Table of Contents

Preliminary Definitions: .............................................................................................................. 7

a. Sup and Inf .............................................................................................................................. 7

b. Lim Inf and Lim Sup of a Sequence of reals ........................................................................... 7

c. Lim Inf and Lim Sup of a Set .................................................................................................. 7

d. Metric Spaces ...................................................................................................................... 7

e. Normed Spaces ................................................................................................................... 7

f. Inner Product Spaces ........................................................................................................... 7

g. Stochastic Order ................................................................................................................ 7

Chapter 1: Probability Measure and Integration ........................................................................ 8

I. Definitions: ............................................................................................................................ 8

a. Probability Space .................................................................................................................. 8

b. (D.1.1.1) Sigma-Algebra: (We want to impose some structure on F) ....................................... 8

c. (D.1.1.2) Measurable Space: A pair \( \Omega, F \) with F a sigma-field is called a measurable space. 8

d. (D.1.1.2) Probability Measure: .......................................................................................... 8

e. (D.1.1.5) Generated Sigma Fields: ..................................................................................... 8

f. (D.1.1.7/E.1.1.8) Borel Sigma-Field on \( \mathbb{R}\): ................................................................. 8

g. (D.1.2.8) Sigma Field Generated by a RV: .......................................................................... 9

h. (D.1.2.8) Sigma Field Generated by a Sequence of R.V.s: .................................................... 9

i. (D.1.2.1) Random Variable: ................................................................................................ 9

j. (E.1.2.2) Indicators: ............................................................................................................. 9

k. (E.1.2.3) Simple Functions: ................................................................................................ 9

l. (D.1.2.7) X,Y Almost Surely the same: We say X,Y def on same \( F,P \) are almost surely the same if \( P : X( ) \neq Y( ) \implies 0 \) .... 9

m. (D.1.2.11) Borel Measurable Function: .............................................................................. 9

n. (D.1.2.18) Mathematical Expectation of a R.V.: ................................................................ 9

o. (D.1.2.24) Positive (X+) and Negative (X-) Parts of a RV: ................................................. 9

p. (D.1.2.25) Integrable R.V.: .................................................................................................. 10

q. (D.1.4.21,L.1.4.25) Uniform Integrability of a Collection of R.V.s ............................................ 10

r. Sufficient Conditions for Uniform Integrability of a Collection of R.V.'s: ......................... 10

s. (D.1.3.1,D.1.3.2,D.1.4.9,P.1.4.10) Convergences of R.V.s ..................................................... 10

t. (D.1.3.4) Complete Probability Space .................................................................................. 11

u. (D.1.3.2) \( L^1 \) space ............................................................................................................. 11

v. (D.1.4.1) Law of a R.V. ........................................................................................................ 11
II. Theorems/Propositions:

a. (E.1.1.3) Properties of Probability Measures

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

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

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

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

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:

h. (P.1.2.31) Properties of Expectation

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

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

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

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

m. (P.1.3.22) Typical Application of BCI: E(X^2_n) = 1 \sum_{n} \frac{1}{n} X_n^2 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 \lim X_n \lim E X_n

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

p. (P.1.4.3) Change of Variables (We can measure a RV either in the input space of in the output space!)

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 \rightarrow E[h(X)] = E[h(Y)] for all h bounded and Borel measurable.

Chapter 2: Conditional Expectation and Hilbert Spaces

I. Definitions:

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

b. Definition of E(X|Y) when X \in L^2, F_P i.e. E(|X|^2)

c. General Definition of E(X|Y)

II. Theorems/Propositions:

a. (P.2.1.2) Projection Theorem:

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

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

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

e. (T.2.3.13,T.2.3.14) Monotone and Dominated Convergence for Conditional Expectations
Chapter 3: General Theory of Stochastic Processes

I. Definitions:

a. (D.3.1.1) Stochastic Process  ................................................................. 19
b. (D.3.1.2) Random Walk  ................................................................. 19
c. (E.3.1.4) S.P. with Independent Increments  ......................................... 19
d. (D.3.1.5) Finite Dimensional Distribution Function:  .......................... 19
e. (D.1.3.12) Consistency of FDD ......................................................... 19
f. (D.3.1.14) Sigma-Field of a S.P .......................................................... 19
g. (D.1.3.7) Versions: ................................................................. 19
h. (D.1.3.8) Modifications: ................................................................. 19
i. (E.3.3.7) Indistinguishable sample path (a.s.): .................................. 19
j. (D.3.2.1) Characteristic Function of a Random Vector  ....................... 20
k. (D.3.2.4) PDF of a Random Vector .................................................. 20
l. (D.3.2.6) Positive Semi-definite Matrix .............................................. 20
m. (D.3.2.7) Gaussian Distribution .......................................................... 20
n. (P.3.2.5) Gaussian Density ................................................................. 21
o. (D.3.2.14) Gaussian Stochastic Process ............................................. 21
p. (D.3.2.19) Strong and Covariance/Weak Stationary Process  ............... 21

II. Theorems/Propositions:

a. Modification → Versions ................................................................. 22
b. X(t), Y(t) modifications of each other does not imply that they’re indistinguishable. ................................................................. 22
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. ................................................................. 22
d. (P.3.2.2) Characteristic Function Uniquely Determines the Law $P_x$ of a Random Vector ................................................................. 22
e. (P.3.2.5) Properties of Characteristic Functions ....................................... 22
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 ................................................................. 22
g. (C.3.2.15, P.3.2.16) Properties of Gaussian Processes ................................... 22
h. (T.3.3.16) Fubini’s Theorem ................................................................. 23

Chapter 4: Martingales and Stopping Times

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 $F$ ................................................................. 24

That is, $F_0 \subset F_1 \subset \cdots \subset F_n \subset \cdots$ where $F_n$ is a sigma field for each n. ................................................................. 24
b. (D.4.1.2) S.P. Adapted to Filtration: A (discrete time) S.P. \( X_{n,n} \) is adapted to a filtration \( F_n \) if \( X_{n,n} = X_n(\omega) \) for each \( n \).

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

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

e. (D.3.1.14) L2 Martingale

f. Martingale Differences

g. (D.4.1.11) Previsible / Predictable Process

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

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

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

k. (D.4.2.9) Right-Continuous Filtration

l. (D.4.3.1, D.4.3.11) Stopping Time \( t \) for a filtration \( \{F_t\} \)
m. (D.4.3.5) Stopped Process

n. (D.4.4.2) Innovation Process

o. (D.4.4.8) Increasing Part of a M.G. \( \{M_t\} \)
p. (D.4.6.1) Branching Process (We use MGs to study the extinction probabilities of branching processes)

q. Probability of Extinction

Theorems

a. is adapted to a filtration \( F_n \) iff

\[
\begin{align*}
X_0, \ldots, X_n, F_n, n
\end{align*}
\]

b. (P. 4.1.7):

c. (T.4.1.12): Martingale Transform

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

e. (R.4.1.21) SubMG (SupMG) have non-decreasing (non-increasing) expectation \( E(X_n) \)

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

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

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

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

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

k. (T.4.4.1) Doob’s Decomposition

l. (E.4.4.3) Doob’s Decomposition for SubMG (A, non-decreasing) and SupMG (A, non-increasing)

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

n. (T.4.4.11) Doob’s Inequality

o. (T.4.5.1) Doob’s Martingale Convergence Theorem
p. (P.4.5.3) L2 and A.S. convergence of MGs (A Stronger Convergence Theorem than Doob’s) .......................................................... 32
q. (P.4.6.2) SP Xn=m’Zn is a MG for the filtration {Fm}, n a Branching Process .......................................................... 32
r. (P.4.6.3) Sub-Critical Process Dies Off / Extinction Probability is One when m < 1 .................................................. 32
s. (P.4.6.5) Critical Process Dies Off / Extinction Probability is One when m = 1 .................................................. 32

Chapter 5: Brownian Motion ................................................................................................................................. 33

Definitions .................................................................................................................................................................. 33
a. (D.5.1.1, T.5.2.1) Brownian Motion ................................................................................................................. 33
b. Comparison between Variation of Nice Calculus and Stochastic Calculus: ................................................................. 33
c. (D.5.3.1) | | | (Length of Longest Interval) and Q-th Variation of f(·) on CALCULUS DEFINITION) .................. 33
d. (D.5.3.2) Q-th variation of a S.P. {X(t)} on the interval [a,b] ................................................................................ 33
e. Lipschitz Sample Path (w.p.1) ......................................................................................................................... 33
f. (D.5.3.3) Quadratic Variation ......................................................................................................................... 34

Theorems .................................................................................................................................................................. 35
a. (P.5.1.2) Brownian Motion has Independent Increments of Zero Mean ........................................................................... 35
b. (E.5.1.4) Interesting Properties of a BM ........................................................................................................... 35
c. (T.5.2.1) Levy’s Martingale Characterization of BM ............................................................................................. 35
g. (E.5.1.5,E.5.1.6) BM-Related Processes and their properties ................................................................................. 35

(SEE PS5 for Work) ............................................................................................................................................... 36
d. (P.5.2.2) X_1 \ W \ W a BM USEFUL FOR REFLECTION PRINCIPLE) ................................................................. 36
e. W(t) \ :max_0 \ T W(z) (t) : T where \ inf(t 0 : W_t) is the First Hitting Time of BM ............................................ 36
f. Reflection Principle (p.106-107) .......................................................................................................................... 36
g. (P.5.3.9) Total Variation of BM W(t) is infinite w.p.1 .......................................................................................... 36
h. (D.5.3.4) For BM W(t), as \ | | | 0 we have that W(t)^2(b a) in 2-mean. .................................................. 37
i. Quadratic Variation of a BM W(t) is t, since its Doob-Meyer Decomposition is W(t)^2=[W(t)^2-t] + t ........................................... 37

Chapter 6: Markov, Poisson, and Jump Processes ............................................................................................ 38

Definitions .................................................................................................................................................................. 38
a. (D.6.1.1) Markov Chain ........................................................................................................................................ 38
b. (D.6.1.2) (Time) Homogeneous Markov Chain .................................................................................................. 38
c. (D.6.1.3, D.6.1.8) Stationary Transition Probabilities (Determines distribution of Homogeneous MCs) ................. 38
d. (D.6.1.4) Initial Distribution of a Markov Chain ................................................................................................ 39
e. 4 Useful Conditions for Defining Poisson Processes: ................................................................................................. 39
f. (D.6.2.1) Counting Process (C0) .......................................................................................................................... 40
g. (D.6.2.1) Jump Times ........................................................................................................................................... 40
h. Poisson R.V. .......................................................................................................................................................... 40
S.P. with Independent Increments ................................................................. 40

(D. 6.2.3) S.P. with Stationary Increments ..................................................... 40

(P.6.2.5) Memoryless Property of Exponential Law ........................................ 40

(D.6.2.1, P.6.2.4) Poisson Process ................................................................. 40

Theorems ............................................................................................................ 41

a. Showing that a SP is a Markov SP ............................................................... 41

b. Showing that a Markov SP is Time-Homogeneous ...................................... 41

c. (Lemma 1) If \( \{X(t)\} \) is a S.P. with independent increments, then \( \{X(t)\} \) is a Markov SP. ................................................................. 41

d. (Lemma 2) If \( \{X(t)\} \) is a stationary process and a Markov Process, then it is Homogeneous ......................................................... 41

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. 41

f. (P.6.1.13) Every Continuous Time SP \( \{X(t)\} \) with Stationary and Independent Increments is a Homogenous Markov Process. ................................................................. 42

g. (P.6.1.5) Strong Markov Property ............................................................... 42

h. (P.6.2.4) Poisson Process is the only S.P. with stationary independent increments that satisfies condition C0 (counting process) ................................................................. 42

i. (P.6.2.5) Memoryless Property of Exponential Law ........................................ 42

j. (P.6.2.6) A S.P. \( N(t) \) that satisfies C0 is a Poisson process of rate \( \lambda \) iff it satisfies C3 ................................................................................ 42

k. (P.6.2.8) Relationship between Poisson Process and Uniform Measure ................................................................. 42

l. (T.6.2.10) Poisson Approximation ............................................................... 42

m. (P.6.2.12) \( N_1(t) + N_2(t) \) Poisson if \( N_1(t), N_2(t) \) independent Poisson ........................................................................... 43
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
\[ \limsup_n X_n = \lim \sup_k x_k \quad \text{and} \quad \liminf_n X_n = \lim \inf_k x_k \]
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, \( (A_i, i.o.) \) \( (A_i \text{ occurs infinitely often}) \)
\[ \limsup A_i = \bigcap_{m \geq 1} \bigcup_{n \geq m} A_n \]
(\text{Intuition: For } n > 1, \text{ Limsup is in ONE of the tail } A_n \text{'s. This is true for all } n. \)
\[ = m, \ n(\ ) \quad m \text{ s.t. } A_n(\ ) \]
\[ = A_n \text{ for infinitely many } n \]

Idea: If \( \text{LimSup } A_i \) \( \bigcap_{m \geq 1} \bigcup_{n \geq m} A_n \) \( \bigcup_{n \geq 1} A_n \) AND \( \bigcup_{n \geq 2} A_n \) AND ... \( \bigcup_{n \geq k} A_n \) \( k \).
That is, \( A_i \) is in \( {-m} \) many \( A_k \)’s. So, a collection of such a collection of \( A_i \) is a collection of \( s.t. A_k \)’s occur i.o.

\( (A_i, \epsilon.v.) \) \( (A_i \text{ eventually}) \)
\[ \liminf A_i = \bigcup_{m \geq 1} \bigcap_{n \geq m} A_n \]
(\text{Intuition: For } n > 1, \text{ Liminf is in ALL of the } A_n \text{'s. This is true for all } n. \)
\[ = m \text{ s.t. } n > m, \quad A_n(\ ) \]
\[ = A_n \text{ for all large } n \]

Idea: If \( \text{lim Inf } A_i \) \( \bigcup_{m \geq 1} \bigcap_{n \geq m} A_n \) \( \bigcap_{n \geq 1} A_n \) AND \( \bigcap_{n \geq 2} A_n \) AND ... \( \bigcap_{n \geq k} A_n \) \( k \).
That is, \( A_i \) is in all the \( A_k \)’s eventually. So, a collection of such a collection of \( A_i \) is a collection of \( s.t. A_k \)’s occur eventually.

(See notes)

d. Metric Spaces
e. Normed Spaces
f. Inner Product Spaces
g. Stochastic Order
\( O(.) : \text{As } x \to ?, \ f(x) = O(g(x)) \iff C \text{ such that eventually } |f(x)| \leq C|g(x)|. \text{ Or, the fraction } \frac{f(x)}{g(x)} \text{ is eventually bounded as } x \text{ approaches limit.} \)
\( o(.) : \text{As } x \to ?, \ f(x) = o(g(x)) \iff \lim_{x \to ?} \frac{f(x)}{g(x)} = 0. \text{ (i.e. } f(x) \text{ goes to 0 faster than } g(x) \text{ as } x \text{ approaches the limit.} \)
Op(.):
Manipulating this notation:
Most of what we have here is common sense. Some useful ideas:
\[ O(g_1(x)) + O(g_2(x)) = O(\max(g_1(x), g_2(x)) \]
\[ O(g_1(x)) O(g_2(x)) = O(g_1(x) g_2(x)) \]

We’d like to be able to say that for reasonable functions \( w, w(O(g(x))) = O(w(g(x)). \text{ There’s no problem with saying } (O(h))^2 = O(h^2), \text{ but } e^{O(ln x)} \text{ isn’t well defined. You have to be careful there.} \)
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: \( \frac{1}{2 + O(x)} = \frac{1}{2} + O(x) \text{ as } x \to 0. \)
Chapter 1: Probability Measure and Integration

I. Definitions:

a. Probability Space

Probability space is a triple \( \Omega, \mathcal{F}, P \) where:

- Sample space \( \Omega \) the set of all possible outcomes of some random experiment or phenomenon.
- Event space \( \mathcal{F} \) a subset of \( 2^\Omega \) (set of all subsets), consisting of all "allowed events". Or, events to which we shall assign prob's.
- Probability Measure \( P = \text{a set function} : \mathcal{F} \rightarrow [0, 1] \)

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

We say that \( \mathcal{F} \) is a -field/ -algebra if

- (a) \( A \in \mathcal{F} \) \( \Rightarrow \) \( \Omega \in \mathcal{F} \) (Complement Rule)
- (b) \( A \in \mathcal{F} \) \( \Rightarrow \) \( A \in \mathcal{F} \) (Countable Union Rule)
- (c) \( A \in \mathcal{F} \) \( \Rightarrow \) \( \Omega \in \mathcal{F} \) \( \Rightarrow \) \( \mathcal{F} \) is always in -field

De Morgan’s Laws:

\[ A \cup B = \neg (\neg A \cap \neg B) \]

\[ A \cap B = \neg (\neg A \cup \neg B) \]

Equivalently:

\[ (a, b) : a \cup b \quad [a, b) : a, b \quad (a, b) : b \quad (b, a) : a \quad Q \quad O \quad \text{open} \]

Note on Borel and \( \mathcal{X} \): 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 \( F = \mathcal{X}(\ ) \rightarrow \mathcal{R} \) so that in the output space we have all the Borel sets (because then we have in the output space \( (a, b) : a, b \rightarrow R \) which is just Borel).

---

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

Pf: Let \( \Omega \).

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

Recall, binomial th. says \( \binom{N}{i} = \frac{N!}{i!(N-i)!} \) by De Morgan's Law (Countable Intersection Rule)

\[ \therefore (\ ) \text{ always in -field} \]

By De Morgan's Laws:

\[ A \cap B = \neg (\neg A \cup \neg B) \]

\[ A \cup B = \neg (\neg A \cap \neg B) \]

Equivalently:

\[ (a, b) : a \cup b \quad [a, b) : a, b \quad (a, b) : b \quad (b, a) : a \quad Q \quad O \quad \text{open} \]

Thus, \( \bigcup_i A_i \cap \bigcup_i A_i^c \cap F \quad \cap_i A_i^c \quad F \quad \cap_i A_i \quad \bigcup_i A_i^c \quad F \) by Complement Rule

\[ \bigcup_i A_i \cap \bigcup_i A_i^c \quad F \text{ by Complement Rule} \]

\[ \bigcup_i A_i \quad F \text{ by Countable Union} \]

\[ \bigcup_i A_i^c \quad F \text{ by Complement.} \]
g. **(D.1.2.8) Sigma Field Generated by a RV:**
Given RV $X$, we denote by $(X)$ the smallest field $\mathcal{G}_F$ s.t. $X(\omega)$ is measurable on $\sigma\mathcal{G}_F$. We call $(X)$ the field generated by $X$, sometimes denoting by $F_X$.
It can be shown that:

$$X \subseteq (X) \subseteq \Omega$$

| Note on $(X)_\sigma$ | Naturally, we may want to consider $\sigma(X) = X(\omega)$ for all Borel sets $B$. But, it can be shown that $\sigma(X) = (X)$. |
---|---|
Note on $(X)$ and information: $\sigma(X)$ 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, $(X)$ contains exactly those events $A$ for which we can say whether $A$ or not. So, if we knew whether all the sets in $(X)$ 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 $X_1, \ldots, X_n$ on the same probability space $(\Omega, F)$, denote by $X_k, k = n$ the smallest field $\mathcal{G}_F$ s.t. $X_k(\omega)$ are measurable on $(\Omega, F)$.
Thus,

$$X_k, k = n \cap \bigcup_{k=1}^{n} \mathcal{H}_k$$

(where $\mathcal{H}_k$ are sigma fields that contain $X_k$ i.e. smallest field containing $(X_k)_{k=1,\ldots,n}$)

| i. **(D.1.2.1) Random Variable:** |

A R.V. on $(\Omega, F)$ is a function $X: \Omega \to R$ s.t. $R$, the set $X(\omega) \in F$ (called F-measurable function)
(Equivalently: $X: \Omega \to R$)

| Note on F-Measurability: Why do we care about this? So that we can induce a probability measure $P_x$ on $\mathbb{R}$.

i.e. $A \subseteq \mathbb{R}, P_x A : X(\omega) = A$ is well-defined because $X(\omega) = A \subseteq F$ and thus we can assign a probability measure! |

| Note on Choice of F: In defining RV, F is implicit (i.e. X on $(\mathbb{W,F,P})$ 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) $I_A(X(\omega)) = \begin{cases} 1, & A \\ 0, & \text{a.w.} \end{cases}$ 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. $X_N = \sum_{n=1}^{N} c_n I_{A_n}(\omega)$ (for finite $N$, non-random $c_n \subseteq \mathbb{R}$, and sets $A_n \in F$) a simple function, denoted by $X_N \subseteq SF$. |

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

(Can show by showing $\text{Var}(X-Y) = 0$) |

| m. **(D.1.2.11) Borel Measurable Function:** |

A function $g: \mathbb{R} \to \mathbb{R}$ is called Borel (measurable) function if $g(\omega)$ is a R.V. on $(\mathbb{R}, F)$ so that $x : g(x) = B$

| 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 $X(\omega)$ is denoted $EX$.

$$EX = \lim_{n} \sum_{k=1}^{n} \int_{k}^{k+1} x \cdot I_{X_k}(x) P : X(\omega) = \int_{k}^{k+1} x \cdot I_{X_k}(x) P$$

(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 $\max X, 0$ and $\min X, 0$.

Then,

$$[X - X, X + |X|, X - |X|, \text{ and EX } E|X| \leq \infty (\text{provided either EX or } E|X|)$$

Note on Expectation Being Undefined: If $E|X|$ and $E|X|$ are both $\infty$, then the expectation is undefined. Where as if only one 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, $E|X| < \infty$ precisely to avoid the un-defined expectation problem. $E|X|$ is always defined (since inf + inf = inf), whereas $EX$ is not always defined (since inf – inf = undefined. E.g. $\lim 1/n - \lim 1/n = 0$ but $\lim 2/n - \lim 1/n = \infty$).

p. (D.1.2.25) Integrable R.V.:
A R.V. $X$ is integrable (or has finite expectation) if $E|X| < \infty$ (i.e. both EX and EX )

$E|X| M \leq 0$ (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 $X$, $I$ is called Uniformly Integrable (U.I.) if $\lim_{M} \sup E|X|_I = 0$

OR EQUIVALENTLY
A collection of RV's $X$, $I$ is U.I. iff
(a) $\lim sup E|X|_I = 0$
(b) $>0$, $>0$ s.t. $\sup E|X|_I < \infty$ and events $A$ s.t. $P(A) < \epsilon$

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 $X(n) \xrightarrow{p} X$, and $|X(n)|^q$ uniformly integrable, then $X(n) \xrightarrow{q.m.} X$.

r. Sufficient Conditions for Uniform Integrability of a Collection of R.V.’s
(Lecture 4 pg.8)

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

b) If $C < \infty$ s.t. $X_n$ C a.s. n, then $X_n$ U.I.

c) If $X_i$, $L^q$ for $i=1,\ldots, n$ (a finite collection of R.V.), then $X_1,\ldots, X_n$ U.I.

 d) If $\sup_n E|X|_I$ for some $>0$, 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: $X_n \xrightarrow{p} X$ if
$X_n(\ ) \xrightarrow{a.s. n} X(\ )$ for each $A$

P : $\lim_{n}X_n(\ ) X(\ ) 1 0 P : |\lim_{n}X_n(\ ) X(\ )| 1 0 P : |\lim_{n}X_n(\ ) X(\ )| 0

0, P $\cupk_n : [X_k(\ ) X(\ ) 0, lim \cupk_n : [X_k(\ ) X(\ ) 0, lim \cupk_n : [X_k(\ ) X(\ ) 0,

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

0, P $\cupk_n : [X_k(\ ) X(\ ) 0, lim \cupk_n : [X_k(\ ) X(\ ) 0,

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

(2) Probability: $X_n \xrightarrow{p} X$ if $P : |X_n(\ ) X(\ )| 0 0$ fixed

(3) Lq/q-Mean: $X_n \xrightarrow{q.m.} X$ if $X_n, X L^q$ and $|X_n - X|_q n 0$ i.e. $E |X_n - X|_q n 0$

(4) Distribution/Law/Weakly: $X_n \xrightarrow{D} X$ if $F_X(\ ) n F_X(\ )$ for each that is a continuity point of $F_X$.

(or equivalently) $X_n \xrightarrow{D} X$ iff $h$ continuous and bounded (on the range of $X$), we have $E h(X_n) n E(h(X))$

(5) Expectation: $X_n \xrightarrow{E} X$ if $E(X_n) n E(X)$

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 2\textsuperscript{nd} 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.

\textbf{t. (D.1.3.4) Complete Probability Space}

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

\textbf{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. \( X_n \rightarrow X \) for each \( \omega \) and \( P(A) = 1 \), then there exists a RV \( \hat{X} \) s.t. \( N = \{ \omega : X_n(\omega) \rightarrow \hat{X}(\omega) \} \) is a subset of \( B = A^C \). By assuming that the probability space is complete, we guarantee that \( N \) is in \( F \), and therefore \( X_n \rightarrow \hat{X} \) is necessarily a random variable.

\textbf{u. (D.1.3.2) \( L^q \) space}

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

Note on convergence: Convergence in \( L^q \) is the convergence of RV’s in the \( L^q \) space (defined above), with the \( \| \cdot \|_q \) metric.

\textbf{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 \( L_q \) (P.1.3.16):

- \( L^q \), \( F, P \) is a complete, normed (topological) vector space (i.e. \( X + Y \) whenever \( X, Y \in L^q \), \( \lambda X \) \( \in \), \( R \))
- The norm \( \| \cdot \|_q \) \( E|X|^q \). If \( \| X_n - X_m \|_q \to 0 \) then \( X_n \rightarrow \) \( q \) \( m \) \( X \) for some \( X \) in \( L^q \).

\textbf{v. (D.1.4.1) Law of a R.V.}

The law of a RV \( X \), denoted \( P_X \), is the probability measure on \( (\Omega, \mathcal{B}, \mu) \) s.t. \( P_X(\Omega) = 1 \).

(\textit{In other words,} \( P_X \) is the measure induced by \( X \) on \( \Omega \).

\textbf{w. (D.1.4.4) Distribution Function of a R.V.}

The distribution function \( F_X \) of a real-valued R.V. \( X \) is \( F_X : \Omega \rightarrow [0,1] \).

Note on Existence of \( F_X \): Since by definition \( X \) must be measurable w.r.t. \( (\Omega, \mathcal{B}, \mu) \), 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).

\textbf{x. (D.1.2.22,P.1.4.7) Probability Density Function (PDF)}:

A R.V. \( X \) has a PDF \( f_X \) if \( P(a < X < b) = \int_a^b f_X(x)dx \) \( a < b \). Such a density must be a non-ive function with \( \int_{-\infty}^\infty f_X(x)dx = 1 \).

OR Equivalently

A R.V. \( X \) has a PDF \( f_X \) iff its distribution function \( F_X \) can be expressed as

\[ F_X(\cdot) | f_X(x)dx \]

where \( f_X \) is 0 and \( f_X(x)dx \) is 1.

Note on Non-Existence of PDF:

If PDF exists, then it must be that \( F_X(b) = \int_a^b f_X(x)dx \), and thus by Fundamental Th. Of Calculus, \( F_X(b) = \int_a^b f_X(x)dx \). 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,

\[ F_X(b) = \int_{X(\omega) < b} \mu(\omega) \, dp(\omega) \]

Note on Existence of PDF and Property of CDF: If PDF exists, then we know that \( F_X \) is continuous and almost everywhere differentiable with \( F_X'(b) = f_X(x) \) for almost every \( x \).
Note on PDF’s, Lebesgue 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 $X(\omega)$ into finitely many small intervals and approximating $X(\omega)$ by a constant on the preimage of each of the intervals. This allows us to deal with rather general domain. In contrast, the Riemann integral splits the domain of integration into finitely many small intervals (hence we’re limited to $\mathbb{R}^d$). Even when $\Omega = [0,1]$ it allows us to deal with measures for which $P$ for which $P([0,1])$ 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) $xf_X(x)dx \overset{\text{Lebesgue}}{=} X(\omega) dP(\omega)$

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 $A_i$ of $F$ are $P$-mutually independent if for any $L < \infty$ and distinct indices $i_1, \ldots, i_n$,

$$P(A_{i_1} \cap \ldots \cap A_{i_n}) = P(A_{i_1}) \cdots P(A_{i_n})$$

2) Independence of $\sigma$-Fields: Two $\sigma$-fields $H,G$ of $F$ are $P$-independent if

$$P(g \circ h \in G, h \in H) = P(g \in G, h \in H) \forall g, h$$

3) Independence of R.V.'s: Two R.V's $X,Y$ defined on $\Omega$, $F$ are independent if $X,Y$ (where $X,Y$ $F$)

OR EQUIVALENTLY

$$P(X \leq x, Y \leq y) \overset{\text{Lebesgue}}{=} P(X \leq x)P(Y \leq y)$$

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

$$E[h(X_1, \ldots, X_m)g(Y_1, \ldots, Y_n)] = E[h(X_1, \ldots, X_m)]E[g(Y_1, \ldots, Y_n)]$$

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

$$A \overset{\text{Lebesgue}}{=} A^c \overset{\text{Lebesgue}}{=} B \overset{\text{Lebesgue}}{=} B^c \overset{\text{Lebesgue}}{=} A^c \overset{\text{Lebesgue}}{=} B^c$$
II. Theorems/Propositions:

a. **(E.1.1.3) Properties of Probability Measures**

   Let \( F, P \) be a probability space and \( A, B, A_i \) events in \( F \). Then, \( P \) satisfies:

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

   (b) Sub-additivity: If \( A \subseteq \bigcup_{i} A_i \) then \( P(A) \leq \sum_{i} P(A_i) \)

   (c) Continuity from Below: If \( A \subseteq \bigcap_{i} A_i \) (i.e. \( \bigcap_{i} A_i \subseteq \Omega \)), then \( P(A) = \lim_{i} P(A_i) \)

   (d) Continuity from Above: If \( A \subseteq \bigcup_{i} A_i \) (i.e. \( \bigcup_{i} A_i \subseteq \Omega \)), then \( P(A) = \lim_{i} P(A_i) \)

   (e) Inclusion-Exclusion Rule: \( P(\bigcup_{i} A_i) = \sum_{i} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \cdots ± \)

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

   Let \( A_i \) be a -field for each \( i \), an arbitrary (possibly uncountable) index set. Then, \( \bigcup_{i} A_i \) is a -field, but not necessarily one.

c. **(P.1.1.19) Borel set does not contain all subsets of \( \mathbb{R} \)**

   There exists a subset of \( \mathbb{R} \) that is not in Borel set. Thus, not all subsets of \( \mathbb{R} \) 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) \( I(\omega) = 0 \) and \( I(\omega) = 0 \)

   (b) \( I_{A^c}(\omega) = 1 \)

   (c) \( I_A(\omega) = I_B(\omega) \) iff \( A \subseteq B \)

   (d) \( I_{A \cap B}(\omega) = I_A(\omega) \)

   (e) If \( A_i \) disjoint then \( I_{\bigcup_{i} A_i}(\omega) = I_{A_i}(\omega) \)

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. \( X_n(\omega) \rightarrow X(\omega) \) for each fixed \( \omega \) as \( n \rightarrow \infty \)

   Use \( f_n(x) = n I_{x} + \frac{n^n}{k} I_{(k2^n,x)} + \frac{n^n}{k} I_{(x,k2^n)} \)

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

   \[ 0, \frac{1}{2^n}, \frac{1}{2^n}, \ldots, \frac{2^n}{2^n}, \frac{4^n}{2^n}, \ldots, \frac{n^n}{2^n} \]

   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 RV. on a measurable space is a RV. on that space

   Let \( \Omega, F \) be a measurable space and let \( X_n \) be a sequence of random variables on it. If \( X(\omega) = \lim_n X_n(\omega) \) exists and is finite. Then, \( X(\omega) \) is a RV on \( \Omega, F \).

   **Note on usefulness:**

   For a typical measurable space with uncountable \( \Omega \), 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 indeed a RV. (i.e. by showing that it’s a limit to a sequence of RV’s on \( \Omega, F \))

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

   If \( g: \mathbb{R} \rightarrow \mathbb{R} \) is a Borel measurable function and \( X_1, \ldots, X_n \) are R.V. on \( \mathbb{R}, F \), then \( g \circ X_1, \ldots, X_n \) also a R.V. on \( \mathbb{R}, F \)

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

   For any \( n \in \mathbb{N} \), \( g: \mathbb{R} \rightarrow \mathbb{R} \) Borel measurable and any R.V. \( Y_1, \ldots, Y_n \) on the same measurable space, then \( F_{g(Y_1, \ldots, Y_n)} \) \( (Y_1, \ldots, Y_n) \).

   If \( g \) invertible, then \( F_{g(Y_1, \ldots, Y_n)} = (Y_1, \ldots, Y_n) \).
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) \( E_X P(A) \) for any \( A \) \( F \)

(2) If \( X = \sum_{n=1}^{N} c_n A_n \) 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 \( c \), the R.V. \( X + Y \) is integrable and \( E(X + Y) = E(X) + E(Y) \)

(4) (Constant) \( E(X) = c \) if \( P(X = c) = 1 \)

(5) (Monotonicity) If \( X \) Y a.s., then \( E(X) \leq E(Y) \). 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. \( g(x+(1-\lambda)y) \leq (1-\lambda)g(x) + \lambda g(y) \) \( x,y \in \mathbb{R}, [0,1] \)

If X is an integrable R.V. and \( g(X) \) is also integrable, then \( E(g(X)) \leq g(E(X)) \) 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. \( X \), \( > 0, P \left| X \left( \right) \right| \leq \frac{1}{\int f(\cdot)} \int f(\left| X \right|) \)

(3) (Usual) Markov’s: Take above with \( f(x) = x \). Then, \( > 0, P \left| X \left( \right) \right| \leq \frac{E(Y-E(Y))^2}{2} Var(Y) \)

(4) Cauchy-Schwartz: Suppose X,Y R.V’s on the same probability space w/ \( E Y^2 \) and \( E X^2 \). Then, \( E|YZ| \leq \sqrt{E Y^2 E Z^2} \)

Note: In particular, for \( X,Y \) with mean 0, \( \text{Jensen’s} |EX| \) \( \text{Cauchy Schwarts} \sqrt{E(Y^2)E(X^2)} \) OR, \( \text{Cov}(X,Y) \sqrt{\text{Var}(X)\text{Var}(Y)} \)

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

Using Euclidean \( d(x,y) = |x-y| \):

\[
\begin{align*}
\| x \| &= \sqrt{\sum_{n=1}^{N} x_n^2} \\
\| y \| &= \sqrt{\sum_{n=1}^{N} y_n^2} \\
\| x - y \| &= \sqrt{\sum_{n=1}^{N} (x_n - y_n)^2}
\end{align*}
\]

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) \( X_n \text{ a.s. } X \) \( X_n \rightarrow X \) (and converse not true)

(b) (BUT) \( X_n \text{ p } X \) subsequence \( n_k \) s.t. \( X_{n_k} \text{ a.s. } X \) for \( n \)

a.s. and q.m. convergence

(c) \( X_n \text{ q.m. } X \) does not imply \( X_n \text{ a.s. } X \) and vice versa.

(d) If \( X_n \text{ q.m. } X \) and \( X_n \text{ a.s. } Y \) then \( X \text{ a.s. } Y \)

q.m. and r.m. convergence

(e) If \( X_n \text{ q.m. } X \) \( X_n \text{ r.m. } X \) \( r \approx q \)

q.m. and p convergence

(f) \( X_n \text{ q.m. } X \) \( X_n \text{ p } X \) (Converse not true except following)

(g) \( X_n \text{ p } X, |X_n|^q \) U.I. \( X_n \text{ q.m. } X \)

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

other convergences and d convergence

(h) \( X_n \text{ p } X \) \( X_n \text{ d } X \) (Converse not true, except for special case when \( X \) is nonrandom constant a.s.)

\[
\begin{align*}
X_n &\text{ d } X \ X_n &\text{ d } X \\
X_n &\text{ a.s. } X \ X_n &\text{ d } X
\end{align*}
\]

(i) If \( X_n \text{ d } X \) and \( X \) is nonrandom constant a.s. \( X_n \text{ p } X \)
Property of the $\| \cdot \|_q$ Norm: $\|X\|_q$ is nondecreasing in $q$.

**L.1.3.10, L.1.3.11** Borel Cantelli Lemmas

BCI: Let $A_k$ and $P(A_k)$ be the probability of $A_k$. Then, $P(\bigcap_{k=1}^\infty A_k) = \lim_{n \to \infty} P(\bigcup_{k=n}^\infty A_k) = P A_k \ i.o.$

BCII (converse with independence): If $A_k$ independent and $P(A_k)$, then $P(\bigcap_{k=1}^\infty A_k) = \lim_{n \to \infty} P(\bigcup_{k=n}^\infty A_k) = P A_k \ i.o.$

**Note on Intuition:** BCI 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:** $E(X_n^2) \to 1$ s.t. $1/n X_n \to 0$ a.s.

Pf: WTS: $P \lim_{n \to \infty} \frac{1}{n} X_n = 0$ or equivalently $\lim_{n \to \infty} P(\left\{ \frac{1}{n} X_n > 0 \right\}) = 0$

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

Using BCI, note $P A_k \to \lim_{n \to \infty} P \left( \bigcup_{k=n}^\infty A_k \right) = \lim_{n \to \infty} \frac{E X_k}{k^2/2} = \frac{1}{k^2}$ by Chebychev.

Then, $P A_k \to \lim_{n \to \infty} \frac{1}{k^2} \frac{1}{k} = \frac{1}{k^2}$

Thus, $P \lim_{n \to \infty} \frac{1}{n} X_n = P \lim_{n \to \infty} \frac{1}{n} X_n = 0$. So, $1/n X_n \to 0$ 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 $E \lim X_n \approx E X_n$

(0) Fatou’s Lemma: If $X \in L^1$ and $X_n \to X$, then $\liminf_n E X_n \leq E \liminf_n X_n$

where $\liminf_n X_n$ denotes $\inf_{k \to \infty} \lim_{n \to \infty} X_n$

(1) Dominated Convergence: Suppose $X_n \to X$ for a.e. $\omega \in \Omega$ (or $X_n \to X$ for a.e. $\omega \in \Omega$) for a.e. $\omega \in \Omega$.

(2) Bounded Convergence: Suppose $X_n \to X$ for a.e. $\omega \in \Omega$ (or $X_n \to X$ for a.e. $\omega \in \Omega$).

(3) Monotone Convergence: Suppose $X_n \to X$ for a.e. $\omega \in \Omega$ (or $X_n \to X$ for a.e. $\omega \in \Omega$).

Note on Fatou’s Lemma: Fatou’s is useful because it tells us something about the expectations even if we don’t know that $X(n) \to X$, or even if we don’t know that the limit of the expectation exists. Since $\{E(X_n)\}$ is just a sequence of reals, then $\liminf / \limsup \ always \ exists \ by \ completeness \ of \ R$. Now, if a limit to $E(X_n)$ exists, then $\liminf$ $E(X_n)$ and $\limsup$ $E(X_n)$ 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 $X_n \to X, X_n \to q.m. X$

**o. (T.1.4.22)** Uniform Integrability + Convergence in Prob $\rightarrow q.m.$ Convergence

If $X_n \to X$ and $|X_n|^q$ are U.I., then $X_n \to q.m. X$ i.e. $E X_n^q \to E X^q$.

Note on Dominated Convergence as a Special Case: Dominated convergence is a special case of $X_n \to X, X_n \to q.m. X$.
Dom. Convergence is a special case because if $\mathbb{E} X_n \to X$, then we have $X_n \to X$ in prob, and we also have convergence in prob, then we have $L^1$ convergence by theorem, which gives us the result of dominated convergence.

Note on Versions of this Theorem: Some versions of this uses $X_n \to X$. 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 $X^q$.

(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 $\mathcal{F}, P$ and let $g$ be a Borel function on $\mathbb{R}$. Suppose either $g$ is nonnegative or $\mathbb{E}[g(X)] < \infty$. Then,

$$\mathbb{E}[g(X)] = \int g(x) \, dP_X(x)$$

Note on Borel-Measurability:

$g(x): (R, B)$ Borel measurable, $x: g(x) : B$. 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.

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

(1.4.3) $X \sim Y \Rightarrow \mathbb{E}[h(X)] = \mathbb{E}[h(Y)]$ for all $h$ bounded and Borel measurable.

We can show this using the delta method.
Chapter 2: Conditional Expectation and Hilbert Spaces

1. Definitions
   a. Expression for Conditional Expectation for Discrete R.V. Y:
   \[ E[Y | X] = \sum_{y \in Y} y \cdot P(Y = y | X) \]
   \[ = \int_{\omega \in \Omega} y \cdot P(Y = y | X(\omega)) \, d\omega \]
II. Theorems/Propositions

a. (P.2.1.2) Projection Theorem:
There exists a unique (a.s.) optional \( Z \in L^2 \), \((Y),P\) s.t. \( E[X Z^2] = \inf_{H \in \mathcal{F}_X} E[X W^2] \).

Further, the optimality of \( Z \) is equivalent to the orthogonality property \( E(XZ) = 0 \), \( V \in H_Y \).

b. (T.2.1.6) Consistency of \( L^2 \) and General Definition of \( E(X|Y) \):
The C.E. of integrable R.V. \( X \) given any \( \mathcal{F} \)-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 \( E(XZ) = 0 \), \( V \in L^2 \), \( G,P \).

(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,E.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 \( X \) ind \( Y \), \( E(X|Y) = E(X) \)
2. (Trivial -Field) If \( F_0 = \emptyset \), then \( E(X|F_0) = E(X) \) for any integrable \( X \)
3. (Linearity) Let \( X,Y \in L^1 \), \( F,P \). Then, \( E(X|G) E(Y|G) = E(X+Y|G) \) for any \( G \), \( R \)
4. (Tower Property) Suppose \( H \subseteq G \) (\( H \) contains less information than \( G \)) and \( X \in L^1 \), \( F,P \).

Then, \( E(X|G) = E(E(X|H)|G) \) (smaller set wins!)

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

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 \( f:R \rightarrow R \) be a convex function. Suppose \( X \in L^1 \), \( F,P \) is s.t. \( E|f(X)| < \infty \). Then, \( E(f(X)|G) \leq f(E(X|G)) \)

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 \( \text{Var}(Y|G) = E\ (Y^2|G) = E\ (Y|G)^2 \) where \( Y \in L^2 \), \( F,P \) and \( G \) a \( \mathcal{F} \)-field. Then, \( \text{Var} Y E(\text{Var}(Y|G)) = \text{Var}(E(Y|G)). \)
I. Definitions:

a. (D.3.1.1) Stochastic Process
Given $F, P, \omega$, a stochastic process $X_t$ is a collection $X_t : t I$ of RVs where the index $t$ belongs to the index set $I$. If $I$ is an interval in $\mathbb{R}$, we say it's continuous time. If $I$ a subset of $\mathbb{Z}$ we say it's discrete time.

Fixing $\omega$, 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 $F, P, \omega$, a stochastic process $X_t$ is a collection $X_t : t I$ of RVs where the index $t$ belongs to the index set $I$. If $I$ is an interval in $\mathbb{R}$, we say it's continuous time. If $I$ a subset of $\mathbb{Z}$ we say it's discrete time.

Fixing $\omega$, 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 $\xi_i$ are iid RV defined on the same probability space $\omega$.

Special Case: If $i, 1, i \in \mathbb{Z}$, then $S_n$ is a simple random walk.

If $i, Z$, 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 $X_s, 0 \leq s \leq t$

(d. (D.3.1.5) Finite Dimensional Distribution Function:
Given $N$ and a collection $t_1, t_2, ..., t_N$ in $I$, we denote the joint distribution of $X_{t_1}, ..., X_{t_N}$ by $F_{t_1, ..., t_N}(\cdot)$.

so $F_{t_1, ..., t_N}(\cdot)$ is the finite dimensional distribution of the S.P.

We call the collection of functions $F_{t_1, ..., t_N}(\cdot)$ 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

$$\lim_{k \to \infty} F_{t_1, ..., t_N}(\cdot) = F_{t_1, ..., t_N}(\cdot)$$

for any $1 \leq k \leq N$, $t_1, ..., t_N$, and $I$ and $R$

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.3.1.7) Versions:
S.P.'s $X_t$, $Y_t$ are versions of one another if they have the same f.d.d.

h. (D.3.1.8) Modifications:
A S.P. $Y_t$ is a modification of another S.P. $X_t$ if $P : X_t(\omega) \Rightarrow Y_t(\omega)$ 1 t 1

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 $P : X_t(\omega) \Rightarrow Y_t(\omega)$ 1 t 1
j. **(D.3.2.1) Characteristic Function of a Random Vector**

A random vector \( X = X_1, X_2, \ldots, X_n \) with values in \( \mathbb{R}^n \) has the characteristic function

\[
\phi_X(t) = E \left[ e^{jX} \right] = \int_{\mathbb{R}^n} e^{jtx} f_X(x) \, dx
\]

where \( j = \sqrt{-1} \) and \( t = (t_1, \ldots, t_n) \).

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

**Note on Properties:** \( \phi_X(0) = 1 \) and \( \phi_X(t) \) is bounded on \( \mathbb{R}^n \) (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 \( f_X(x) \), then the characteristic function is

\[
\phi_X(t) = E \left[ e^{jX} \right] = \int_{\mathbb{R}^n} e^{jtx} f_X(x) \, dx
\]

which is just the Fourier transform of the density function \( f_X(x) \).

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

A random vector \( X = X_1, X_2, \ldots, X_n \) has a PDF \( f_X(x) \) if for every \( a_1, b_1, \ldots, a_n, b_n \),

\[
P \left[ a_1 < X_1 < b_1, \ldots, a_n < X_n < b_n \right] = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f_X(x_1, \ldots, x_n) \, dx_1 \cdots dx_n
\]

Note: Such a function (also called the joint density of \( X(1) \ldots X(n) \)) 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 nonzero vector \( \mathbb{R}^n, A \mathbf{x} = 0 \).

m. **(D.3.2.7) Gaussian Distribution**

We say that a random vector \( X = X_1, X_2, \ldots, X_n \) has a Gaussian distribution if

\[
X - \mu \sim \mathcal{N}(\mu, \Sigma)
\]

for some psd symmetric nxn matrix \( \Sigma \) (VarCov matrix), some \( \mu \) (vector of means) and all \( \mathbb{R}^n \).

**Note on R:**

A R.V \( X \) is Gaussian if for some \( \mu, \Sigma \), and all \( \mathbb{R} \)

\[
E \left[ e^{jX} \right] = \int_{\mathbb{R}^n} e^{jtx} f_X(x) \, dx = e^{\frac{1}{2} x^T \Sigma^{-1} x + j \mu^T x}
\]

**Note on Non-Degeneracy (D.3.2.8):**

We say that \( X \) has a nondegenerate Gaussian distribution if the matrix \( \Sigma \) is invertible, or, when \( X \) 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 \( L^2 \), i.e. contains all its limit points. (For example, a sequence of normal R.V’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

\[
\frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \frac{1}{\sigma^2} (X - \mu)^2 \right)
\]

In particular, if \( \sigma^2 = 0 \), then a Gaussian RV has density

\[
\frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} X^2 \right)
\]

o. **(D.3.2.14) Gaussian Stochastic Process**

A stochastic process \( X_t \) is Gaussian if for all \( n < \), and all \( t_1, ..., t_n \), the random vector \( X_{t_1}, ..., X_{t_n} \) 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 \( \mu(t) = E(X_t) \) of the process and its autocovariance function \( \rho_{tt} \).

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

\[
E(e^{i\theta X_t}) = e^{i\theta \mu(t) - \frac{1}{2} \sigma^2 \theta^2}
\]

which is parameterized by the mean vector and the Var-Cov matrix. Elements of the mean vector and var-cov function are exactly what \( \mu(t) \) and \( \rho_{tt} \) are.

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

Strong

A stochastic process \( X_t \) is called (strong) stationary if its FDD's satisfy:

\[
R_{ij} = \{R_i, R_{i+1}, ..., R_{N-i}\} \quad \forall i, j
\]

Weak

A stochastic process \( X_t \) is called (weak/covariance) stationary if

\[
R_{ij} = \{R_i, R_{i+1}, ..., R_{N-i}\} \quad \forall i, j
\]

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

\[
F(Z_i, ..., Z_k) \text{ does not depend on } t \text{ for all } K \quad 0, 1, 2, ..., \quad F(Z_{i_1}, ..., Z_{k_1}) = F(Z_{i_1}, ..., Z_{k_1}) \quad \forall t, K
\]

\( \Rightarrow F(Z_i) \text{ does not depend on } t! \) (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(z) \text{ does not depend on } i \text{ and } \quad \text{Cov}(z_i, z_j) \exists, \text{ finite, and depends only on } j \text{ but not on } i \text{ (e.g. Cov}(z_i, z_j) = \text{Cov}(z_i, z_{i+j}) \)
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 PX of a Random Vector

The characteristic function determines the law of a random vector. That is, if \( X \sim Y \sim R^n \), 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. (Independence) If \( X, X_1, X_2, \ldots, X_n \) mutually independent RV’s iff

\[
E e^{i\sum_j X_j} = E e^{i\sum_j X_j} E e^{i\sum_j X_j}
\]

2. (Linearity) For \( a, b \in R \), \( aX + bX \)

3. (Gaussian)

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 Independent: If X, Y jointly Gaussian, then \( \text{Cov} X, Y = 0 \)

(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. \( X_{i=1}^n \) Gaussian is a m-dimensional Gaussian random vector, for linear combinations \( \sum_{j=1}^m a_j \).

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

3. X Gaussian has PDF:

\[
\frac{1}{\sqrt{\det(\Sigma)}} \exp \left( -\frac{1}{2} \sum_{t=1}^T - \sum_{k=1}^n \right)
\]

4. Distribution of X Gaussian has parameters and the variance-covariance matrix and the mean vector, respectively.

5. Closure: If a sequence of n-dimensional Gaussian random vectors \( X^{(k)} \) converges in L2 to a n-dimensional vector X, then, X is a Gaussian random vector whose parameters and are the limits of the corresponding parameters of the sequence of random vectors, \( X^{(k)} \) and \( X^{(k)} \) of \( X^{(k)} \).

A stochastic process \( X_t, t \in I \) is Gaussian if for all \( n < \), and all \( t_1, \ldots, \)

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

1. (t) and \( t, s \) of a Gaussian SP: All distributional properties of a Gaussian processes are determined by the mean \( \mu = \text{E}(X_t) \)

and the autocovariance function \( \text{Cov}(X_t, X_s) \) of the process and its autocovariance function \( t, s \in I \)

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

are such that for each fixed \( t, l, I \),

\( \text{E} \left[ X_t X^{(k)}_t \right] = 0 \) (each function converges pointwise on t, based on \( \| \cdot \|_2 \) metric),

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

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

\[
(i.e. \ (t) = \text{(non-random constant)} \text{ and } t, s = r, s \text{ (i.e. a function of the time difference).}
\]

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 \( L^2 \). 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. \( \omega \), the function \( t \mapsto X_t(\omega) \) is measurable from \( R \) to \( R \).

Moreover, for any interval I s.t. \( E|X_t| \, dt \), we have that \( E X_t \, dt \), \( E X_t \, dt \) (can exchange the order of integration). Further, if \( E|X_t| \) for all \( t \) in I then the function \( t \mapsto E X_t \) 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 \( F \). That is, \( F_0 \subseteq F_1 \subseteq \ldots \subseteq F_n \subseteq \ldots \subseteq F \) 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. \( X_n, n \in \{0, 1, \ldots\ \} \) is adapted to a filtration \( F_n \) if \( X_n(\omega) \) is a R.V. on \( F_n \) for each \( n \). i.e.) \( (X_n) \subseteq F_n \) for each \( n \).

Note: It can be shown that \( X_n, n \in \{0, 1, \ldots\ \} \) is adapted to a filtration \( F_n \) iff \( X_0, X_1, \ldots, X_n \subseteq F_n \) for all \( n \).

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

The filtration \( G_n = X_0, \ldots, X_n \) 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. \( A_n \) s.t. S.P. \( X_n \) is adapted to \( A_n \), \( A_n \supseteq G_n \).

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. \( E[X_n] \) S.P. adapted to this filtration s.t.

\[
E[X_{n+1} | F_n] = X_n, \text{ a.s.}
\]

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 \( H_n \subseteq F_n \) and S.P. \( X_n \) is adapted to filtration \( H_n \) then \( X_n \) is a MG.

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

Note on the Meaning of Filtration: \( E[X_{n+1} | F_n] = E[X_n | F_n] \).

(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 \( \sigma \)-Fields \( F_t \), \( F_{t+h} \) form a continuous time filtration.

That is, \( F_t \subseteq F_{t+h} \) for \( t \geq 0 \) and \( h > 0 \).

(b) The continuous time S.P. \( X_t \) is integrable and adapted to this filtration.

That is, \( E[X_t] \) and \( (X_t) \subseteq F_t \).

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

e. (D.3.1.14) L2 Martingale

A \( L^2 \)-MG (or square-integrable MG) is a MG \( X_n \) s.t. \( E[X_n^2] < \infty \) for all \( n \).

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 \( n \).
h. (D.4.1.16) Orthogonal Sequence of R.V.’s
We say that \( D_n, L^2, F, P \) is an orthogonal sequence of RV if
\[
\begin{align*}
   E[D_n h(D_0, D_1, ..., D_{n-1})] &= E[D_n] h(D_0, D_1, ..., D_{n-1}) = 2 \quad \text{for any } n \geq 1 \text{ and every Borel function } h:R^n \rightarrow R \text{ s.t. } E[h(D_0, D_1, ..., D_{n-1})] = 2
\end{align*}
\]
for any \( n \geq 1 \) and every Borel function \( h:R^n \rightarrow R \) s.t. \( E[h(D_0, D_1, ..., D_{n-1})] = 2 \)

Equivalently,
\[
E[D_n | D_0, ..., D_{n-1}] = E[D_n]
\]

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.
\[
E[X_n | F_n] = X_n \quad \text{n. a.s. i.e. } P[X_n = X_n(\omega)] = 0 \quad 0
\]
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.
\[
E[X_n | F_n] = X_n \quad \text{n. a.s. i.e. } P[X_n = X_n(\omega)] = 0 \quad 0
\]
Alternatively: \( X_n \) subMG iff \( -X_n \) superMG

Note on SubMG, SupMG, and MG: All the results about SubMG are dual staments 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 \( X_t \), if \( X_t \) is measurable on \( F_\infty \) and for each \( t > 0 \), a.s. E \( X_t \).

Similarly for supMG, but with E \( X_t \).

Note: Last element preserves sub/supMG

k. (D.4.2.9) Right-Continuous Filtration
A filtration is called right-filtration if for any \( t > 0 \),
\[
\bigcap_{h \leq t} F_h = F_t
\]

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

Note2: Note that not all filtrations are right-continuous. Consider \( F_0 = F_{\infty} = \emptyset \), \( F_{\infty} \neq \Omega \), \( F_{t} = F_{t+} = F_0 \) is not right-continuous at \( t=0 \).

l. (D.4.3.1, D.4.3.11) Stopping Time \( t \) for a filtration \( \{F_t\} \)
Discrete Time
A RV taking values in \( 0, 1, ..., n, ..., \) is a stopping time for the filtration \( F_n \) if the event \( (\omega) \cap (n) = F_n \) 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 \( (\omega) \) is called stopping time wrt the continuous filtration \( F_t \) if \( (\omega) \cap (t) = F_t \) 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 \( w:t(w)\leq n \) to check stopping time (DISCRETE TIME): Let \( (\omega) \min k \leq n : X_k(\omega) = B \) for some set \( B \) Borel
\[
(\omega) \min \bigcup_{k=1}^{n} : X_k(\omega) = B
\]
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 – If \( t \rightarrow X(t) \) is continuous a.s. and \( B \) is a closed set, then \( t(w) = \inf(t\leq 0; X(t) \in B) \) 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 $\min n, (\cdot)$, the stopped at stochastic process $X_n$ is given by
\[ X_n = X_{\min n, (\cdot)}(\cdot), \quad n (\cdot) \]

n. (D.4.4.2) Innovation Process
When using Doob's decomposition for the canonical filtration $\mathcal{X}_k,k \ n$, the MG $Y_n$ is called the innovation process associated with $X_n$.
Note: The reason for this name is that $X_n, Y_n, A_n, A_n, Y_n, Y_n, Y_n, Y_n$, where $A_n,Y_n$ is measurable on $\mathcal{X}_k,k \ n$ while $Y_n, Y_n$ describes the "NEW" part of $X_n, Y_n$.

o. (D.4.4.8) Increasing Part of a M.G. $\{M_t\}$
The S.P. $A_t$ in the Doob-Meyer decomposition of $M^2_t$ 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.
\[ Z_0 = 1 \text{ and } Z_n = N_j^{(n)} n 1 \]
where $N$ and $N_j^{(n)}$ for $j=1,2,...$ are iid non-negative integer valued R.V. with finite mean $E(N)$, and where we use the convention that if $Z_n 0$ 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 $F_n, N_j^{(k)} n,j 1,2,...$
Note that in general, $G_n, Z_k,k n F_n$
(STRICT subset, since we cannot in general recover the number of offsprings of each individual knowing only the total population sizes at different generations)

\[
\begin{align*}
Z_0 &= 1 \\
N_1^{(0)} \quad Z_1 \\
N_1^{(1)} \quad N_2^{(1)}
\end{align*}
\]

q. Probability of Extinction
\[ p_{ex} \quad P Z_n 0 \text{ for some } n 0 \]
Theorems

a. \( X_n, n \in \{0, 1, \ldots, n\} \) is adapted to a filtration \( F_n \) iff \( X_0, \ldots, X_n, F_n \) are adapted to \( F_n \).

b. (P. 4.1.7):
   If \( X_n \) is adapted to \( F_n \) for \( n \geq 0 \), 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., adapted to \( F_n \) s.t. \( E | F_n | \) is a.s. \( n \).

c. (T.4.1.12) : Martingale Transform
   Let \( X_n, F_n \) be a MG and \( V_n \) be a reasonable sequence for the same filtration. The sequence of R.V.
   \[ Y_n^{(1)} = \sum_{k=0}^{n} V_k X_k^{(1)} \]
called the martingale transform of \( V \) wrt \( X \), is then a MG with respect to the filtration \( F_n \), provided \( |V_n| \leq C_n \) for all \( n \) and some \( 1 < p < q < \infty \) 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. \( X, Z \) adapted to the filtration \( F_n \) is a MG iff \( E X_n | F_n | = X_n \) for any \( Z \) \( L^2 \)-adapted to \( F_n \).

   Note on Intuition: This result follows from the fact that \( X_n \) is a MG iff \( E[X_n | F_n] = X_n \). But the idea is that no random variables that live in the space \( L^2 \)-adapted to \( F_n \) 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 \( L^2 \)-adapted to \( F_n \) is correlated with the difference, and hence we know something about \( X_{n+1} \).

(P.4.1.17)
   A S.P. \( X, Z \) is a MG for its canonical filtration iff it has an orthogonal, zero-mean differences sequence \( D_n \), \( n \Rightarrow \) from previous.

   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 \( E D_n = 0 \) and \( E D_n D_l = 0 \) \( i \neq l \) \( n \).

   With the Gaussian vector \( D_0, \ldots, D_n \) having uncorrelated elements, we know that the var-cov matrix is diagonal and the characteristic function \( \Phi \) \( \prod_{k=0}^{n} \Phi_k \).

   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 \( E(X_n) \)
   If \( X_n \) is a subMG, then necessarily \( n \mapsto E(X_n) \) is non-decreasing. (non-increasing for supMG).

   Pf: \( E X_n \) \( E E X_n | F_n_1 \) \( E X_n | F_n_1 \) \( A \) s.t. \( P(A) = 1 \)

   So, \( E(X_n | F_n_1)(\cdot) dP(\cdot) = E(X_n | F_n_1)(\cdot) \)

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 \( \tau \) is a stopping time for \( F_n \), then \( X_n, F_n \) is also a subMG (or supMG or MG).

   Pf: \( X_n^{(\cdot)}, n \mapsto X_n^{(\cdot)} \) from previous.

   (C.4.3.7)
   If \( X_n, F_n \) is a subMG and \( \tau \) is a stopping time for \( F_n \), then \( E X_n | F_n | = X_0 \).

   If in addition \( X_n, F_n \) is a MG, then \( E X_n | = X_0 \).
Continuous Time

\( f \) 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 \( X_t \) 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 and \( B \) is a closed set, then \( \inf t \geq 0 : X_t(\omega) \in B \) is a stopping time for the canonical filtration \( G_t \).

Note on Proof: Details in pg 83.

\( : B(\omega) = \bigcup_{t \in t} : X_t(\omega) B \)

Now, each : \( X_t(\omega) \in G_t \). 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 \( \bigcup_{t \in t} : X_t(\omega) \in B \cap_k \bigcup_{t \in Q_k} : X_t(\omega) \in B_k \) for \( F_t \)

since : \( X_t(\omega) \in B_k \) and we have countable union and intersection

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

and \( B_k \) is an open set containing \( B \) (so that as \( k \to \), \( B_k \to B \)).

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) a.s., 3) \( \inf \tau \geq 0 \) is a stopping time for the filtration \( F_n \) and 4) the sequence \( X_n \) is uniformly integrable.

Then, \( E X \in X_0 \).

If in addition, 5) \( X_n, F_n \) is a MG, then \( E X \in X_0 \).

Pf: Observe that whenever \( \omega \in \omega \), then \( X_n \in X_0 \).

Since, a.s. (by assumption), then \( X_n \) converges in prob \( X_0 \) by Th.1.4.22

Thus, we have \( X_n \) U.I. and converges in prob \( X_n \to X_0 \)

\( E X_n \) by corollary 4.3.7 above, since \( X_n | F_n \) a subMG

(T.4.3.16) Continuous Time

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

3) a.s., 4) \( \inf \tau \geq 0 \) is a stopping time for the filtration \( F_t \) and 5) the sequence \( X_t \) is uniformly integrable.

Then, \( E X \in X_0 \).

If in addition, 6) \( X_t, F_t \) is a MG, then \( E X \in X_0 \).

i. (E.4.3.9) Typical Example of Application of Doob’s Optional Stopping
Consider the simple random walk \( S_n \) with \( k \) iid s.t. \( P_k = \frac{1}{2} \) so that \( E(\delta_k) = 0. \\

Let \( \inf n : S_n = \text{a time that } S_n \text{ exits the interval}. \)

We can show that...

1) \( S_n, F_n \) is a MG
2) \( S_n \) has right continuous sample paths
3) \( \tau > \) a.s.
4) \( a,b \) is a \( F_n \) Stopping Time
5) \( \max(a,b) \) so that \( S_n \) U.I.

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

\[ E S E S_0 = 0 \text{ where } S \leq a, b \text{ necessarily by construction.} \]

Thus, \( E S aP S bP S bP aP b(0 P_a) 0 P_a b a b a b \)

\[ S_i \text{ hits } -a \text{ before } b \]

\[ S_i \text{ hits } b \text{ before } -a \]

\[ \text{(E.4.3.17) A Proof for why } t < \inf \]

Let \( a,b \) \( \inf t 0: W_t < (a,b) \)

\[ P a b \lim aP a b \]

Now, \( a,b \) \( \inf t 0: W_t b \)

It's sufficient to show that \( b \) a.s.

\[ P b T P \max_0 s T W_s b \] by reflection principle \( \frac{2}{\sqrt{2}} \int_0^T \exp x^2 / 2 \ dx \tau 1 \)

Thus, \( a,b \) b a.s.

\[ \text{(T.4.4.1) Doob's Decomposition} \]

Given an integrable S.P. \( Y_n \) adapted to a discrete parameter filtration \( F_n \), \( n \) 0, there exists a decomposition

\[ X_n Y_n A_n \text{ s.t. } Y_n, F_n \text{ is a MG and } A_n \text{ is a previsible S.P. This decomposition is unique up to the value of } Y_0, \]

a R.V. measurable on \( F_0 \).
Pf: Let $A_0 = 0$, and for all $n \geq 1$, let $A_n = A_{n-1} + E X_n | F_{n-1}$. Also, since $F_k \subset F_{k+1}$ for all $k$, $A_k$ is $F_{k-1}$ measurable, by def of C.E.

Now, $A_n = A_{n-1} + A_n = \ldots = A_0 + \sum_{k=1}^{n} A_k$ and by def. of filtration, so $A_k$ is $F_{k-1}$ measurable.

Then, $A_n$ is measurable w.r.t. $F_1$ and thus a previsible process by def. To check $A_n$ is a martingale, need to show that $A_n$ is integrable, by assumption, thus it suffices to show that $A_0$ integrable. Since $X_n$ is an adapted to $F_n$, $A_n$ is integrable by Jensen's Inequality.

Now, $A_n$ is integrable. Hence, $A_n$ is integrable by the fact that $A_n$ is $F_{n-1}$ measurable and $A_n$ is previsible.

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

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

If $X_n, F_n$ a subMG, then $A_n = A_{n-1} \in$ Doob's decomposition.

$X_n, F_n$ a supMG, then $A_n = A_{n-1} \in$ Doob's decomposition.

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 $n \to \inf$ with that of the non-decreasing sequence $A_n$ in the decomposition of $\{X_n^2\}$. 

---

 discontinuous.
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. $E M_t^2 < \infty$. Then, 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 monotonicity 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, $x \in 0$ and $N$, $P \cdot \max_{0 \leq n \leq N} X_n(\omega, x) \leq \frac{E[X_N]}{x}$

(b) Suppose $X_n$ is a subMG with a last element $(X_n, F_n)$ with $F_n$ with $X$ integrable and measurable wrt $F$ and $t > 0$, $E X_t | F_t$. Then, $x \in 0$, $P \cdot \sup_{0 \leq n \leq N} X_n(\omega, x) \leq \frac{E[X_N]}{x}$

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

Then, $x \in 0$, $P \cdot \sup_{0 \leq t \leq T} X_t(\omega, x) \leq \frac{E[X_N]}{x}$

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

Pf: Consider the stopping time $x \in \min_{0 \leq n \leq N} X_n(\omega, x)$.

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

(””: Take $A = \max_{0 \leq n \leq N} X_n(\omega, x)$ at least one $n \neq N$ is s.t. $X_n(\omega, x) : x \in x N = B$)

(“”: Take $B = \min_{0 \leq n \leq N} X_n(\omega, x)$ at least one $n \neq N$ is s.t. $X_n(\omega, x) : x \in x N = B$)

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 $\sup_{0 \leq t} E X_t$ exists w.p.1, i.e. $X_t$ s.t. $X_t, a.s.$ $X$ Furthermore, in this case $E[X_t] \lim_{t \to \infty} E[X_t]$

(b) If $X_t$ is uniformly integrable, then $X_t$ also in $L^1$.

Further, $X_t \in L^1, X_t, F_t$ is a subMG with $X$ its last element i.e. for any fixed $t \in 0, X_t, E X_t | F_t$ 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 \( E X_t \) for some \( C < \) and all \( t \).

By Def 1.4.21 a necessary condition for U.I. is that \( \sup_t E|X_t| \leq \infty \) and all \( t \).

Further, by Th. 1.4.22, \( X_t \) U.I. and \( X_t \) a.s. \( X \) (and thus in prob) \( X_t \) or \( E X_t \) 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 \( X_t \) \( E X_t | F_t \) 0.

Note on Non-Converging MGs: Many important martingales do not converge. (example: Brownian Motion! It can be shown that LimSup \( W(t) = \infty \) and Lim Inf \( W(t) = -\infty \). Indeed, Doob's convergence assumptions do not hold because it can be shown that \( E[W(t)+] = \sqrt{t/2\pi} \) 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. \( EX_n \leq C \) for some \( C \) and all \( n \), then a R.V. \( X \) s.t. \( X_n \) a.s. \( X \) AND \( X_n \) L2 \( X \).

Moreover, \( EX^2 \leq C \).

Note on No L1 Analog: There is no L1 analog to this proposition!

That is, there exists a MG \( \{X_n\} \) s.t. \( EX_n < C \) for some \( C \) finite, but the L1 and a.s. limits are different!

Example: There exists a non-negative MG \( \{X_n\} \) s.t. \( EX_n = 1 \) for all \( n \) and \( X_n \) a.s. \( 0 \). So, the a.s. and L1 limits are different. (See P.4.6.5)

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

Pf: \[
E Z_{n+1} | F_n = E \sum_{j=1}^{N_j^{(n+1)}} | F_n = E N_{j}^{(n+1)} | F_n \text{ by linearity of conditional expectation and since } Z_n \text{ measurable w.r.t. } F_n \\
Z_n = E N_{j}^{(n+1)} \text{ since } N_j^{(n+1)} \text{ is independent of } F_n \text{ (the number of offsprings this gen for any ind. of the past)} \\
E N Z_n \text{ since } N_j^{(n+1)} \text{ iid with finite mean } m = E(N) < mZ_n \\
\text{Thus, } E X_{n+1} | F_n = E \frac{Z_n}{m^n+1} | F_n = \frac{1}{m^n} E Z_{n+1} | F_n = \frac{1}{m^n} m Z_n = \frac{m^n}{m^n} X_n \\
\]

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 nonnegative right-continuous MG, then \( X_t \) a.s. \( X \) for some \( X \) where \( X < \) a.s.

\( Z_n \) \( m^n X_n \) and \( m^n \) \( 0 \) (since \( m < 1 \)) \( Z_n \) a.s. 0.

But \( Z_n \) is integer valued, so \( Z_n = 0 \) 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. \( W_t, 0 \leq t \leq T \) is called a Brownian Motion (or a Wiener Process) if:

(a) \( W_t \) is a Gaussian Process

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

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

**(T.5.2.1) Levy’s Martingale Characterization of BM**

Let \( G_t \) \( X_t, 0 \leq t \leq T \). If \( X_t, G_t \) is a MG w/ continuous sample path and \( X_t^2, t, G_t \) also a MG, 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.

\[
E[W_t - W_s] = E(W_t) - E(W_s) = 0
\]

Now, \( W_t \) Gaussian all lin comb are Gaussian \( W_t, W_s \) Gaussian (s<t)

So, they're independent iff uncorrelated.

\[
\text{Cov}(W_t, W_s) = 0, \quad \text{Cov}(W_t, W_s) = \min(t,s) - \min(t,s)
\]

**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:

<table>
<thead>
<tr>
<th></th>
<th>“Nice” Calculus</th>
<th>“Stochastic” Calculus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Variation</td>
<td>(&lt;\infty)</td>
<td>(+\infty)</td>
</tr>
<tr>
<td>Quadratic Variation</td>
<td>0</td>
<td>(&lt;\infty)</td>
</tr>
<tr>
<td>(P\geq3) Variation</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**c.** (D.5.3.1) \( \| \| \) (Length of Longest Interval) and Q-th Variation of f(.) on \([a,b]\) on CALCULUS DEFINITION)

For any finite partition \( \{a = t_0 < t_1 < \cdots < t_k = b\} \), that is,

\[
|\| = \max_{i \neq j} |t_i - t_j|
\]

denote the length of the longest interval in

\[
\text{and let } V^{(q)}(f) = \frac{1}{\|}\sum_{i=1}^{k} |t_i - t_{i-1}|^q f(t_{i-1})^q
\]

denote the q-th variation of \( f(.) \) on \([a,b]\) .

The q-th variation of \( f(.) \) on \([a,b]\) is

\[
\lim_{\|\| \to 0} V^{(q)}(f)
\]

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 \( X(t) \) 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 \( \text{L}(\cdot) \) which is finite a.s. s.t. \( |X(t) - X(s)| \leq L(t - s) \) \( a.s. \) \( t,s \) \( a,b \)

**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 $L(\cdot)$ s.t. $|X(t)-X(s)| \leq L|t-s|$)

Then, $V^{(2)}(X) = \sum_{i}^{\infty} \left( X(t_{i+1}) - X(t_{i}) \right)^{2}$

So, $X(t)$ has 0 quadratic variation.

f. (D.5.3.3) Quadratic Variation

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

\[ f. \text{ (D.5.3.3) Quadratic Variation} \]

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

\[ f. \text{ (D.5.3.3) Quadratic Variation} \]
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 \( t \leq s, h > 0 \),

\[
(B_t, B_s + h) - (B_t, B_s)
\]
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 $(t,s) = \min(t,s) - \min(s,1) \min(t,1)$
   (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 $W_t | W_1 = 0$

2. (Geometric BM): $Y_t = e^{W_t}$
   (a) $Y_t$ a Gaussian S.P.
   (b) $E(Y_t) = \exp(1/2)$ and $(t,s) = \exp(\min(t,s))[(\exp(1/2)(t+s))-1]$ (USEFUL FOR REFLECTION PRINCIPLE)
   (c) Continuous sample paths a.s.
   (d) Not adapted to the canonical filtration of $W_t$ and not stationary

3. (Ornstein-Uhlenbeck): $U_t = e^{t/2 \ W^2_t}$
   (a) $U_t$ a Gaussian S.P.
   (b) $E(U_t) = 0$ and $(t,s) = \exp(1/2 - |s|)$
   (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): $X_t = x + t \ W_t$
   (a) $X_t$ a Gaussian S.P.
   (b) $E(X_t) = x + t \ E(W_t)$ and $(t,s) = \exp(\min(t,s))$
   (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) Let $X_t$, $W_t$, $W$ a BM (USEFUL FOR REFLECTION PRINCIPLE)
   $X_t \ W_t \ W$ a BM, where is a $G_t$-stopping time, $G_t$ the canonical filtration for a brownian. And $X_t$ independent of stopped field $G$

   $W_T(\ ) \colon \max_{0 \leq s \leq T} W(s) \colon (\ ) T$ where $\inf(t : W(t) )$ 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

\[
P_{\max_0 s T \ W_s} P_{T, W_T} \quad P_{T, X_T} \quad P_{T, W_T} \quad P_{T, X_T} \quad P_{T, W_T} \quad P_{\max_0 s T \ W_s} P_{T, W_T}
\]

\[
\text{true by above} \quad \text{bc if first hitting time is } T \text{ and } T \text{ W(T) still } \text{, so the difference between the first hitting time and } T \text{ must be positive}
\]

Using the fact that $X_T$ is a BM and thus for each $t$, distribution is symmetric around 0

B

\[
P_{\max_0 s T \ W_s} P_{\max_0 s T \ W_s} P_{\max_0 s T \ W_s} \quad 2P_{\max_0 s T \ W_s} \quad 2P_{W_T}
\]

by reflection principal A

SMOOTHNESS AND VARIATIONS OF THE BROWNIAN MOTION PATH

g. (P.5.3.9) Total Variation of BM $W(t)$ is infinite a.s.
Pf.: Let \( (h) = \sup_{t \in [a,b-h]} |W(t) - W(t+h)| \) (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, \( (h) \to 0 \) a.s. as \( h \to 0 \).

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

Then,

\[
(\ast) \quad \left\{ V_{(\infty)}^{(2)}(W) \right\} \stackrel{2^n}{i=0} \left( W \ a \ (i \ 1) \right) \|n\| \ W \ a \ i \|n\| \ 2^n \left( \|n\| \right) \ W \ a \ (i \ 1) \|n\| \ W \ a \ i \|n\| \\
\rightarrow \frac{2^n}{i=0} \left( \|n\| \right) \ W \ a \ (i \ 1) \|n\| \ W \ a \ i \|n\| \ \text{and} \ \left( \|n\| \right) \to 0
\]

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

Furthermore, we have \( \|n\| \) a.s. \( 0 \) from above,

So, \( V_{(\infty)}^{(2)}(W) \stackrel{2^n}{i=0} \left( W \ a \ (i \ 1) \right) \|n\| \ W \ a \ i \|n\| , \text{and} \ \left( \|n\| \right) \to 0
\]

\[
V_{(\infty)}^{(1)}(W) \quad 2^n \left( \|n\| \right) \ W \ a \ (i \ 1) \|n\| \ W \ a \ i \|n\| \ \text{a.s.}
\]

\[
V_{(\infty)}^{(1)}(W) \ = \ \text{wp1}
\]

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

<table>
<thead>
<tr>
<th>discrete values</th>
<th>continuous time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrete Time</td>
<td>Continuous Time</td>
</tr>
<tr>
<td>Continuous Values</td>
<td></td>
</tr>
</tbody>
</table>

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 \( X_n, n = 0, 1, \ldots \) with each RV \( X_n \) taking values in measurable space \( S, B \) is called a Markov Chain if

   \[
   \begin{align*}
   \text{or } & n \in \mathbb{N}, A, B, P X_n \mid A \mid X_1, \ldots, X_n, P X_n \mid A \mid X_n \quad \text{a.s.} \\
   \text{or } & n \in \mathbb{N}, A, B, E I X_n \mid A \mid X_1, \ldots, X_n, E I X_n \mid A \mid X_n \quad \text{a.s.}
   \end{align*}
   \]

   or

   \[
   (\text{or } P X_n \mid x_n \mid X_1, \ldots, X_n, x_n, P X_n \mid x_n \mid X_n, x_n \text{ if } B \text{ is the Borel sigma field, bc } B = X_n) \]

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

   **Continuous Time MC**

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

   \[
   \begin{align*}
   \text{or } & P X_t \mid (X_s, s \leq t) \quad P X_t \mid X_t \\
   \text{or } & E I X_t \mid (X_s, s \leq t) \quad E I X_t \mid X_t
   \end{align*}
   \]

   for any \( t, u \in [0, \infty) \) and any bounded measurable function \( f(.) \) on \( (S, B) \),

   \[
   E f(X_t) \mid (X_s, s \leq t) \quad E f(X_t) \mid X_t \quad \text{a.s.}
   \]

   The set \( S \) is called the state space of the Markov Process.

   **Note on Markov Process and Joint Law**: MARKOV PROPERTY 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 MARKOV.

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

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

   \[
   A, B, P X_n \mid A \mid X_n \quad \text{does not depend on } n \text{ (except via the value of } X_n\text{)}.
   \]

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

   **Discrete Time**

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

   \[
   p(A|x) \text{ is measurable on } B \text{ for any } A \in B, \text{ and } p(A|X_n) \mid P X_n \mid A \mid X_n \quad n = 0, 1, \ldots
   \]

   **Continuous Time**

   For each \( t > s \) and fixed \( x \in S \), a probability measure \( p_t(x) \) s.t. for each fixed \( A \in B \),

   \[
   \text{the function } p_t(x)(A|.) \text{ is measurable and } p_t(x)(A|X_s) \mid E I X_t \mid A \mid X_t, P X_t \mid A \mid X_t
   \]

   Such a collection \( p_t(x)(A|.) \) is called the transition probabilities for the Markov process \( X_t \).

   **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 of $X_0$. That is, it's the probability measure $A \ P X_0 \ A$ 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 and the transition probability function $p(A | x), A B, x S$.

If discrete state space and time homogenous, we can find the distribution of $X_n$ given $A$ is: $p^n$

**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 nonegative integer $k$, every $0=t_0<t_1<...<t_k$ and $A_0,...,A_k$ we have that

\[
P X(t_k) \ A_k, ..., X(t_0) \ A_0 = \int_{A_k, ..., A_0} p_{t_k, t_{k-1}} (dx_k | x_{k-1}) ... p_{t_1, t_0} (dx_1 | x_0) (dx_0) \quad \text{(using Lebesgue Integral)}
\]

**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 $p_t(\cdot | \cdot)$ 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 may often express $p_t(\cdot | \cdot)$ 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 reason that 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 $N_t(\cdot)$ is piecewise constant, nondecreasing, and right-continuous

b) $N_0(\cdot) = 0$

c) All jump discontinuities are of size 1

**C1 (Independent and Stationary Poisson Increments):**

For any $k$ and any $0 = t_1 < t_2 < ... < t_k$, the increments $N_{t_1}, N_{t_2}, ..., N_{t_k}$ are independent R.V.'s.

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

(i.e. stationary and independent poisson distributed increments)

**C2 (Independent and Stationary Increments):**

For any $k$ and any $0 = t_1 < t_2 < ... < t_k$, the increments $N_{t_1}, N_{t_2}, ..., N_{t_k}$ are independent R.V.'s.

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

(i.e. stationary increments)

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

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

**C4**

a) The S.P. $N_t$ has no fixed discontinuities, i.e. $P(T_k = t) = 0$ for $k$ and $t > 0$.

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

\[
P(N_{t_k} = n_k \mid N_{t_1} = n_1, j \mid k) = o(h).
\]

\[
P(N_{t_k} = n_k \mid N_{t_1} = n_1, j \mid k) = o(h).
\]

where $o(h)$ denotes a function $f(h)$ s.t. $f(h) \rightarrow 0$ as $h \rightarrow 0$. 

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 jump discontinuities are of size 1

or equivalently, \( N_t \sup_{k} k : T_k \leq t \) 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 \( t \geq 0 \), 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 \( 0 = T_0 < T_1 < ... \) such that

\[ T_k = \inf t : N_t \geq k \quad \forall k \]

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

h. Poisson R.V.

A R.V. \( N \) has the Poisson(\( \lambda \)) law if

\[ P( N = k ) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k \text{ integer} \]

Has \( \text{E}(N) = \lambda \) and \( \text{Var}(N) = \lambda \)

i. S.P. with Independent Increments

We say that a S.P. \( N_t \) has independent increments if

\[ N_{t+h} - N_t \sim \text{Poisson}(h \lambda) \]

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

We say that the S.P. \( N_t \) 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(\( \lambda \)) law if \( P(T \geq t) = e^{-\lambda t} \) for all \( t \geq 0 \) and some \( \lambda > 0 \).

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

\[ P(T \geq x | T \geq y) = P(T \geq x) \]

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 \( t_1 < t_2 < ... < t_k \),

a) the increments, \( N_{t_1} - N_{t_0}, N_{t_2} - N_{t_1}, ..., N_{t_k} - N_{t_{k-1}} \) are independent R.V.’s

b) for some \( \lambda > 0 \) (rate of intensity) and all \( t \geq 0 \), the increment \( N_t \sim \text{Poisson}(\lambda t) \)

(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 \( \lambda \), 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 \( N(t) - N(0,t) \) and \( N(t) - N(s) \sim \text{Poisson}(\lambda(t-s)) \).

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 \( \lambda \) iff it is a counting process (C0) with Jump Time Increments~Exp(\( \lambda \)) (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 \( \lambda \) iff it is a counting process (C0) that satisfies C4.
Theorems

a. Showing that a SP is a Markov SP

1. Compute \( P \left( X_{t+h} \mid A \right) F_t \) or \( P \left( X_{n+1} \mid A \right) F_n \) 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 \( X(t) = f[Y(g(t))] \) 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 \( F_{t,n} \) be the canonical filtration of \( X_t, t \geq 0 \).

Independent increments means that for any \( t,h \geq 0 \), the random variable \( X_{t+h} - X_h \) is \( F_t \).

To show that \( X_t \) is a Markov process, from Denition 6.1.7 it sufﬁces to show that

\[
E[f(X_{t+h}) \mid F_t] = E[f(X_{t+h} \mid X_t)],
\]

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,

\[
E[f(X_{t+h}) \mid F_t] = E[f(X_{t+h} \mid X_t)],
\]

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 \( X_n \).

Thus,

\[
P(X_{n+1} \mid x, X_n) = \frac{P(X_{n+1} \mid y, X_n)}{P(X_n \mid x)} = \frac{P(X_{n+1} \mid Y_{n+1}, X_n)}{P(X_n \mid x)} = \frac{P(X_{n+1} \mid Y_{n+1}, X_n)}{P(X_n \mid x)},
\]

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 Æ 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 \( Y(t) = f(X_{g(t)}) \) is also a Markov Process for \( f \) invertible and \( g \) strictly increasing.

Theorem: If \( X(t) \) a Markov Process, then \( Y(t) = f(X_{g(t)}) \) also a Markov Process for \( f \) invertible and \( g \) strictly increasing.

If further that \( X(t) \) is homogeneous and \( f \) not time dependent, then \( f(X_{g(t)}) \) also homogeneous.
Pf.
Take $A \subseteq F_t$ and $y_s \in \mathbb{R}$.

Note that since $f$ invertible, $y_s$ is Markovian.

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

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($\lambda$) law if $P(T > t) = e^{-\lambda t}$ for all $t \geq 0$ and some $\lambda > 0$.

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

$$P(T > x, y | T > y) = P(T > x)$$

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

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

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

$$P(T_k \leq t_k \leq \ldots \leq t_n \mid N_t = n) = \prod_{i=1}^{n} P(U_i \leq \frac{t_i - t_{i-1}}{n}) = \frac{1}{n!} \int_0^t \ldots \int_0^{t_n} dx_1 dx_2 \ldots dx_n$$

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, $E \left[ T_1 \mid N_t = n \right]$ by theorem $= E \left[ n \sum_{i=1}^{n} U_i \right]$ since it doesn't matter how we sum

$$= \sum_{i=1}^{n} E[U_i] = \frac{n t}{2}$$

l. (T.6.2.10) Poisson Approximation

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

$$P(Z^{(n)}_l = 0) > 0 \text{ and } P(Z^{(n)}_l > 0) > 0 \text{ and } (iii) \max_{l} P_l^{(n)} = 0$$

Then, $S_n \overset{D}{\to} \text{Poisson}$ when $n$
Take $Z_i^{(n)} \sim \frac{1}{n}$ w.p. $\frac{1}{n}$

So, $Z_i^{(n)} \sim \text{Bernoulli} \left( \frac{1}{n} \right)$, $S_n \sim \text{Binomial} \left( n, \frac{1}{n} \right)$ (since sum of iid bernoulli)

Here, $\frac{1}{n} p_{i}^{(n)} \sim \frac{1}{n}$ and $P \left( Z_i^{(n)} \right) \sim \frac{2}{n}$ 0 so $\frac{1}{n} p_{i}^{(n)} \sim \frac{1}{n}$ 0 and $\max_{1 \leq i \leq n} \frac{1}{n} p_{i}^{(n)} \sim \frac{1}{n}$ 0

Thus, by theorem above, we have $S_n \sim \sqrt[n]{\max_{1 \leq i \leq n} Z_i^{(n)} \sim \text{Poisson}}$ when $n$

**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 $N_1^{(1)}$ and $N_2^{(2)}$ are two independent Poisson processes of rates $\lambda_1$ and $\lambda_2$ respectively, then $N_1^{(1)} + N_2^{(2)}$ 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 $\lambda$ corresponds to a Poisson process of rate $p\lambda$. 