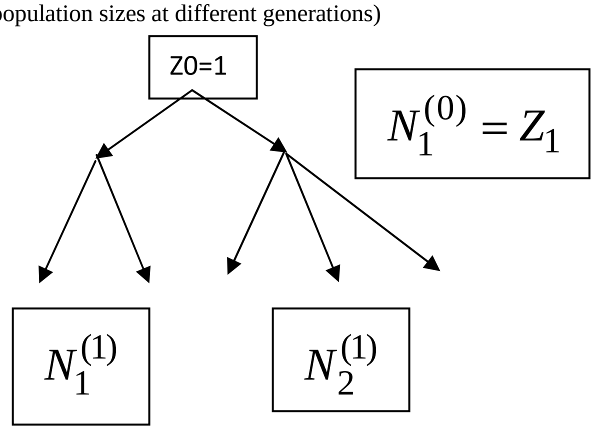
Associated with the branching process is the family tree with the root denoting the 0-th generation and having $N_j^{(n)}$ edges from the vertex j at distance (n-1) from the root to vertices of distance n from the root.

Random trees generated in such a fashion are called Galton-Watson trees.

We shall use throughout the filtration $F_n = \sigma\left(\left\{N_j^{(k)}, k \le n, j = 1, 2, ...\right\}\right)$

Note that in general, $G_n = \sigma(\{Z_k, k \le n\}) \subset F_n$

(STRICT subset, since we cannot in general recover the number of offsprings of each individual knowing only the total population sizes at different generations)



q. **Probability of Extinction**

$p_{ex} = P(Z_n = 0 \text{ for some } n \ge 0)$

a. $\{X_n, n=0,1,\ldots\}$ **is adapted to a filtration** $\{F_n\}$ **iff** $\boxed{\sigma(X_0,\ldots,X_n)\subseteq F_n \quad \forall n}$

b. **(P. 4.1.7):**

If $X_n = \sum_{i=0}^{n} D_i$ then the canonical filtration for $\{X_n\}$ is the same as the canonical filtration for $\{D_n\}$.

Furthermore, (X_n, F_n) is a martingale iff $\{D_n\}$ is an integrable S.P., adopted to $\{F_n\}$ s.t. $\mathrm{E}(D_{n+1} \mid F_n) = 0$ a.s. $\forall$n.

c. **(T.4.1.12) : Martingale Transform**

Let (X_n, F_n) be a MG and $\{V_n\}$ be a previsible sequence for the same filtration. The sequence of R.V.

$$\mathrm{Y}_n = \sum_{k=1}^{n} V_k \left(X_k - X_{k-1}\right)$$

called the martingale transform of V. wrt X. , is then a MG with respect to the filtration $\{F_n\}$, provided $|V_n| \le C_n$

for some non-random constatnst $C_n < \infty$, or more generally $\mathrm{E}|V_n|^q < \infty$ for all n and some $1 \le$ p, q<∞ s.t. $\frac{1}{q} + \frac{1}{p} = 1$.

d. **(P.4.1.15,P.4.1.17) Alternative and Equivalent Characterizations of M.G. S.P.**

(P.4.1.15)

A S.P. $\mathrm{X}_n \in L^2(\Omega, F, P)$ adapted to the filtration $\{F_n\}$ is a MG iff $\mathrm{E}\left[(X_{n+1} - X_n)Z\right] = 0$ for any $\mathrm{Z} \in L^2(\Omega, F_n, P)$

Note on Intuition: This result follows from the fact that Xn is a MG iff $\mathrm{E}[X_{n+1}|F_n]=X_n$. But the idea is that no random variables that live in the space $L^2(\Omega, F_n, P)$ is "correlated" (in the sense of having a non-zero inner product, or non-orthogonal in the L2 norm) to the difference between X_{n+1}-X_n. Because if so, then it means that some random variable in the space $L^2(\Omega, F_n, P)$ is correlated with the difference, and hence we know something about X_{n+1}.

(P.4.1.17)

A S.P. $\mathrm{X}_n \in L^2(\Omega, F, P)$ is a MG for its canonical filtration iff it has an orthogonal, zero-mean differences sequence $D_n = X_n - X_{n-1},\ n \ge 1$

Note on Intuition: This is just a simple reformulation of P.4.1.15 above, using the definition of orthogonal sequence of R.V.

Note on Application to Gaussians:

From above, a necessary condition for the MG property is to have $\mathrm{E}(D_n) = 0$ and $\mathrm{E}(D_n D_i) = 0\ \forall 0 \le i < n$.

With the Gaussian vector $D = (D_0,\ldots,D_n)$ having uncorrelated elements, we know that the var-cov matrix Σ is diagonal

and the characteristic function $\Phi_{\mathbf{D}}(\theta) = \prod_{k=0}^{n} \Phi_{D_k}(\theta_k)$. This is true iff D_k's are independent (by P.3.2.5).

Thus, for a Gaussian S.P. having independent, orthogonal or uncorrelated differences are equivalent properties, which together with each of these differences having a zero mean is also equivalent to the MG condition.

e. **(R.4.1.21) SubMG (SupMG) have non-decreasing (non-increasing) expectation E(Xn)**

If $\{X_n\}$ a subMG, then necessarly $n \mapsto \mathrm{E}(X_n)$ is non-decreasing. (non-increasing for supMG)

Pf: $E[X_n] = E[E(X_n \mid F_{n-1})] \ge E[X_{n-1}]$ since $E(X_n \mid F_{n-1})(\omega) \ge X_{n-1}(\omega)\ \forall \omega \in$ A s.t P(A)=1

$$\text{So, } \int_{\omega\in A} E(X_n \mid F_{n-1})(\omega) dP(\omega) \ge \int_{\omega \in A} E(X_n \mid F_{n-1})(\omega)$$

f. **(T.4.3.6) Preservation of (sub/sup)MG of Stopped Processes**

Discrete Time

If (X_n, F_n) is a subMG (or supMG or a MG), and τ is a stopping time for $\{F_n\}$, then $(X_{n\wedge\tau}, F_n)$ is also a subMG (or supMG or MG).

Pf:

$$X_{n\wedge\tau}(\omega) = \begin{cases} X_n(\omega), & n \le \tau(\omega) \\ X_{\tau(\omega)}(\omega), & n > \tau(\omega) \end{cases} \Rightarrow E(X_{n\wedge\tau} \mid F_n) = \begin{cases} X_n(\omega), & n \le \tau(\omega) \\ E\left(X_{\tau(\omega)}(\omega) \mid F_n\right), & n > \tau(\omega) \end{cases} \ge \begin{cases} X_n(\omega), & n \le \tau(\omega) & \text{from previous} \\ X_n(\omega), & n > \tau(\omega) & \text{from the fact that } X_n \text{ is subMG} \end{cases}$$

(C.4.3.7)

If (X_n, F_n) is a subMG and τ is a stopping time for $\{F_n\}$, then $E(X_{n\wedge\tau} \mid F_n) \ge X_0\ \forall n$.

If in addition (X_n, F_n) is a MG, then $E(X_{n\wedge\tau}) = E(X_0)$.

Continuous Time

f τ is a stopping time for the filtration $\{F_t\}$ and the S.P. $\{X_t\}$ of right-continuous sample path is a subMG (or supMG or a MG) for $\{F_t\}$, then $X_{t\wedge\tau} = X_{t\wedge\tau(\omega)}(\omega)$ is also a subMG (or supMG or MG) for this filtration.

g. (P.4.3.13) Checking Stopping Time for a Continuous-Time Process

If the sample path $t \mapsto X_t(\omega)$ is continuous $\forall\omega\in\Omega$ and B is a closed set, then $\tau_B(\omega) = \inf\{t \ge 0 : X_t(\omega) \in B\}$ is a stopping time for the canonical filtration $G_t = \sigma(X_s, s \le t)$.

Note on Proof: Details in pg 83.

$\{\omega : \tau_B(\omega) \le t\} = \bigcup_{s\le t}\{\omega : X_s(\omega) \in B\}$

Now, each $\{\omega : X_s(\omega) \in B\} \in G_t$. However, we have an uncountable union, and therefore we don't know if this is in fact in G_t.

It can be shown that $\bigcup_{s\le t}\{\omega : X_s(\omega) \in B\} = \bigcap_{k\in Q_t}\bigcup_{s\in Q_t}\{\omega : X_s(\omega) \in B_k\} \in F_t$

since $\{\omega : X_s(\omega) \in B_k\} \in F_t$ and we have countable union and intersection

where Q_t is the set of all rational numbers in [0,t) together with t

and $B_k = \bigcup_{y\in B}\left(y - \frac{1}{k}, y + \frac{1}{k}\right)$ is an open set containing B (so that as $k \to \infty$, $B_k \to B$).

h. (T.4.3.8, T.4.3.16) Doob's Optional Stopping Theorem

(T.4.3.8) Discrete Time

If 1) (X_n, F_n) is a subMG, 2) $\tau < \infty$ a.s., 3) τ is a stopping time for the filtration $\{F_n\}$ and 4) the sequence $\{X_{n\wedge\tau}\}$ is uniformly integrable.

Then, $E(X_\tau) \ge E(X_0)$.

If in addition, 5) (X_n, F_n) is a MG, then $E(X_\tau) = E(X_0)$.

Pf: Observe that whenever $\tau(\omega) < \infty$, $X_{n\wedge\tau(\omega)}(\omega) \to_{n\to\infty} X_{\tau(\omega)}(\omega)$.

Since, $\tau < \infty$ a.s. (by assumption), thus $X_{n\wedge\tau(\omega)}(\omega) \to_{a.s.} X_{\tau(\omega)}(\omega) \Rightarrow X_{n\wedge\tau(\omega)}(\omega) \to_P X_{\tau(\omega)}(\omega)$

Also, we have $\left\{X_{n\wedge\tau(\omega)}(\omega)\right\}$ U.I.

Thus, we have $E(X_\tau)$=$E(\lim_{n\to\infty} X_{n\wedge\tau})$ by a.s. convergence above = $\lim_{n\to\infty}$E$(X_{n\wedge\tau})$ by Th.1.4.22

$X_{n\wedge\tau}$U.I. and converges in prob $\Rightarrow X_{n\wedge\tau} \to_{L_1} X_\tau$

$\ge E(X_0)$ by corollary 4.3.7 above, since $(X_{n\wedge\tau} \mid F_n)$ a subMG

(T.4.3.16) Continuous Time

If 1) (X_t, F_t) is a subMG with 2) right-continuous sample path ($t \mapsto X_t(\omega)$ is right continuous $\forall\omega\in$ A s.t. P(A) = 1)

3) $\tau < \infty$ a.s., 4) τ is a stopping time for the filtration $\{F_t\}$ and 5) the sequence $\{X_{t\wedge\tau}\}$ is uniformly integrable.

Then, $E(X_\tau) \ge E(X_0)$.

If in addition, 6) (X_t, F_t) is a MG, then $E(X_\tau) = E(X_0)$.

i. (E.4.3.9) Typical Example of Application of Doob's Optional Stopping

Consider the simple random walk $S_n = \sum_{k=1}^{n} \xi_k$ with $\xi_k \in \{-1,1\}$ iid s.t. $P(\xi_k = 1) = P(\xi_k = -1) = \frac{1}{2}$ so that $E(\xi_k) = 0$.

Let $\tau_{a,b}$=inf$\{n \geq 0 : S_n \notin (-a,b)\}$ (first time that S_n exits the interval).

We can show that...

1) (S_n, F_n) is a MG

2) S_n has right continuous sample paths

3) $\tau<\infty$ a.s.

4) $\tau_{a,b}$ is a $\{F_n\}-$Stopping Time

5) $|S_{n\wedge\tau}| \leq \max(a,b)$ so that $\{S_{n\wedge\tau}\}$ U.I.

Thus, by Doob's optional Stopping, we know that...

$E(S_\tau) = E(S_0) = 0$ where $S_\tau \in \{-a,b\}$ necessarily by construction.

Thus, $E(S_\tau) = \underbrace{-aP(S_\tau = -a)}_{S_t \text{ hits -a before b}} + \underbrace{bP(S_\tau = b)}_{S_t \text{ hits b before -a}} = -aP_a + b(1-P_a) = 0 \Rightarrow P_a = \frac{b}{a+b}$ and $P_b = \frac{a}{a+b}$

j. (E.4.3.17) A Proof for why t <inf

Let $\tau_{a.b} = \inf\{t \geq 0 : W_t \notin (-a,b)\}$

$P(\tau_{a.b} = \infty) = \lim_{n\to\infty} P(\tau_{a.b} > n)$

Now, $\tau_{a.b} \leq \tau_b = \inf\{t \geq 0 : W_t \geq b\}$

It's sufficient to show that $\tau_b < \infty$ a.s.

$P(\tau_b \leq T) = P(\max_{0\leq s\leq T} W_s \geq b)$ by reflection principle $= \frac{2}{\sqrt{2\pi}} \int_{\alpha/\sqrt{T}}^{\infty} \exp(-x^2/2)dx \to_{T\to\infty} 1$

Thus, $\tau_{a.b} \leq \tau_b < \infty$ a.s.

k. (T.4.4.1) Doob's Decomposition

Given an integrable S.P. $\{X_n\}$, adapted to a discrete parameter filtration $\{F_n\}$, $n \geq 0$, there exists a decomposition $X_n = Y_n + A_n$ s.t. (Y_n, F_n) is a MG and $\{A_n\}$ is a previsible S.P. This decomposition is unique up to the value of Y_0, a R.V. measurable on F_0.

Pf: Let $A_0 = 0$, and for all $n \ge 1$, let $A_n = A_{n-1} + E\left(X_n - X_{n-1} \mid F_{n-1}\right)$.

$A_n - A_{n-1} = E\left(X_n - X_{n-1} \mid F_{n-1}\right)$ is F_{n-1} measurable, by def of C.E.

Also, since $F_{k-1} \subset F_{n-1}\ \forall k \le n$ by def. of filtration, so $A_k - A_{k-1}$ is F_{n-1} measurable.

Now, $A_n = A_n \pm A_{n-1} \pm A_{n-2} \pm ... \pm A_0 = \left(A_n - A_{n-1}\right) + \left(A_{n-1} - A_{n-2}\right) + ... + \left(A_1 - A_0\right) + A_0 = A_0 + \sum_{k=1}^{n}\left(A_k - A_{k-1}\right)$

Then, $\{A_n\}$ is measurable w.r.t. $\{F_{n-1}\}\ \forall\ n \ge 1$ and thus a previsible process by def.

To check $Y_n = X_n - A_n$ is a martingale, need to show that $\left(Y_n, F_n\right)$, where $\{F_n\}$ is a filtration and Y_n integrable $\left(\text{i.e. } E\left|Y_n\right| < \infty\right)$

S.P. adapted to this filtration s.t. $E\left[Y_n \mid F_{n-1}\right] = Y_{n-1}$.

- *WTS* Y_n integrable. We know that X_n is integrable, by assumption, thus it suffices to show that A_n integrable.

  $A_n - A_{n-1} = E\left(X_n - X_{n-1} \mid F_{n-1}\right) \Rightarrow \left|A_n - A_{n-1}\right| = \left|E\left(X_n - X_{n-1} \mid F_{n-1}\right)\right|$

  $\Rightarrow E\left|A_n - A_{n-1}\right| = E\left|E\left(X_n - X_{n-1} \mid F_{n-1}\right)\right| \le EE\left(\left|X_n - X_{n-1}\right| \mid F_{n-1}\right)$ by Jensen's Inequality $= E\left(\left|X_n - X_{n-1}\right|\right) \le \infty$

  since X_n integrable $\forall n \Rightarrow X_n - X_{n-1}$ integrable.

  Now, we have $\{A_n - A_{n-1}\}$ and $\{X_n - X_{n-1}\}$ integrable.

  $E\left|Y_n\right| = E\left|X_n - A_n\right| = E\left|X_n - \left(X_{n-1} + X_{n-1}\right) + \left(A_{n-1} - A_{n-1}\right) - A_n\right| = E\left|\left(X_n - X_{n-1}\right) + \left(X_{n-1} - A_{n-1}\right) + \left(A_{n-1} - A_n\right)\right|$

  $\le E\left|X_n - X_{n-1}\right| + E\left|X_{n-1} - A_{n-1}\right| + E\left|A_{n-1} - A_n\right|$ by triangle inequality of the L1 norm

  $\le C_n + E\left|Y_{n-1}\right|$ where $E\left|X_n - X_{n-1}\right| + E\left|A_{n-1} - A_n\right| \le C_n \le \infty$ by integrability of these 2 terms

  $\le \sum_{k=1}^{n} C_k + E\left|Y_0\right| = \sum_{k=1}^{n} C_k + E\left|X_0 - A_0\right| = \sum_{k=1}^{n} C_k + E\left|X_0\right|$ since $A_0 = 0$ $\le \infty$

  since $\sum_{k=1}^{n} C_k \le \infty$ and X_0 integrable.

  So, $\{Y_n\}$ integrable.

- $\{Y_n\}$ adapted to $\{F_n\}$ since X_n adapted and A_n previsible

- $E\left[Y_n \mid F_{n-1}\right] - Y_{n-1} = E\left[X_n - A_n \mid F_{n-1}\right] - \left(X_{n-1} - A_{n-1}\right) = E\left[(X_n - X_{n-1}) \mid F_{n-1}\right] - \left(A_n - A_{n-1}\right)$

  by the fact that X_{n-1} is F_{n-1} measurable and A_n is previsible

  $= (A_n - A_{n-1}) - \left(A_n - A_{n-1}\right) = 0$ since by construction, $E\left[(X_n - X_{n-1}) \mid F_{n-1}\right] = A_n - A_{n-1}$

  So, $E\left[Y_n \mid F_{n-1}\right] = Y_{n-1}$.

- WTS uniqueness up to the value of Y_0:

  Suppose $\exists$ two such decompositions, $X_n = Y_n + A_n = \tilde{Y}_n + \tilde{A}_n$. Then, $Y_n - \tilde{Y}_n = \tilde{A}_n - A_n$.

  $A_n - \tilde{A}_n = E\left(A_n - \tilde{A}_n \mid F_{n-1}\right)$ by previsibility $= E\left(\tilde{Y}_n - Y_n \mid F_{n-1}\right) = \tilde{Y}_{n-1} - Y_{n-1}$ by M.G.

  $= \tilde{A}_{n-1} - A_{n-1} = E\left(A_{n-1} - \tilde{A}_{n-1} \mid F_{n-2}\right) = E\left(\tilde{Y}_{n-1} - Y_{n-1} \mid F_{n-2}\right) = \tilde{Y}_{n-2} - Y_{n-2}$

  $= ...$

  $= \tilde{Y}_0 - Y_0$

  Thus, if $\tilde{Y}_0 = Y_0$ then $\Rightarrow A_n = \tilde{A}_n \Rightarrow Y_n = \tilde{Y}_n \Rightarrow$ *SAME DECOMPOSITION*.

  That is, the decomposition is unique up to the value of Y_0.

l. (E.4.4.3) Doob's Decomposition for SubMG (A_n non-decreasing) and SupMG (A_n non-increasing)

If $\left(X_n, F_n\right)$ a subMG, then $A_n \le A_{n+1}\ \forall n$ in Doob's decomposition.

$\left(X_n, F_n\right)$ a supMG, then $A_n \ge A_{n+1}\ \forall n$ in Doob's decomposition.

Pf: $\left(X_n, F_n\right)$ a subMG $\Rightarrow E\left[X_{n+1} \mid F_n\right] \ge X_n \Rightarrow E\left[Y_{n+1} + A_{n+1} \mid F_n\right] \ge Y_n + A_n \Rightarrow Y_n + A_{n+1} \ge Y_n + A_n \Rightarrow A_{n+1} \ge A_n$

$\left(X_n, F_n\right)$ a supMG $\Rightarrow E\left[X_{n+1} \mid F_n\right] \le X_n \Rightarrow E\left[Y_{n+1} + A_{n+1} \mid F_n\right] \le Y_n + A_n \Rightarrow Y_n + A_{n+1} \le Y_n + A_n \Rightarrow A_{n+1} \le A_n$

Note: For this reason, Doob's decomposition is very attractive and useful for subMG and supMG.

Note2: Doob's is also particularly useful in connection with square-integrable martingales $\{X_n\}$, where one can relate the limit of X_n as $n \rightarrow \inf$ with that of the non-decreasing sequence A_n in the decomposition of $\{X_n^2\}$.

m. (T.4.4.7) Doob-Meyer Decomposition (continuous time analog of Doob's decomposition and fundamental to stochastic integration)

Suppose $\{F_t\}$ is right-continuous filtration and the MG (M_t, F_t) having continuous sample path is s.t. $E(M_t^2) < \infty$ $\forall t \geq 0$.

Then, $\exists$ a unique SP $\{A_t\}$ s.t.

(a) $A_0 = 0$

(b) $\{A_t\}$ has continuous sample path w.p. 1

(c) $\{A_t\}$ is adapted to $\{F_t\}$.

(d) $t \mapsto A_t$ is non-decreasing w.p.1

(e) $(M_t^2 - A_t, F_t)$ is a MG

Note: This is just the Doob decomposition of the sub-MG $X_t=M_t^2$, where (a) resolves the issue of the uniqueness of the R.V. A_0 measurable on F_0, (b) specifies the smoothness of the sample-path of the continuous-time S.P. $\{A_t\}$ and (d) is an analogy of the monoticity property we saw above.

Note on "Increasing Part": The S.P. $\{A_t\}$ in the Doob-Meyer decomposition of $\{M_t^2\}$ is called the increasing part of the increasing process associated with the MG $\{M_t\}$. Note that this is a decomposition of M_t^2 s.t. $M_t^2 = M_t^2 - A_t + A_t$ for some $\{A_t\}$ nondecreasing and $M_t^2 - A_t$ a MG.

Note on Quadratic Variation and Increasing Part: Later we will see that the increasing part gives us the quadratic variation (up to t) of M_t. We know for B-M that $W_t^2=(W_t^2-t)+t$ is the Doob Meyer decomposition, so that t is the increasing part. We also know that the quadratic variation for W_t is t!!!

n. (T.4.4.11) Doob's Inequality

(a) Suppose $\{X_n\}$ is a subMG. Then, $\forall x > 0$ and $N < \infty$, $P(\{\omega : \max_{0 \leq n \leq N} X_n(\omega) > x\}) \leq \frac{E|X_N|}{x}$

(b) Suppose $\{X_n\}$ is a subMG with a last element (X_∞, F_∞) with $F_t \subset F_\infty$ with X_∞ integrable and measurable wrt F_∞ and $\forall t>0$, $E(X_\infty | F_t) \geq X_t$.

Then, $\forall x > 0$, $P(\{\omega : \sup_{0 \leq n \leq \infty} X_n(\omega) > x\}) \leq \frac{E|X_\infty|}{x}$

(c) Suppose $\{X_t\}, t \in [0,T]$ is a continuous-parameter, right continuous subMG (i.e. each sample path $t \mapsto X_t(\omega)$ is right continuous).

Then, $\forall x > 0$, $P(\{\omega : \sup_{0 \leq t \leq T} X_t(\omega) > x\}) \leq \frac{E|X_T|}{x}$

Note: This theorem states that the maximum of a subMG does not grow too quickly.

Pf: Consider the stopping time $\tau_x(\omega) = \min\{n \geq 0 : X_n(\omega) > x\}$.

Observe that $A = \{\omega : \max_{0 \leq n \leq N} X_n(\omega) > x\} = \{\omega : \tau_x(\omega) \leq N\}$ =B

("$\subseteq$": Take $\omega \in A$. $\max_{0 \leq n \leq N} X_n(\omega) > x \Rightarrow$ at least one $n \in [0,N]$ is s.t. $X_n(\omega) > x \Rightarrow \tau_x(\omega) = \min\{n \geq 0 : X_n(\omega) > x\} \leq N$

"$\supseteq$": Take $\omega \in$ B. $\tau_x(\omega) = \min\{n \geq 0 : X_n(\omega) > x\} \leq N \Rightarrow$ at least one $n \in [0,N]$ is s.t. $X_n(\omega) > x$

$\Rightarrow$ the maximum X_n in $[0,N]$ is s.t. $X_n(\omega) > x \Rightarrow \max_{0 \leq n \leq N} X_n(\omega) > x$.)

Thus, ...

o. (T. 4.5.1) Doob's Martingale Convergence Theorem
(Idea: The fact that the maximum of a subMG does not grow too rapidly is closely related to convergence properties of subMG (also supMGs and MGs).

Suppose (X_t, F_t) is a right continuous subMG.

(a) If $\sup_{t \geq 0}\{E\ (X_t)_+\} < \infty$, then $X_\infty(\omega) = \lim_{t \to \infty} X_t(\omega)$ exists w.p.1. i.e. $\exists X_\infty(\omega)$ s.t. $X_t(\omega) \to_{a.s.} X(\omega)$

Furthermore, in this case $E|X_\infty| \leq \lim_{t \to \infty} E|X_t| < \infty$

(b) If $\{X_t\}$ is uniformly integrable, then $X_t \to X_\infty$ also in L^1.

Further, $X_t \to_{L1} X_\infty \Rightarrow (X_t, F_t)$ is a subMG with X_∞ its last element i.e. for any fixed $t \geq 0$, $X_t \leq E(X_\infty | F_t)$ also

Note on the Difference between (a) and (b):

To understand the difference between parts (a) and (b) of Doob's convergence, recall that if $\{X_t\}$ U.I. then E $(X_t)_+ \le C$ for some C<∞ and all t.

by Def 1.4.21 a necessary condition for U.I. is that $\sup_t E|X_t| = \sup_t \mathrm{E}(X_t)_+ + \mathrm{E}(X_t)_- < \infty \Rightarrow \mathrm{E}(X_t)_+ < \infty\ \forall t$

Further, by Th. 1.4.22, $\{X_t\}$ U.I. and $X_t \to_{a.s.} X$ (and thus in prob) $\Rightarrow X_t \to_{L1} X$ or $E(X_t) \to E(X)$.

Thus, part (b) 's assumption implies part (a) assumptions, so that we have a.s. convergence. PLUS U.I. we have L1 convergence.
This all comes for free using theorems we already have seen.

But the content of (b) i.e. requiring proof is that $X_t \to_{L1} X \Rightarrow X_t \le E(X_\infty | F_t)\ \forall t \ge 0$.

Note on Non-Converging MGs: **Many important martingales do not converge.**
(example: Brownian Motion! It can be shown that LimSup W(t) = inf and Lim Inf W(t) = -inf. Indeed, Doob's convergence assumptions do not hold because it can be shown that E[W(t)+] = sqrt(t/2pi) which is not bounded, so Doob's does not apply.)

p. (P.4.5.3) L2 and A.S. convergence of MGs (A Stronger Convergence Theorem than Doob's)

If the MG $\{X_n\}$ is s.t. $EX_n^2 \le C$ for some $C < \infty$ and all n, then ∃ a R.V. X_∞ s.t. $X_n \to_{a.s.} X_\infty$ AND $X_n \to_{L2} X_\infty$.

Moreover, $\mathrm{E}X_\infty^2 \le C < \infty$.

Note on No L1 Analog: There is no L1 analog to this proposition!
That is, there exists a MG $\{X_n\}$ s.t. $E[X_n]<C$ for some C finite, but the L1 and a.s. limits are different!
Example: There exists a non-negative MG $\{X_n\}$ s.t. $E(X_n)=1$ for all n and $X_n \to 0$ a.s. So, the a.s. and L1 limits are different. (See P.4.6.5)

q. (P.4.6.2) SP $X_n=m^{-n}Z_n$ is a MG for the filtration $\{F_n\}$, Zn a Branching Process

Pf:

$$E(Z_{n+1} | F_n) = E\left[\sum_{j=1}^{Z_n} N_j^{(n+1)} | F_n\right] = \sum_{j=1}^{Z_n} E\left(N_j^{(n+1)} | F_n\right)$$ by linearity of conditional expectation and since Z_n measurable w.r.t. F_n

$$= \sum_{j=1}^{Z_n} E\left(N_j^{(n+1)}\right)$$ since $N_j^{(n+1)}$ is independent of $\{F_n\}$ (the number of offsprings this gen for any ind ind. of the past)

$$= E(N) Z_n$$ since $N_j^{(n+1)}$ iid with finite mean m = E(N)<∞

$$= mZ_n$$

Thus, $E(X_{n+1} | F_n) = E\left[\frac{Z_{n+1}}{m^{n+1}} | F_n\right] = \frac{1}{m^{n+1}} E(Z_{n+1} | F_n) = \frac{1}{m^{n+1}} mZ_n = \frac{Z_n}{m^n} = X_n$

r. (P.4.6.3) Sub-Critical Process Dies Off / Extinction Probability is One when m < 1

If m<1 then $p_{ex} = 1$, i.e. with probability 1 the population eventually dies off.

Pf:

Recall that $X_n = \frac{Z_n}{m^n}$ is a non-negative MG. From Ex 4.5.6, we know from part (a) of Doob's convergence theorem that if $\{X_t\}$ is a nonnegative right-continuous MG, then $X_t \to_{as} X_\infty$ for some X_∞ where X_∞<∞ a.s..

$Z_n = m^n X_n$ and $m^n \to 0$ (since m<1) $\Rightarrow Z_n \to_{a.s.} 0$.

But Z_n integer valued, so $Z_n \to 0$ if $Z_n = 0$ for some n eventually.

s. (P.4.6.5) Critical Process Dies Off / Extinction Probability is One when m = 1

If m=1 then $p_{ex} = 1$, i.e. with probability 1 the population eventually dies off.

Chapter 5: Brownian Motion

BM is the most fundamental continuous time SP. It is both a martingale and a Gaussian S.P. It has continuous sample path (but differentiable nowhere), independent increments, and the strong Markov property.

Definitions

DEFINITION AND CONSTRUCTION

a. (D.5.1.1, T.5.2.1) Brownian Motion

(D.5.1.1) Definition

A S.P. $\{W_t, 0 \le t \le T\}$ is called a Brownian Motion (or a Wiener Process) if:

(a) W_t is a Gaussian Process

(b) $E(W_t) = 0$, $E(W_t W_s) = \min(t,s)$

(c) For almost every ω, the sample path $t \mapsto W_t(\omega)$ is continous on [0,T]

(T.5.2.1) Levy's Martingale Characterization of BM

Let $G_t = \sigma(X_s, 0 \le s \le t)$. If (X_t, G_t) is a MG w/ continuous sample path and $\underbrace{(X_t^2 - t, G_t) \text{ also a MG}}_{\substack{i.e. \text{ its Doob-Meyere decomp} \\ \text{is } X_t^2 - t}}$, then X_t a BM.

Note on FDD: (a) and (b) completely characterize the FDD of the Brownian Motion (because Gaussian processes are characterized by their mean and auto-covariance functions by C.3.2.15). Adding (c) allows us to characterize the sample path as well.

Note on Independent Increments of 0 Mean: BM has independent increments of 0 mean.

$E(W_{t+h} - W_h) = E(W_{t+h}) - E(W_h) = 0$

Now, $\{W_t\}$ Gaussian $\Rightarrow$ all lin comb are Gaussian $\Rightarrow (W_{t+h} - W_t,\, W_s)$ Gaussian (s<t)

So, they're independent iff uncorrelated.

$Cov(W_{t+h} - W_t,\, W_s) = Cov(W_{t+h},\, W_s) - Cov(W_t,\, W_s) = \min(t+h, s) - \min(t,s) = s - s = 0$

Note on Constructing BM: Read pg 102 – 104. The idea is that we can construct Gaussian process with the distributional properties (a) and (b) of a BM. Then, by Kolmogorov's continuity theorem, we say there exists a modification of this Gaussian process that has continuous sample path – this modification is the BM.

SMOOTHNESS AND VARIATION OF THE BROWNIAN SAMPLE PATH

b. Comparison between Variation of Nice Calculus and Stochastic Calculus:

	"Nice" Calculus	"Stochastic" Calculus
Total Variation	<Inf	+Inf
Quadratic Variation	0	<Inf
P≥3 Variation	0	0

c. (D.5.3.1) || π || (Length of Longest Interval) and Q-th Variation of f(.) on π (CALCULUS DEFINITION)

For any finite partition π of [a,b], that is, $\pi = \left\{a = t_0^{(\pi)} < t_1^{(\pi)} < \ldots < t_k^{(\pi)} = b\right\}$,

let $\boxed{\|\pi\| = \max_i \left\{t_{i+1}^{(\pi)} - t_i^{(\pi)}\right\}}$ denote the length of the longest interval in π

and let $\boxed{V_{(\pi)}^{(q)}(f) = \sum_i \left| f\left(t_{i+1}^{(\pi)}\right) - f\left(t_i^{(\pi)}\right)\right|^q}$ denote the q-th variation of f(.) on π.

The q-th variation of f(.) on [a,b] is $\boxed{\lim_{\|\pi\| \to 0} V_{(\pi)}^{(q)}(f)}$

d. (D.5.3.2) Q-th variation of a S.P. {X(t)} on the interval [a,b]

The q-th variation of a S.P. on the interval [a,b] is the random variable $V^{(q)}(X)$ obtained when replacing f(t) by $X_t(\omega)$ in the above definition, provided the limit exists.

e. Lipschitz Sample Path (w.p.1)

We say that a S.P. X(t) has Lipschitz sample path with probability 1 if ∃ a RV L(ω) which is finite a.s. s.t. $|X(t) - X(s)| \le L|t-s|$ ∀t,s∈ a,b

Note on Quadratic Variation and Smoothness of the Sample Path (A Lipschitz Continuous RV has 0-Quadratic Variation):

The quadratic variation is affected by the smoothness of the sample path.

Suppose a SP X(t) has Lipschitz sample path w.p. 1. (∃ RV L(ω) s.t. |X(t)-X(s)| ≤ L|t-s|.)

Then, $V_{(\pi)}^{(2)}(X)=\sum_i \left|X\left(t_{i+1}^{(\pi)}\right)-X\left(t_i^{(\pi)}\right)\right|^2 \le L^2\sum_i\left(t_{i+1}^{(\pi)}-t_i^{(\pi)}\right)^2 \le L^2\|\pi\|\sum_i\left(t_{i+1}^{(\pi)}-t_i^{(\pi)}\right)$ bc $\|\pi\| \le t_{i+1}^{(\pi)}-t_i^{(\pi)}$ ∀i

$\le L^2\|\pi\|(b-a)\rightarrow_{a.s.} 0$ as $\|\pi\|\rightarrow 0$.

So, X(t) has 0 quadratic variation.

f. (D.5.3.3) Quadratic Variation

The quadratic variation of a S.P. X, denoted $V_t^{(2)}(X)$, is the non-decreasing, non-negative SP corresponding to the quadratic variation of X on the intervals [0,t].

Theorems

DEFINITION AND CONSTRUCTION

a. (P.5.1.2) Brownian Motion has Independent Increments of Zero Mean

BM has independent increments with mean 0

Pf: From part (b) of the definition, we get that for t s, h>0,

( ) ( ) ( ) ( ) ( ) ( )

1. (Brownian Bridge): $B_t = W_t - \min(1,t,)W_t$

(a) $\{B_t\}$ a Gaussian S.P.

(b) $E(B_t) = 0$ and ρ(t,s)=min(t,s)-min(s,1)min(t,1)

(c) Continuous sample paths a.s.

(d) Not adapted to the canonical filtration of W_t and not stationary

(e) B_t has same distribution as $W_t \mid W_1 = 0$

2. (Geometric BM): $Y_t = e^{W_t}$

(a) $\{Y_t\}$ a Gaussian S.P.

(b) $E(Y_t) = \exp(-1/2)$ and ρ(t,s)=exp(min(t,s))[exp((1/2)(t+s)-1]

(c) Continuous sample paths a.s.

(d) Not adapted to the canonical filtration of W_t and not stationary

3. (Ornstein-Uhlenbeck): $U_t = e^{-t/2}W_{e^t}$

(a) $\{U_t\}$ a Gaussian S.P.

(b) $E(U_t) = 0$ and $\rho(t,s) = \exp\ -\frac{1}{2}|t-s|$

(c) Continuous sample paths a.s.

(d) Not adapted to the canonical filtration of W_t

(e) Stationary process with stationary increments

4. (BM with Drift): $X_t = x + \mu t + \sigma W_t$

(a) $\{X_t\}$ a Gaussian S.P.

(b) $E(X_t) = x + \mu t$ and $\rho(t,s) = \sigma^2 \min(t,s)$

(c) Continuous sample paths a.s.

(d) Adapted to the canonical filtration of W_t and not stationary

(SEE PS5 for Work)

REFLECTION PRINCIPLE AND BROWNIAN MOTION HITTING TIMES

d. (P.5.2.2) $X_t = W_{t+\tau} - W_\tau$ a BM **(USEFUL FOR REFLECTION PRINCIPLE)**

$X_t = W_{t+\tau} - W_\tau$ a BM, where τ is a G_t-stopping time, G_t the cononical filtration for a brownian. And X_t independent of stopped σ-field G_τ

e. $\{\omega : W_T(\omega) \ge \alpha\} \subseteq \{\omega : \max_{0\le s\le T} W_s(\omega) \ge \alpha\} = \{\omega : \tau_\alpha(\omega) \le T\}$ **where** $\tau_\alpha = \inf(t > 0 : W_t = \alpha)$ **is the First Hitting Time of BM**

Pf: The first inclusion should be obvious. The second equality is intuitively clear, because the LHS is the set of outcomes in which the BM has hit a by time T, which is precisely what the RHS is.

f. Reflection Principle (pp 106-107)

A

$$\boxed{P\left(\max_{0\le s\le T} W_s \ge \alpha, W_T \ge \alpha\right)} = \underbrace{P\left(\tau_\alpha \le T, W_T \ge \alpha\right)}_{\text{true by above}} = \underbrace{P\left(\tau_\alpha \le T, X_{T-\tau_\alpha} \le 0\right)}_{\text{bc if first hitting time is } \le T \text{ and by T W(T) still } \ge\alpha \text{, so the difference bewteen the first hitting time and T must be positive}} = \underbrace{P\left(\tau_\alpha \le T, X_{T-\tau_\alpha} \ge 0\right)}_{\text{Using the fact that } X_{T-\tau_\alpha} \text{ is a BM and thus for each t, distribution is symmetric around 0}} = P\left(\tau_\alpha \le T, W_T \le \alpha\right) = \boxed{P\left(\max_{0\le s\le T} W_s \ge \alpha, W_T \le \alpha\right)}$$

B

$$\boxed{P\left(\max_{0\le s\le T} W_s \ge \alpha\right)} = P\left(\max_{0\le s\le T} W_s \ge \alpha, W_T \ge \alpha\right) + P\left(\max_{0\le s\le T} W_s \ge \alpha, W_T \le \alpha\right) = 2\underbrace{P\left(\max_{0\le s\le T} W_s \ge \alpha, W_T \ge \alpha\right)}_{\text{by reflection principal A}} = \boxed{2P\left(W_T \ge \alpha\right)}$$

SMOOTHNESS AND VARIATIONS OF THE BROWNIAN MOTION PATH

g. (P.5.3.9) Total Variation of BM W(t) is infinite w.p.1

Pf:

Let $\alpha(h)=\sup_{a\le t\le b-h}|W(t+h)-W(t)|$. (the "largest" difference in BM values in a window of size h, for $t\in[a,b-h]$. Or, this is the difference in BM values when we look over the set of h-width windows where the left side of the window does not go below a and right side of the window does not go above b. We take the sup because there are uncountably many of these windows that we need to look at).

W.P. 1, the sample path W(t) is continuous and hence uniformly continuous on the closed and bounded interval [a,b]. Therefore, $\alpha(h)\to_{a.s.} 0$ as $h\to 0$.

Let π_n divide [a,b] to 2^n equal parts, so $\|\pi_n\|=\frac{(b-a)}{2^n}$.

Then,

$$(*)\quad V_{(\pi_n)}^{(2)}(W)=\sum_{i=0}^{2^n-1}\left|W\left(a+(i+1)\|\pi_n\|\right)-W\left(a+i\|\pi_n\|\right)\right|^2\le\alpha(\|\pi_n\|)\sum_{i=0}^{2^n-1}\left|W\left(a+(i+1)\|\pi_n\|\right)-W\left(a+i\|\pi_n\|\right)\right|$$

From Ex.5.3.6, we know that $V_{(\pi_n)}^{(2)}(W)\to(b-a)<\infty$.

Furthermore, we have $\alpha\left(\|\pi_n\|\right)\to_{a.s.} 0$ from above,

So, $V_{(\pi_n)}^{(2)}(W)\le\alpha(\|\pi_n\|)\sum_{i=0}^{2^n-1}\left|W\left(a+(i+1)\|\pi_n\|\right)-W\left(a+i\|\pi_n\|\right)\right|<\infty$, and $\alpha(\|\pi_n\|)\to 0$

$\Rightarrow V_{(\pi_n)}^{(1)}(W)=\sum_{i=0}^{2^n-1}\left|W\left(a+(i+1)\|\pi_n\|\right)-W\left(a+i\|\pi_n\|\right)\right|\to_{a.s.}\infty$

$\Rightarrow V^{(1)}(W)=\infty$ wp1

h. **(D.5.3.4)** For BM W(t), as $\|\pi\|\to 0$ we have that $V_{(\pi)}^{(2)}(W)\to(b-a)$ in 2-mean.

i. **Quadratic Variation of a BM W(t) is t, since its Doob-Meyer Decomposition is $W(t)^2=[W(t)^2-t]+t$**

Chapter 6: Markov, Poisson, and Jump Processes

These are 3 important families of stochastic processes.

Definitions

	Discrete Time	Continuous Time
Discrete Values		
Continuous Values		

MARKOV PROCESSES

a. (D.6.1.1) Markov Chain
Discrete MC (We mean discrete time, but can be continuous or discrete values)

A discrete time stochastic process $\left(X_n, n=0,1,...\right)$ with each RV X_n taking values in measurable space $\left(S,B\right)$ is called a Markov Chain if

$$\boxed{\forall n\in N_+,\ \forall A\in B,\ P\left(X_{n+1}\in A\mid X_1,...,X_n\right)=P\left(X_{n+1}\in A\mid X_n\right)\ \text{a.s.}}$$

or $$\boxed{\forall n\in N_+,\ \forall A\in B,\ E\left[I\left(X_{n+1}\in A\right)\mid X_1,...,X_n\right]=E\left[I\left(X_{n+1}\in A\right)\mid X_n\right]\ \text{a.s.}}$$

(or $P\left(X_{n+1}\le x_{n+1}\mid X_1=x_1,...,X_n=x_n\right)=P\left(X_{n+1}\le x_{n+1}\mid X_n=x_n\right)$ if B is the Borel sigma field, bc B = $\sigma\left(\left\{X_{n+1}\le\alpha\right\}\right)\ \forall\alpha$)

$$\Leftrightarrow \boxed{\forall n\in N_+, \forall m\in N_+,\ \forall A\in B,\ P\left(X_{n+m}\in A\mid X_1,...,X_n\right)=P\left(X_{n+m}\in A\mid X_n\right)\ \text{a.s.}}$$

$$\Leftrightarrow \boxed{\forall n\in N_+,\text{for any bounded measurable function } f:S\mapsto R,\ E\left(f\left(X_{n+1}\right)\mid X_1,...,X_n\right)=E\left(f\left(X_{n+1}\right)\mid X_n\right)\ \text{a.s.}}$$

The set S is called the state space of the Markov chain.

Continuous Time MC

A S.P. X(t) indexted by $t\in[0,\infty)$ and taking values in a measurable space (S,B) is called a Markov Process if for any $t,u\ge 0$ and $A\in B$ we have that a.s.

$$\boxed{P\left(X_{t+u}\in A\mid\sigma(X_s,s\le t)\right)=P\left(X_{t+u}\in A\mid X_t\right)}$$

or $$\boxed{E\left(I\left(X_{t+u}\in A\right)\mid\sigma(X_s,s\le t)\right)=E\left(I\left(X_{t+u}\in A\right)\mid X_t\right)}$$

$$\Leftrightarrow \boxed{\begin{array}{c}\text{for any } t,u\ge 0 \text{ and any bounded measurable function f(.) on (S,B),}\\ E\left(f(X_{t+u})\mid\sigma(X_s,s\le t)\right)=E\left(f(X_{t+u})\mid X_t\right)\ \text{a.s.}\end{array}}$$

The set Sis called the state apace of the Markov Process.

Note on Markov Process and Joint Law: MARKOV PRPERTY IS A PROPERTY OF THE JOINT LAW. SO IF YOU KNOW ANOTHER PROCESS WITH SAME JOINT LAW THAT YOU CAN SHOW TO BE MARKOV, THEN WE CAN USE THAT THE SHOW THAT THE ORIGINAL PROCESS IS MORKOV.

b. (D.6.1.2) (Time) Homogeneous Markov Chain

A homogeneous Markov chain is a Markov chain that has a modification for which

$$\boxed{\forall A\in B, P\left(X_{n+1}\in A\mid X_n\right) \text{ does not depend on n (except via the value of } X_n)}.$$

c. (D.6.1.3, D.6.1.8) Stationary Transition Probabilities (Determines distribution of Homogeneous MCs)
(D.6.1.3) Discrete Time

To each homogenoues Markov chain $\left\{X_n\right\}$ with values in a closed subset S of R correspond its stationary transition probabilities p(A|x) s.t. p(.|x) is a probability measure on (S,B) for any $x\in S$.

$$\boxed{p(A|x) \text{ is measurable on B for any } A\in B, \text{ and } p(A|X_n)=P\left(X_{n+1}\in A\mid X_n\right)\ \ \forall n\ge 0}$$

(D.6.1.8) Continuous Time

For each t>s and fixed $x\in S$, $\exists$ a probability measure $p_{t,s}\left(.\mid x\right)$ on $\left(S,B\right)$ s.t. for each fixed $A\in B$,

$$\boxed{\text{the function } p_{t,s}(A\mid .) \text{ is measurable and } p_{t,s}(A\mid X_s)=E\left(I\left\{X_t\in A\right\}\mid X_s\right)=P\left(X_t\in A\mid X_s\right)}$$

Such a collection $p_{t,s}(A\mid x)$ is called the transition probabilities for the Markov process $\left\{X_t\right\}$.

Example of stationary homogeneous MCs: Random Walk

d. (D.6.1.4) Initial Distribution of a Markov Chain

The initial distribution of a M.C. is the distribution π of X_0.

That is, it's the probability measure $\pi(A) = P(X_0 \in A)$ on (S, B)

Note on Using Initial Distribution to Determine the Distribution of X_n:

The distribution of X_n for any n is determined by the initial distribution π and the transition probability function $\{p(A \mid x), A \in B, x \in S\}$.

If discrete state space and time homogenous, we can find the distribution of X_n given π is: $\pi' P^n$

Note on Initial Distribution + Probability Function Determine FDD of the M.C.:

To get FDD, we just need to give initial distribution and the transition probability function (or the transition matrix P if discrete state space).

In particular, for each nonnegative integer k, every $0=t_0<t_1<...<t_k$ and $A_0,...,A_k \in B$ we have that

$$P(X(t_k) \in A_k, ..., X(t_0) \in A_0) = \int_{\omega \in A_k} ... \int_{\omega \in A_0} p_{t_k, t_{k-1}}(dx_k \mid x_{k-1}) ... p_{t_1, t_0}(dx_1 \mid x_0) \pi(dx_0)$$ (using Lebesgue Integral)

WHAT DOES THIS MEAN?!?!?

Note on Homogenoues Markov Chains:

Homogenous MC's are fully characterized by the initial distributions and the (one-step) transition probabilities $p_t(.|.)$ for all $t > 0$ in order to determine all distributional properties of the associated homogeneous Markov process.

In view of the Chapman-Kolmogorov relationship, using functional analysis one wmay often express $p_t(.|.)$ in terms of a single operator, called the "generator" of the Markov process. For example, the generator of BM is closely related to the heat equation, hence the reasont hat many computations can be simplified via the theory of PDE.

POISSON PROCESS, EXPONENTIAL INTER-ARRIVALS, AND ORDER STATISTICS

e. 4 Useful Conditions for Defining Poisson Processes:

C0 (Counting Process):

a) Each sample path $N_t(\omega)$ is piecewise constant, nondecreasing, and right-continuous

b) $N_0(\omega) = 0$

c) All jumnp discontinuities are of size 1

C1 (Independent and Stationary Poisson Increments):

For any k and any $0 < t_1 < ... < t_k$,

a) the increments, $N_{t_1}, N_{t_2} - N_{t_1}, ..., N_{t_k} - N_{t_{k-1}}$ are independent R.V.'s

b) for some $\lambda > 0$ and all $t > s \geq 0$, the increment $N_t - N_s \sim Poisson(\lambda(t-s))$

(so stationary and independent poisson distributed increments)

C2 (Independent and Stationary Increments):

For any k and any $0 < t_1 < ... < t_k$,

a) the increments $N_{t_1}, N_{t_2} - N_{t_1}, ..., N_{t_k} - N_{t_{k-1}}$ are independent R.V.'s

b) for all $t > s \geq 0$, the increment $N_t - N_s$ have law that depends only on (t-s)

(i.e. stationary increments)

C3 (Jump Time Increments ~ Exponential(λ)):

The gaps / increments between jump times $T_k - T_{k-1}$ for k=1,2,... are iid RV, each distributed Exp(λ)

C4

a) The S.P. N_t has no fixed discontinuities. i.e. $P(T_k = t) = 0$ $\forall$k and $t \geq 0$.

b) For any fixed k, $0<t_1<t_2<...<t_k$ and nonnegative integers $n_1, n_2, ..., n_k$,

$P(N_{t_k+h} - N_{t_k} = 1 \mid N_{t_j} = n_j, j \leq k) = \lambda h + o(h)$.

$P(N_{t_k+h} - N_{t_k} \geq 2 \mid N_{t_j} = n_j, j \leq k) = o(h)$.

where o(h) denotes a function f(h) s.t. $\frac{f(h)}{h} \to 0$ as $h \downarrow 0$.

f. (D.6.2.1) Counting Process (C0)

N_t is a counting process if...

a) Each sample path $N_t(\omega)$ is piecewise constant, nondecreasing, and right-continuous

b) $N_0(\omega) = 0$

c) All jumnp discontinuities are of size 1

or equivalently, $N_t = \sup\{k \geq 0 : T_k \leq t\}$ where T_k are the jump-times associated with N_t.

Note on Application: We use N(t) as counting the number of discrete events / occurrences in the interval [0,t] for each t>0, with T(k) denoting the arrival or occurrence time of the k-th such event. That's why we call N(t) a counting process.

g. (D.6.2.1) Jump Times

Associated with each sample path of a counting process, $N_t(\omega)$, are jump times $0 = T_0 < T_1 < \ldots$ such that

$T_k = \inf\{t \geq 0 : N_t \geq k\}$ for each k. (T_k is the time when the k-th occurrence arrives/occurs, or, the first time when the counting process reaches k)

Thus, as stated above, counting time is equivalently stated as: $N_t = \sup\{k \geq 0 : T_k \leq t\}$ (bc N_t is the number of occurrences at time t)

h. Poisson R.V.

A RV N has the Poisson(λ) law if $\boxed{P(N = k) = \frac{\lambda^k e^{-\lambda}}{k!},\ k \geq 0 \text{ integer}}$

Has E(N) = λ and Var(N) = λ

i. S.P. with Independent Increments

We say that a S.P. N_t has independent increments if $\boxed{N_{t+h} - N_t \perp \sigma(N_d, 0 \leq s \leq t)\ \forall h\ \forall t}$

j. (D. 6.2.3) S.P. with Stationary Increments

We say that the S.P. N_t $t \geq 0$ has stationary increments if the law of N_{t+h}-N_t is independent (i.e. not a function) of t, but a function of h.

k. (P.6.2.5) Memoryless Property of Exponential Law

We say that a R.V. T has Exp(λ) law if $P(T > t) = e^{-\lambda t}$ for all $t \geq 0$ and some λ>0.

Except for the trivial case of T = 0 w.t. 1, these are the ONLY laws for which

$$\boxed{P(T > x + y \mid T > y) = P(T > x)\ \ \forall x,y \geq 0}$$

l. (D.6.2.1, P.6.2.4) Poisson Process

Definition 1(D.6.2.1): C0 + C1

A poisson process is a counting process (C0) that also satisfies C1 (independent and stationary Poisson increments):

i.e. for any k and any $0 < t_1 < \ldots < t_k$,

a) the increments, $N_{t_1}, N_{t_2} - N_{t_1}, \ldots, N_{t_k} - N_{t_{k-1}}$ are independent R.V.'s

b) for some $\lambda > 0$ (rate of intensity) and all $t > s \geq 0$, the increment $N_t - N_s \sim Poisson(\lambda(t-s))$

(so stationary and independent poisson distributed increments)

In other words, among the processes satisfying C0, the Poisson Process is the unique S.P. having also the property C1.

Note on Terminology: The Poisson process has independent increments, each having a Poisson law, where the parameter of the count N(t)-N(s) is proportional to the length of the corresponding interval [s,t], proportional by exactly a constant λ, called the rate of **intensity of the Poisson process**.

Note on Comparison with Brownian: Just as Brownian S.P., for which B(t)~N(0,t) and B(t)-B(s)~N(0,t-s), then Poisson S.P. is such that N(t)~Poi(λt), and N(t) – N(s) ~ Poi(λ(t-s)).

Definition 2 (P.6.2.4): C0 + C2

A S.P. is a Poisson process iff it is a counting process (C0) with stationary independent increments (C2).

Note on equivalence: **By (P.6.2.4), the poisson process is the only S.P. that satisfies C0 & C2.**

Definition 3 (P.6.2.6): C0 + C3

A S.P. is a Poisson process with rate λ iff it is a counting process (C0) with Jump Time Increments~Exp(λ) (C3)

Note on equivalence: **By (P.6.2.6).**

Definition 4 (P.6.2.6): C0 + C4

A S.P. is a Poisson process of rate λ iff it is a counting process (C0) that satisfies C4.

Theorems

a. Showing that a SP is a Markov SP

1. Compute $P(X_{t+h} \in A | F_t)$ *or* $P(X_{n+1} \in A | F_n)$ if discrete-time only depends on t, and not on $\{X_u, u < t\}$
2. Show that the SP $\{X_t\}$ has independent increments (using following lemma) (See "c" below)
3. Show that it's a function X(t) = f_t[Y(g(t))] of a Markov Process Y(t), for f invertible and g strictly increasing. (See "e" below)
4. Since Markov property is a property of the joint distribution, show that the SP has the same FDD/ Joint Distribution as some other stochastic process which is Markov.

b. Showing that a Markov SP is Time-Homogeneous

1. Use P.6.1.13: Every continuous time stochastic process of stationary and independent increments is a homogenous Markov Process.
2. If X(t) = f[Y(g(t))] for f invertible and time independent, g(t) strictly increasing, and Y(t) is homogeneous, then X(t) also. (See "e" below)
3. If {X(t)} a Markov Process and Stationary Process, then it's homogeneous. (See "d")

c. (Lemma 1) If {X(t)} is a S.P. with independent increments, then {X(t)} is a Markov SP.

PF: Let $\{F_t, t \geq 0\}$ be the canonical filtration of $\{X_t, t \geq 0\}$.

Independent increments means that for any t,h ≥ 0, the random variable X_{t+h} - X_h $\perp F_t$.

To show that $\{X_t\}$ is a Markov process, from Denition 6.1.7 it suffices to show that

$$E[f(X_{t+h}) | F_t] = E[f(X_{t+h}) | X_t].$$

for any bounded measurable function f on (S,B), where S is state space and B Borel.

Let f be an arbitrary bounded measurable function.

Then,

$$E[f(X_{t+h}) | F_t] = E[f(X_{t+h} - X_t + X_t) | F_t] = E[f(X_{t+h} - X_t + X_t) | X_t]$$

since $X_{t+h} - X_t$ is ind of F_t and X_t is X_t measurable.

Note on Converse: Converse is NOT necessarily true. Consider process in Ex 6.1.14.
Note on Intuition : Markov Process is one which the conditional distribution on the past only depends on where you were last. But that's precisely what it means to have independent increments. Imagine a simple random walk. Given that you're at some point, because we have independent increments, it doesn't matter where you were before, just where you are now.

d. (Lemma 2) If {X(t)} is a stationary process and a Markov Process, then it is Homogeneous.

Pf: Here proof only for discrete time, countable state process $\{X_n\}$.

$\{X_n\}$ stationary process $\Rightarrow X_n =_D X_0$ and $(X_{n+1}, X_n) =_D (X_1, X_0)$.

Thus,

$$P(X_{n+1} = y | X_n = x) = \frac{P(X_{n+1} = y, X_n = x)}{P(X_n = x)} = \frac{P(X_1 = y, X_0 = x)}{P(X_0 = x)} = \frac{P(X_1 = y, X_0 = x)}{P(X_0 = x)} = P(X_1 = y | X_0 = x)$$

So the transition probabilities do not depend on n, i.e. chain is time-homogeneous.

Note on Homogeneous Markov Process Has Nothing to Do with Stationary Increments (Except Inc & Stat Inc → Hom Mark Process):
It can be shown that a Markov Process with stationary increments is not necessarily time-homogeneous, and also that a Homogeneous Markov Process may not necessarily have stationary increments.

e. (Extra) If X(t) a Markov Process, then Y(t) = $f_t(X_{g(t)})$ is also a Markov Process for f invertible and g strictly increasing.
Theorem: If X(t) a Markov Process, then Y(t) = $f_t(X_{g(t)})$ also a Markov Process for f invertible and g strictly increasing.
If further that X(t) is homogenous and f not time dependent, then $f(X_{g(t)})$ also homogeneous.

Pf:

Take $A \in F_t \equiv \sigma(Y_s, s \le t)$

Note that since f invertible, $\sigma(Y_s, s \le t) = \sigma(X_{g(s)}, s \le t) = G_{g(t)}$

$P(Y_{t+u} \in A \mid F_t) = P(f_{t+u}(X_{g(t+u)}) \in A \mid F_t)$

$= P(X_{g(t+u)} \in f_{t+u}^{-1}(A) \mid F_t)$ since f invertible (and this is measurable w.r.t. F_t since the canonical filtrations are the same)

$= P(X_{g(t+u)} \in f_{t+u}^{-1}(A) \mid G_{g(t)})$

$= P(X_{g(t+u)} \in f_{t+u}^{-1}(A) \mid \sigma(X_{g(t)}))$ by X_t Markovian

$= P(f_{t+u}(X_{g(t+u)}) \in A \mid \sigma(f(X_{g(t)})))$ again b.c. f invertible, $\sigma(X_{g(t)}) = \sigma(f(X_{g(t)}))$

$= P(Y_{t+u} \in A \mid \sigma(Y_t))$

So, Y_t is Markovian.

Furthermore, if X_t time-homogenous, and f is not time dependent, then f(X_t) also time-homogenoues.

f. **(P.6.1.13) Every Continuous Time SP {X(t)} with Stationary and Independent Increments is a Homogenous Markov Process.**

g. **(P.6.1.5) Strong Markov Property**

h. **(P.6.2.4) Poisson Process is the only S.P. with stationary independent increments that satisfies condition C0 (counting process).**

i. **(P.6.2.5) Memoryless Property of Exponential Law**

We say that a R.V. T has Exp(λ) law if $P(T > t) = e^{-\lambda t}$ for all $t \ge 0$ and some $\lambda > 0$.

Except for the trivial case of T = 0 w.t. 1, these are the ONLY laws for which

$$\mathrm{P}(T > x + y \mid T > y) = P(T > x) \quad \forall x, y \ge 0$$

j. **(P.6.2.6) A S.P. N(t) that satisfies C0 is a Poisson process of rate λ iff it satisfies C3.**

k. **(P.6.2.8) Relationship between Poisson Process and Uniform Measure**

Fixing any integer n and $0 \le t_1 \le t_2 \le \ldots \le t_n \le t$, we have that

$$P(T_k \le t_k, k = 1, \ldots, n \mid N_t = n) = P(T_1 \le t_1, T_2 \le t_2, \ldots, T_n \le t_n) = \frac{n!}{t^n} \int_0^{t_1} \int_{x_1}^{t_2} \cdots \int_{x_{n-1}}^{t_n} dx_1 dx_2 \ldots dx_n$$

That is, conditional on N_t=n, the first n arrival times $\{T_k : k = 1, \ldots, n\}$ have the distribution of the order statistic of a sample of n iid Unif[0,t] R.V.s

WHY???

Note on Application:

For example, $E\left[\sum_{i=1}^{N_t} T_i \mid N_t = n\right] = E\left[\sum_{i=1}^{n} U_{(i)}\right]$ by theorem $= E\left[\sum_{i=1}^{n} U_i\right]$ since it doesn't matter how we sum

$= \sum_{i=1}^{n} E(U_i) = n\frac{t}{2}$ since each $U_i \sim_{iid} Unif[0,t]$

l. **(T.6.2.10) Poisson Approximation**

Suppose that for each n, the random variables $Z_t^{(n)}$ are independent, nonnegative integers where

$P(Z_l^{(n)} = 1) = p_l^{(n)}$ and $P(Z_l^{(n)} \ge 2) = \varepsilon_l^{(n)}$ are such that as $n \to \infty$,

$$\text{(i) } \sum_{l=1}^{n} p_l^{(n)} \to \lambda \in (0, \infty) \quad \text{(ii) } \sum_{l=1}^{n} \varepsilon_l^{(n)} \to 0 \quad \text{and} \quad \text{(iii) } \max_{l=1,\ldots,n}\{p_l^{(n)}\} \to 0$$

Then, $S_n = \sum_{l=1}^{n} Z_l^{(n)} \to_D Poisson(\lambda)$ when $n \to \infty$

Note on Approximating Poisson by Binomial:

Take $Z_l^{(n)} = \begin{cases} 1 \text{ w.p. } \frac{\lambda}{n} \\ 0 \text{ w.p. } 1-\frac{\lambda}{n} \end{cases}$

So, $Z_l^{(n)} \sim Bernoulli\left(\frac{\lambda}{n}\right)$, $\Rightarrow S_n = \sum_{l=1}^{n} Z_l^{(n)} \sim Binomial\left(n, \frac{\lambda}{n}\right)$ (since sum of iid bernoulli)

Here, $\sum_{l=1}^{n} p_l^{(n)} = \sum_{l=1}^{n} \frac{\lambda}{n} = \lambda$ and $P\left(Z_l^{(n)} \geq 2\right) = \varepsilon_l^{(n)} = 0$ so $\sum_{l=1}^{n} \varepsilon_l^{(n)} = 0$ and $\max_{l=1,\ldots,n}\left\{p_l^{(n)}\right\} = \frac{\lambda}{n} \to_{n\to\infty} 0$

Thus, by theorem above, we have $S_n = \sum_{l=1}^{n} Z_l^{(n)} \to_D Poisson(\lambda)$ when $n \to \infty$

Note on Brownian Motion and Functional CLT: T.6.2.10 plays for the Poisson process the same role that CLT plays for the BM. i.e. It provides a characterization of the Poisson process that is very attractive for the purpose of modeling real world phenomena.

m. (P.6.2.12) N1(t) + N2(t) Poisson if N1(t), N2(t) independent Poisson

If $N_t^{(1)}$ and $N_t^{(2)}$ are two independent Poisson processes of rates λ_1 and λ_2 respectively, then $N_t^{(1)}+N_t^{(2)}$ is a Poisson process of rate $\lambda_1 + \lambda_2$.

Conversely, the sub-sequence of jump times obtained by independently keeping with probability p of each of the jump times of a Poisson process of rate λ corresponds to a Poisson process of rate λp.