UC Berkeley Qualifying Exam

Anya Michaelsen, October 2021

Probability Theory Study Guide

Major topic: Probability Theory (Probability)

References: Durrett, *Probability: Theory and Examples 5th Ed.*, Chapters 1–5

- **Preliminaries:** σ-algebras, Dynkin's π-λ theorem, independence, Borel–Cantelli lemmas, Kolmogorov's 0-1 law, Kolmogorov's maximal inequality, strong and weak laws of large numbers
- **Central limit theorems:** weak convergence, characteristic functions, tightness, I.I.D. central limit theorem, Lindeberg–Feller central limit theorem
- **Conditioning:** conditional probability and expectation, regular conditional probabilities
- **Martinagles:** stopping times, upcrossing inequality, uniform integrability, A.S. convergence, Doob's decomposition, Doob's inequality, L^p convergence, L^1 convergence, reverse martingale convergence, optional stopping theorem, Wald's identity
- **Markov chains:** countable state space, stationary measures, convergence theorems, recurrence and transience, asymptotic behavior

Additional References:

- *Probability with Martingales* by David Williams

Contents

Memorization (– key terms –)

Chapter 1 - Preliminaries

1.1 Probability Spaces

1 probability space, measure space

probability space: $(\Omega, \mathcal{F}, P)$ - Ω outcomes, $\mathcal{F}$ events, and $P : \mathcal{F} \to [0, 1]$ assigns probabilities to events
measure space: $(\Omega, \mathcal{F})$ - Ω outcomes, $\mathcal{F}$ events

2 σ-field/algebra, σ-field generated by $\mathcal{A}$

σ-field: $\mathcal{F}$ a non-empty collection of subsets of Ω satisfying:

i $A \in \mathcal{F} \implies A^C \in \mathcal{F}$

ii $A_i \in \mathcal{F}$ countable sequence, then $\cup_i A_i \in \mathcal{F}$

σ-field generated by $\mathcal{A}$: smallest σ-field containing the collection $\mathcal{A}$, denoted $\sigma(\mathcal{A})$

3 measure, probability measure

measure: "non-negative countable additive set function", i.e. $\mu : \mathcal{F} \to \mathbb{R}$ such that

i $\mu(A) \geq \mu(\emptyset)$ for all $A \in \mathcal{F}$

ii $A_i \in \mathcal{F}$ countable sequence of *disjoint* sets, then $\mu(\cup_i A_i) = \sum_i \mu(A_i)$

probability measure: $\mu(\Omega) = 1$, usually denoted P

4 monotonicity, subadditivity

μ a measure on $(\Omega, \mathcal{F})$
monotonicity: $A \subseteq B \implies \mu(A) \leq \mu(B)$
subadditivity: $A \subset \cup_i A_i \implies \mu(A) \leq \sum_i \mu(A_i)$

5 continuity from below/above

μ a measure on $(\Omega, \mathcal{F})$
if $A_i \uparrow A$ ($A_1 \subset A_2 \subset \cdots$ and $\cup_i A_i = A$) then $\mu(A_i) \uparrow \mu(A)$
if $A_i \downarrow A$ ($A_1 \supset A_2 \supset \cdots$ and $\cap_i A_i = A$) then $\mu(A_i) \downarrow \mu(A)$

6 discrete probability spaces

Ω a countable set, $\mathcal{F}$ all subsets of Ω

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

where $p(\omega) \geq 0$ and $\sum_{\omega \in \Omega} p(\omega) = 1$ [i.e. each ω gets assigned its own point probability and sets are simply sums of the point probabilities]
Discrete *uniform* probability - Ω finite and $p(\omega) = 1/|\Omega|$ for all $\omega \in \Omega$.

7 Borel sets

the smallest σ-algebra containing the open sets in $\mathbb{R}^d$ (with the usual Euclidean topology)

8 Stieltjes measure function

A function $F : \mathbb{R} \to \mathbb{R}$ such that F is (i) nondecreasing and (ii) right continuous ($\lim_{y \downarrow x} F(y) = F(x)$)

9 Lebesgue measures on $\mathbb{R}$ and $\mathbb{R}^d$

$\mathbb{R}$: The unique measure on $(\mathbb{R}, \mathcal{R})$ such that $\mu((a, b]) = b - a$.

$\mathbb{R}^d$: The unique measure on $(\mathbb{R}, \mathcal{R})$ such that $\mu(A) =$ area of A for all finite rectangles A.

10 semi-algebra, algebra (field), algebra generated by $\mathcal{S}$

semi-algebra: $\mathcal{S}$ such that (i) closed under finite intersection, (ii) $S \in \mathcal{S}$ implies S^C is a *finite disjoint union of sets* in $\mathcal{S}$

algebra: $\mathcal{A}$ such that (i) closed under finite intersections, (ii) closed under complements (it follows closed under finite unions)

algebra generated by $\mathcal{S}$: $\overline{\mathcal{S}}$, collection of finite disjoint unions of sets in $\mathcal{S}$ (is an algebra)

11 measure on an algebra

given algebra $\mathcal{A}$ a measure on $\mathcal{A}$, μ is a set function $\mu : \mathcal{A} \to \mathbb{R}$ such that

(i) $\mu(A) \geq \mu(\emptyset) = 0$ for all $A \in \mathcal{A}$ and

(ii) $A_i \in \mathcal{A}$ are disjoint *and their union is in* $\mathcal{A}$, then $\mu(\cup_i A_i) = \sum_i \mu(A_i)$.

12 σ-finite

a measure μ on an algebra $\mathcal{A}$ is σ-finite if there is a sequence of sets $A_n \in \mathcal{A}$ such that $\mu(A_n) < \infty$ for all n and $\cup_n A_n = \Omega$ (could also assume that $A_n \uparrow \Omega$ or the A_n are disjoint)

13 countably generated σ-field/algebra

$\mathcal{F}$, a σ-field is countably generated if there is a countable collection $\mathcal{C} \subset \mathcal{F}$ such that $\sigma(\mathcal{C}) = \mathcal{F}$

1.2 Distributions

14 random variable ($\mathcal{F}$-measurable)

a real valued function $X : \Omega \to \mathbb{R}$ such that for every Borel set $B \subset \mathbb{R}$, $X^{-1}(B) \in \mathcal{F}$, the specific σ-field on Ω (if specification needed, X is $\mathcal{F}$-measurable)

15 indicator function of a set

example of a random variable where $A \in \mathcal{F}$

$$1_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

16 distribution (function) of a random variable

When X is a random variable on a probability space $(\Omega, \mathcal{F}, P)$ then its distribution is a probability measure, μ, on $\mathbb{R}$ given by

$$\mu(A) = P(X \in A) = P(X^{-1}(A))$$

the associated distribution function is given by $F(x) = P(X \leq x) = P(X^{-1}((-\infty, x]))$

17 equal in distribution

two random variables whose resulting distributions (measures) on $\mathbb{R}$ are the same, this occurs exactly when they have the same distribution function also, denoted by $X \overset{d}{=} Y$

18 density functions

a function f such that the distribution function $F(x) = P(X \leq x)$ satisfies $F(x) = \int_{-\infty}^{x} f(y)dy$.

19 uniform distribution on $(0,1)$

density function $f(x) = 1$ where $x \in (0,1)$ (0 everywhere else)

$$F(x) = \begin{cases} 0 & x \leq 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

20 absolutely continuous distribution function

a distribution function on $\mathbb{R}$ is absolutely continuous if it has a density function

21 discrete probability measure/distribution function

probability measure P with a countable set S such that $P(S^C) = 0$ (example point mass below)

22 point mass distribution function

$F(x) = 1$ for $x \geq 0$ (or another point of your choosing) and $F(x) = 0$ for $x < 0$.
this is a discrete probability measure realized by the set $S = \{0\}$

1.3 Random Variables

23 measurable map

a function $X : \Omega \to S$ between measurable spaces $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$ ($\mathcal{F}$ and $\mathcal{S}$ σ-fields) such that for all $B \in \mathcal{S}$

$$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{F}$$

24 random variable, random vector

random variable: a measurable function $\Omega \to (\mathbb{R}, \mathcal{R})$
random vector: a measurable function $\Omega \to (\mathbb{R}^d, \mathcal{R}^d)$, $d > 1$

25 σ-field generated by a measurable map

Given $X : \Omega \to S$ with $(S, \mathcal{S})$,

$$\sigma(X) = \{\{X \in B\} : B \in \mathcal{S}\} = \{\{\omega : X(\omega) \in B\} : B \in \mathcal{S}\}$$

26 properties of combining measurable maps

composition of measurable maps is measurable
summation of a finite number of measurable maps is measurable
$X_1, \ldots, X_n$ random variables, and $f : \mathbb{R}^d \to \mathbb{R}$ measurable, then $f(X_1, \ldots, X_n)$ measurable
$\inf, \sup \limsup, \liminf$ of sequences of random variables are random variables

27 random variables converging almost surely

Given random variables $X_i : \Omega \to \mathbb{R}$,

$$\Omega_0 = \{\omega : \lim_n X_n(\omega) \text{ exists } \}$$

the X_i's converge almost surely (or 'almost everywhere') when $P(\Omega_0) = 1$.

28 extended real line

$\mathbb{R}^* = [-\infty, \infty]$ with Borel sets generated by $[-\infty, a), (a, b), (b, \infty]$

1.4 Integration

29 simple function

$\varphi = \sum_{i=0}^{n} a_i 1_{A_i}$ where A_i are disjoint sets with $\mu(A_i) < \infty$

30 integration of simple functions

$$\int \varphi d\mu = \int \sum_{i=0}^{n} a_i 1_{A_i} d\mu = \sum_{i=0}^{n} a_i \mu(A_i)$$

31 $\phi \geq \psi$ almost everywhere

$\phi \geq \psi$ almost everywhere $\iff \mu(\{\omega : \phi(\omega) < \psi(\omega)\}) = 0$

32 integral of bounded functions

$$\int f d\mu = \sup_{\varphi \leq f} \int \varphi d\mu = \inf_{\psi \geq f} \int \psi d\mu$$

33 integral of non-negative functions

$$\int f d\mu = \sup_{0 \leq h \leq f} \left\{ \int h d\mu : h \text{ bounded}, \mu(\{x : h(x) > 0\}) < \infty \right\}$$

34 integrable functions

$\int |f| d\mu < \infty$ (since $|f|$ is non-negative function)

35 integral of (integrable) functions

$f^+ = \max(f, 0)$ and $f^- = \max(-f, 0)$
so that $f = f^+ - f^-$ and f^+, f^- are both non-negative functions
$\int f d\mu = \int f^+ d\mu - \int f^- d\mu$

36 basic properties of integrals

If f, g are both integrable/non-negative/bounded/simple:

(i) If $f \geq 0$ a.e. then $\int f d\mu \geq 0$

(ii) For all $a \in \mathbb{R}$, $\int a f d\mu = a \int f d\mu$

(iii) $\int f + g d\mu = \int f d\mu + \int g d\mu$

(iv) If $g \leq f$ a.e. $\int g d\mu \leq \int f d\mu$

(v) If $g = f$ a.e. $\int g d\mu = \int f d\mu$

(vi) $\left| \int f d\mu \right| \leq \int |f| d\mu$

1.5 Properties of the Integral

37 Jensen's Inequality (integral)

If φ is convex (technical: $\lambda\varphi(x) + (1-\lambda)\varphi(y) \geq \varphi(\lambda x + (1-\lambda)y)$, $\lambda \in (0,1)$)
μ probability measure, and f and $\varphi(f)$ integrable

$$\varphi\left(\int f d\mu\right) \leq \int \varphi(f) d\mu$$

38 $||f||_p$

$$||f||_p = \left(\int |f|^p d\mu\right)^{1/p}$$

for $1 \leq p < \infty$

39 Hölder's Inequality (integral)

If $p, q \in (1, \infty)$ and $\frac{1}{p} + \frac{1}{q} = 1$ then

$$\int |fg| d\mu \leq ||f||_p ||g||_q$$

40 Cauchy-Schwarz Inequality

$$\int |fg| d\mu \leq ||f||_2 ||g||_2 = \sqrt{\int f^2 d\mu}\sqrt{\int g^2 d\mu}$$

41 Bounded Convergence Theorem

Finite measure set E ('bounded': $\mu(E) < \infty$) and f_n supported on E (vanishes on E^C)
f_n bounded (i.e. $|f_n| \leq M$ for some M)
$f_n \to f$ in measure (measure zero set in the limit $|f_n(\omega) - f(\omega)| > \varepsilon$)

$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu$$

42 Fatou's Lemma

$$f_n \geq 0 \qquad \Longrightarrow \qquad \liminf_{n\to\infty} \int f_n d\mu \geq \int \left(\liminf_{n\to\infty} f_n\right) d\mu$$

43 Monotone Convergence Theorem

$$f_n \geq 0 \text{ and } f_n \uparrow f \qquad \Longrightarrow \qquad \int f_n d\mu \uparrow \int f d\mu$$

44 Dominated Convergence Theorem

If $f_n \to f$ a.e., $|f_n| \leq g$ for all n where g is integrable then

$$\int f_n d\mu \to \int f d\mu$$

1.6 Expected Value

45 expected Value (mean) of R.V. (and basic properties)

$X \geq 0$ a random variable on $(\Omega, \mathcal{F}, P)$ then

$$EX = \int X dP$$

(may be infinite)

By integral properties, $E(X+Y) = EX+EY$, $E(aX+b) = aE(X)+b$ and $X \geq Y$ implies $EX \geq EY$.

46 Jensen's Inequality for EX

If φ is convex, then $E(\varphi(X)) \geq \varphi(E(X))$ where both exist

47 Hölder's Inequality for EX

$p, q \in [1, \infty]$ with $1/p + 1/q = 1$

$$E|XY| \leq ||X||_p ||Y||_q$$

where $||X||_p = (E|X|^p)^{1/p}$ for $p < \infty$ and $||X||_\infty = \inf\{M : P(|X| > M) = 0\}$

48 Chebyshev's Inequality

$\varphi : \mathbb{R} \to \mathbb{R}$ with $\varphi \geq 0$. For Borel set $A \in \mathcal{R}$ let $i_A = \inf\{\varphi(y) : y \in A\}$ ('min' value of φ on A)

$$i_A P(X \in A) \leq E(\varphi(X); X \in A) \leq E(\varphi(X))$$

A common version:

$$P(X \geq y) f(y) \leq Ef(x)$$

for example when $EX = 0$

$$a^2 P(|X| \geq a) \leq EX^2 \implies P(|X| \geq a) \leq a^{-2}\operatorname{var}(X)$$

49 Fatou's Lemma for EX

$$\liminf_{n\to\infty} EX_n \geq E(\liminf_{n\to\infty} X_n)$$

50 Monotone convergence theorem for EX

$X_n \geq 0$ and $X_n \uparrow X$ then $EX_n \uparrow EX$

51 dominated convergence theorem for EX

$X_n \to X$ and $|X_n| \leq Y$ for all n and $EY < \infty$ then $EX_n \to EX$

52 bounded convergence theorem for EX

$X_n \to X$ and $|X_n| \leq M$ for all n then $EX_n \to EX$

53 mean and variance of a R.V.

mean is just expected value, $\mu = EX$

if EX^2 exists, then $\operatorname{var}(X) = E(X - \mu)^2$

54 kth moment of X

$E(X^k)$

55 computing EX integrals (change of variable formula)

Measure space $(S, \mathcal{S}, P)$ X a random element (variable?) on $(S, \mathcal{S})$ with $\mu(A) = P(X \in A) = P(X^{-1}(A))$

If f measurable function $(S, \mathcal{S}) \to (\mathbb{R}, \mathcal{R})$ with $f \geq 0$ or $E|f(X)| < \infty$ then

$$Ef(x) = \int_S f(y)\mu(dy) = \int f d\mu$$

If X has density function $F(x) = \int_{-\infty}^{x} g(x)dx$ then

$$Ef(X) = \int_{-\infty}^{\infty} f(x)g(x)dx$$

56 Bernoulli Distribution

this is a discrete distribution

p some parameter, $P(X = 1) = p$ and $P(X = 0) = 1 - p$

57 Poisson Distribution

this is a discrete distribution

λ some parameter, $P(X = k) = e^{-\lambda}\lambda^k/k!$ for $k = 0, 1, 2, \dots$

58 Other formulas for Expected Values

If $X \geq 0$ $EX = \sum_{i=0}^{\infty} P(X \geq i)$ or $EX = \int_0^{\infty} P(X \geq x)dx$ (useful when nice formula for $P(X \geq x)$) so for example $E|X| = \int_0^{\infty} P(|X| > x)dx$ (can be derived by Fubini's Theorem)

1.7 Product Measures, Fubini's Theorem

59 product measure

Take $(X, \mathbb{A}, \mu_1)$ and $(Y, \mathcal{B}, \mu_2)$ with σ-finite measures, then $\mu = \mu_1 \times \mu_2$ is the unique measure on $X \times Y$ such that $\mu(A \times B) = \mu_1(A)\mu_2(B)$

60 Fubini's Theorem

Fubini's gives conditions for switching multiple integrals using product spaces

Fubini's Theorem: *Let μ_1, μ_2 be σ-finite with $\mu = \mu_1 \times \mu_2$. If $f \geq 0$ or $\int |f| d\mu < \infty$ then*

$$\int_X \int_Y f d\mu_2 d\mu_1 = \int_{X \times Y} f d\mu = \int_Y \int_X f d\mu_1 d\mu_2$$

Typical application to summation/sum+integral combinations

Chapter 2 - Law of Large Numbers

2.1 Independence

61 independence of σ-fields, random variables

independence for σ-fields: (finite version)

$\mathcal{F}_1, \dots, \mathcal{F}_n$ σ-fields (all contained in some larger σ-field with P probability measure) are **independent** if for any choice of $A_i \in \mathcal{F}_i$ for all $i = 1, \dots, n$

$$P(\cap_i A_i) = \prod_i P(A_i)$$

(infinite collections are independent if all finite sub-collections are independent)

Independence for random variables:
$X_1, \dots, X_n$ random variables from $(\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{R})$ are independent if $\sigma(X_i)$'s are all independent which is equivalent to when

$$P(X_1 \in C_1, \dots, X_n \in C_n) = P(\cap_i \{X_i \in C_i\}) = \prod_i P(X_i \in C_i)$$

for any collection of $C_i \in \mathcal{R}$

62 Independence of events and arbitrary collections of events

Most simply, $P(A \cap B) = P(A)P(B)$

Independence of Events:
Generally, $A_1, \dots, A_n$ are independent events in $(\Omega, \mathcal{F}, P)$ if for any *sub-collection* of sets (i.e. $I \subseteq \{1, 2, \dots, n\}$

$$P(\cap_{i \in I} A_i = \prod_{i \in I} P(A_i)$$

Independence of Collections of Sets: Given $\mathcal{A}_1, \dots, \mathcal{A}_n$ collections of sets, these are independent if for any choice of $A_i \in \mathcal{A}_i$ of a subcollection (i.e. $I \subseteq \{1, \dots, n\}$) we have A_is are independent. Can always assume $\Omega \in \mathcal{A}_i$ and take the full collection every time.

63 Pairwise Independent, how it differs from Independent

Any *pairs* are independent ($P(A \cap B) = P(A)P(B)$)

Independent $\implies$ Pairwise Independent but pairwise is *strictly weaker*

Example:
X_1, X_2, X_3 with $P(X_i = 0) = P(X_i = 1) = 1/2$
$A_1 = \{X_2 = X_3\}$, $A_2 = \{X_1 = X_3\}$, and $A_3 = \{X_1 = X_2\}$

$A_1 \cap A_2 = A_1 \cap A_2 \cap A_3$ so the probabilities are the same, but $P(A_i) = 1/2$ so adding A_3 changes the RHS of the independence equation.

64 π-system, λ-system, relationship to σ-fields

π-system - closed under intersection

λ-system - $\Omega \in \mathcal{L}$, countable unions of increasing sets contained, and set subtraction contained ($A \subseteq B \Rightarrow B \cap A^C \in \mathcal{L}$)

$\pi - \lambda$ systems $\iff$ σ-algebra, so π,λ kinda split up the σ algebra properties

65 Dynkin's π-λ Theorem

π-λ Theorem: *If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system with $\mathcal{P} \subseteq \mathcal{L}$ then $\sigma(\mathcal{P}) \subset \mathcal{L}$.*

π-system - closed under intersection

λ-system - $\Omega \in \mathcal{L}$, countable unions contained, and set subtraction contained ($A \subseteq B \Rightarrow B \cap A^C \in \mathcal{L}$)

66 distribution of collections of independent variables

$Y = (X_1, \dots, X_n)$ for independent random variables X_i each with distribution $\mu_i(A_i) = P(X_i \in A_i)$ (so Y a random vector) then the distribution measure for Y is $\mu = \mu_1 \times \cdots \mu_n$ where $\mu(A_1 \times A_n) = \prod_i \mu_i(A_i)$.

67 expected value of product of independent variables

If $X_1, \dots, X_n$ are independent variables and either all $X_i \geq 0$ OR all $E|X_i| < \infty$ then

$$E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E(X_i)$$

68 convolution (of distribution functions)

F and G distribution functions

$$(F * G)(z) = \int F(z-y)dG(y) = \int F(z-y)d\mu$$

69 distributions of sums of variables

If X, Y are independent, with distribution functions $F(x) = P(X \leq x)$ and $G(y) = P(Y \leq y)$ (and $\mu(A) = P(Y \in A)$ distribution measure for Y),

$$P(X + Y \leq w) = \int F(w-y)dG(y) = \int F(w-y)d\mu$$

70 density function for sums of variables

Given X, Y independent R.V.s with density functions f, g respectively, the density function for $X+Y$ is $\boxed{\int_{\mathbb{R}} f(x-y)g(y)dy}$

71 Gamma density (parameters α, λ)

Density function $\boxed{f(x) = \lambda^{\alpha} x^{\alpha-1} e^{-\lambda x}/\Gamma(\alpha)}$ for $x \geq 0$ (and $f = 0$ when $x < 0$) where $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1}e^{-x}$, related to factorials.

72 sums of exponential distributions

If $X_1, \dots, X_n$ are all independent exponential distributions with λ, then $X_1 + \cdots + X_n$ is Gamma distribution with $\alpha = n$ and $\lambda = \lambda$

73 sums of normal distributions

$N(\mu, a) + N(\nu, b) = N(\mu + \nu, a + b)$ (assuming independence!)

2.2 Weak Laws of Large Numbers

74 convergences almost surely (almost everywhere)

$\mu(\{x : f_n(x) \neq f(x)\}) \to 0$

75 L^p convergence

$E|f_n - f|^p \to 0$

76 convergence in probability

for all $\varepsilon > 0$, $P(\{x : |f_n(x) - f(x)| > \varepsilon\}) \to 0$ (equivalent to $\geq \varepsilon$)

77 i.i.d variables

independent and identically distributed variables

same probabilities/expected values/variances etc *and* independent

example: repeated coin tosses, all have the same probabilities but do not affect each other

78 Weak Law of Large Numbers

Weak Law of Large Numbers: *Let $X_1, X_2, \ldots$ be i.i.d with finite variance (can weaken to $E|X_i| < \infty$). Let $S_n = X_1 + X_2 + \cdots + X_n$ and $\mu = EX_1$. Then* $\boxed{S_n/n \to \mu \text{ converges in probability}}$.

2.3 Borel-Cantelli Lemmas

79 Borel-Cantelli Lemma

Borel-Cantelli Lemma

$$\boxed{\sum_{n=1}^{\infty} P(A_n) < \infty \quad \Rightarrow \quad P(A_n \text{ i.o.}) = 0} = P(\limsup_{n\to\infty} A_n) = P(\lim_{n\to\infty} \cup_{m=n}^{\infty} A_m)$$

80 Second Borel-Cantelli Lemma

Second Borel-Cantelli Lemma: *If A_n are independent and $\sum P(A_n) = \infty$ then*

$$P(\{x : x \in A_n \text{ infinitely often}\}) = P(\limsup_{n\to\infty} A_n) = P(\lim_{n\to\infty} \cup_{m=n}^{\infty} A_m) = \boxed{P(A_n \text{ infinitely often}) = 1}$$

2.4 Strong Law of Large Numbers

81 Strong Law of Large Numbers

SLLN: *Let $X_1, X_2, \ldots$ be i.i.d. with $EX_i = \mu$ and $EX_i^4 < \infty$. (can weaken to $E|X_i| < \infty$) Then*

$$\frac{S_n}{n} = \frac{X_1 + \cdots X_n}{n} \xrightarrow{a.s.} \mu$$

Generalizations:
Let $X_1, X_2, \ldots$ be pairwise independent and identically distributed with $EX_i = \mu$ and $E|X_i| < \infty$. Then

$$\frac{S_n}{n} = \frac{X_1 + \cdots X_n}{n} \xrightarrow{a.s.} \mu$$

Let $X_1, X_2, \ldots$ be i.i.d with $EX_i^+ = \infty$ and $EX_i^- < \infty$ (hence $EX_i = \infty$). Then

$$\frac{S_n}{n} = \frac{X_1 + \cdots X_n}{n} \xrightarrow{a.s.} EX_i = \infty$$

2.5 Convergence of Random Series

82 tail σ-field

$\mathcal{T}$ only depends on tail behavior, i.e. changing finitely many values does not affect it

Formally, $\mathcal{F}_n = \sigma(X_n, X_{n+1}, \ldots)$ and $\mathcal{T} = \cap_n \mathcal{F}_n$

Examples: $\{\lim_{n\to\infty} S_n \text{ exists}\} \in \mathcal{T}$, $B_n \in \mathcal{R}$ then $\{X_n \in B_n \text{ i.o.}\} \in \mathcal{T}$

83 Kolmogorov's 0-1 Law

$X_1, X_2, \ldots$ are independent, then $A \in \mathcal{T}$ implies $P(A) \in \{0, 1\}$ (almost always or almost never)

84 exchangeable σ-field and Hewitt-Savage 0-1 Law

exchangeable σ-field invariant sets under finite permutation of values/variables (contains tail σ-field)
Hewitt-Savage 0-1 Law: If $X_1, X_2, \ldots$ i.i.d then $A \in \mathcal{E}$ implies $P(A) \in \{0, 1\}$

85 Kolmogorov's Maximal Inequality

$X_1, X_2, \ldots, X_n$ independent with $EX_i = 0$ and $\text{var}(X_i) < \infty$, $S_n = X_1 + \cdots X_n$ as usual

$$P\left(\max_{1 \le k \le n} |S_k| \ge x\right) \le x^{-2} \text{var}(S_n)$$

86 General Idea of Kolmogorov's Three Series Theorem

gives (3) equivalent conditions on X_i and their truncations to show that the series converges a.s.

Chapter 3 - Central Limit Theorems

3.1 The De Moivre-Laplace Theorem

87 Stirling's Formula

$n! \sim n^n e^{-n} \sqrt{2\pi n}$

3.2 Weak Convergence

88 weak convergence

$F_n \Rightarrow F_\infty$ when $\lim_n F_n(y) = F_\infty(y)$ for all points y where F_∞ is continuous.
Equivalently, $X_n \Rightarrow X_\infty$ or $\mu_n \Rightarrow \mu_\infty$ where μ_i is the probability measure for X_i.

89 (equivalent) properties of weak convergence

$X_n \Rightarrow X_\infty$ if and only if $Eg(X_n) \to Eg(X_\infty)$ for all bounded continuous functions g.
$X_n \Rightarrow X_\infty$ if and only if there exists Y_n with the same distribution such that $Y_n \to Y_\infty$ a.s.
Also some equivalent conditions in terms of open/closed/Borel sets

90 Helly's Selection Theorem/vague convergence

Given a sequence F_n of dist. fun.
There is a subsequence $F_{n(k)}$ that 'weakly' converges to a function G that is right continuous and nondecreasing (but may not go to 0,1 in the limits).

91 tight

the sequence F_n is tight if for every $\varepsilon > 0$ there is an M_ε so that

$$\limsup_{n\to\infty} 1 - F_n(M_\varepsilon) + F_n(-M_\varepsilon) \le \varepsilon \quad \iff \quad 1 - \varepsilon \le \liminf_{n\to\infty} F_n(M_\varepsilon) - F_n(-M_\varepsilon)$$

92 tightness criteria

F_n has a subsequence converging weakly to G and G is a distribution function if and only if F_n are tight.

3.3 Characteristic Functions

93 characteristic function

$\varphi_X(t) = E(e^{itX}) = E(\cos(tX)) + iE(\sin(tX))$

94 properties of characteristic functions

$\varphi(0) = 1 \quad \varphi(-t) = \overline{\varphi(t)} \quad |\varphi(t)| \le E|e^{itX}| \le 1 \quad \varphi_{aX+b}(t) = \varphi(at)e^{itb}$

95 combining characteristic functions

X, Y independent $\implies \varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$

96 characteristic function examples

Normal distribution $\varphi(t) = e^{-t^2/2}$
coin flip $\varphi(t) = Ee^{itX} = \frac{1}{2}e^{it} + \frac{1}{2}e^{-it} = \cos(t)$

97 inversion formula for characteristic functions

1-1 correspondence between φ and distribution functions
If $\varphi(t) = \int e^{itx}\mu(dx)$ then

$$\mu(a,b) + \tfrac{1}{2}\mu(\{a,b\}) = \lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\varphi(t)dt$$

If $\int |\varphi(t)|dt < \infty$ then μ has density

$$f(y) = \frac{1}{2\pi}\int e^{-ity}\varphi(t)dt$$

3.4 Central Limit Theorems

98 i.i.d. Central Limit Theorem

$X_1, X_2, \ldots$ i.i.d. with $EX_i = \mu$ and $\mathrm{var}(X_i) = \sigma^2 \in (0,\infty)$

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \Rightarrow \chi$$

where χ is the standard normal distribution ($\mu = 0$, $\sigma = 1$)

99 Lindeberg-Feller Central Limit Theorem

Triangular array $X_{n,m}$ independent variables for $1 \le m \le n$ where $EX_{n,m} = 0$ and

(i) $\sum_{m=1}^{n} EX_{n,m}^2 \to \sigma^2 > 0$ as $n \to \infty$

(ii) $\forall\varepsilon > 0 \sum_{m=1}^{n} E(|X_{n,m}|^2; |X_{n,m}| > \varepsilon) \to 0$ as $n \to \infty$

$$S_n = X_{n,1} + \cdots + X_{n,n} \Rightarrow \sigma\chi$$

3.6 Poisson Convergence

100 Poisson Convergence Theorem

$X_{n,m}$ independent Bernoulli variables for $1 \le m \le n$ with $P(X_{n,m} = 1) = p_{n,m} = 1 - P(X_{n,m} = 0)$

(i) $\sum_{m=1}^{n} p_{n,m} \to \lambda > 0$ as $n \to \infty$

(ii) $\max_{1 \le m \le n} p_{n,m} \to 0$ as $n \to \infty$

$$S_n = X_{n,1} + \cdots + X_{n,n} \Rightarrow Poisson(\lambda)$$

101 Poisson Processes

Generalizes Poisson convergence to larger class of variables taking non-negative integer values and still converging to $Poisson(\lambda)$.

3.10 Limit Theorems in $\mathbb{R}^d$

102 distribution functions in $\mathbb{R}^d$

$F(y) = P(X \le y) = P(X_i \le y_i)$ for all i.

Note that $F_i(y_i) = \lim_{n\to\infty} F(n, \ldots, n, y_i, n, \ldots, n)$

103 characteristic functions in $\mathbb{R}^d$

$\phi(t) = Ex^{it\cdot X} = Ee^{i}(t_1 X_1 + \cdots + t_d X_d)$

104 Central Limit Theorem in $\mathbb{R}^d$

$X_1, X_2, \ldots$ i.i.d random vectors with $EX_i = \mu \in \mathbb{R}^d$ with finite covariance $\Gamma_{i,j} = E(X_i - \mu_i)(X_j - \mu_j)$,

$$\frac{S_n - n\mu}{\sqrt{n}} \Rightarrow \mathcal{N}_d(0, \Gamma)$$

where $\mathcal{N}_d(0, \Gamma)$ is the multivariate Gaussian with mean 0 and covariance Γ.

Chapter 4 - Martingales

4.1 Conditional Expectation

105 conditional expectation

If $E|X| < \infty$ then $E(X|\mathcal{F})$ is any random variable such that

$$(i)\ E(X|\mathcal{F}) \in \mathcal{F} \qquad (ii)\ \forall A \in \mathcal{F},\ \int_A E(X|\mathcal{F})dP = \int_A X dP$$

This exists (by Radon-Nikodym Thm/derivatives) and is unique up to a.e.

106 major examples of conditional expectation

- Perfect information, $X \in \mathcal{F}$ then $E(X|\mathcal{F}) = X$
- No information, X and $\mathcal{F}$ are independent then $E(X|F) = EX$
- $\Omega_1, \Omega_2, \ldots$ disjoint partition of Ω then for $\mathcal{F} = \sigma(\Omega_1, \Omega_2, \ldots)$

$$E(X|\mathcal{F}) = \frac{E(X; \Omega_i)}{P(\Omega_i)} \text{ on } \Omega_i$$

107 $P(A|\mathcal{F})$, $P(A|B)$, $E(X|Y)$

If A, B are events, $\mathcal{F}$ is a σ-field, and X, Y random variables,

$$P(A|\mathcal{F}) = E(1_A|\mathcal{F}) \qquad P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad E(X|Y) = E(X|\sigma(Y))$$

108 Properties of $E(X|\mathcal{F})$

Where all conditional expectations are defined (i.e. $E|X| < \infty$ for $E(X|\mathcal{F})$

- $E(aX + Y|\mathcal{F}) = aE(X|\mathcal{F}) + E(Y|\mathcal{F})$ (regardless of independence)
- $X \leq Y$ implies $E(X|\mathcal{F}) \leq E(Y|\mathcal{F})$
- $X_n \geq 0$ and $X_n \uparrow X$ (and $E|X| < \infty$) then $E(X_n|\mathcal{F}) \uparrow E(X|\mathcal{F})$
- If $\mathcal{F} \subset \mathcal{G}$ then $E(E(X|\mathcal{F})|\mathcal{G}) = E(E(X|\mathcal{G})|\mathcal{F}) = E(X|\mathcal{F})$
- If $X \in \mathcal{F}$ (and $E|XY|, E|Y| < \infty$) then $E(XY|\mathcal{F}) = XE(Y|\mathcal{F})$

109 Jensen's Inequality for Conditional Probability

φ convex, X random variable with $E|X|, E|\varphi(X)| < \infty$ (so that both cond. exp exist) then

$$\varphi(E(X|\mathcal{F})) \leq E(\varphi(X)|\mathcal{F})$$

110 regular conditional probabilities

$X : (\Omega, \mathcal{F}) \to (\Omega, \mathcal{F})$ the identity map, then $\mu : \Omega \times \Omega \to [0, 1]$ is a regular conditional probability if

- for every $A \in \mathcal{F}$, $\omega \mapsto \mu(w, A)$ is a version of $P(A|\mathcal{F}) = E(1_A|\mathcal{F})$
- for every $\omega \in \Omega$, $A \mapsto \mu(w, A)$ is a probability measure

These reg. cond. prob. exist for $(\mathbb{R}, \mathcal{R})$ and other measure spaces that are measurably isomorphic to it (i.e. 'nice')

Motivation: If μ is a regular conditional probability, $E(g(X)|\mathcal{F}) = \int g(x)\mu(w, dx)$.

4.2 Martingales, Almost Sure Convergence

111 filtration, and adapted to filtration

$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n \subseteq \cdots$ an increasing sequence of σ-fields is a filtration

if X_n is a sequence of random variables with $X_n \in \mathcal{F}_n$ (i.e. X_n is $\mathcal{F}_n$-measurable) then X_n is adapted to the filtration F_n

112 martingale, submartingale, supermartingale

X_n is **martingale** w.r.t $\mathcal{F}_n$ filtration if

(i) $E|X_n| < \infty$ for all n (so cond. exp. exist)

(ii) X_n adapted to $\mathcal{F}_n$ (so $X_n \in \mathcal{F}_n$)

(iii) $E(X_{n+1}|\mathcal{F}_n) = X_n$ [$E(X_{n+1}|\mathcal{F}_n) \leq X_n$ is supermartingale, $E(X_{n+1}|\mathcal{F}_n) \geq X_n$ is submartingale]

By induction, for any $n > m$, $E(X_n|\mathcal{F}_m) = X_m$

X is submartingale $\iff$ $-X$ is supermartingale

113 linear martingales

Take S_0 a constant, and Y_n i.i.d random variables with mean 0, then $X_n = S_0 + Y_1 + \cdots + Y_n$ is martingale with respect to $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$. (if mean ≤ 0 then supermartingale, and mean ≥ 0 gives submartingale).

This is also an example of a random walk!

114 functions of martingales

If φ is (increasing) convex function, and X_n is (sub)martingale, then $\varphi(X_n)$ is submartingale. (proof by Jensen's Inequality for Conditional Expectations)

Example: $|X_n|^p$ submartingale, $(X_n - a)^+$ submartingale, X_n supermartingale then $\min(X_n, a)$ is supermartingale too

115 predictable sequence

H_n random variables such that $H_n \in \mathcal{F}_{n-1}$ (i.e. H_n only depends on information at time $n-1$)

If H_n is a betting scheme, and X_n is the martingale of money earned at time n betting a single unit each round, then

$$(H \cdot X)_n = \sum_{m=1}^{n} H_m(X_m - X_{m-1})$$

is the earnings of H_n at time n.

Theorem: If X_n is a (sub/super)martingale, and H_n is a predictable sequence with $H_n \geq 0$ and each H_n bounded, then $(H \cdot X)_n$ is a (sub/super)martingale as well.

116 Classical Martingale Betting Strategy

"Double down on losses"

$$H_n = \begin{cases} 2H_{n-1} & X_{n-1} - X_{n-2} = -1 \text{ (loss @ } n-1) \\ 1 & X_{n-1} - X_{n-2} = 1 \text{ (win @ } n-1) \end{cases}$$

117 stopping times

N, a random variable, is a stopping time if $\{N = n\} \in \mathcal{F}_n$ for all n (i.e. the decision to stop at time n must be decidable with the information at time n)

Theorem: If X_n is a (sub/super)martingale, then $X_{N \wedge n} = X_{\min(N,n)}$ is a (sub/super)martingale.

118 Upcrossing Inequality

If X_m is a submartingale, and $a < b$ with U_n the number of upcrossings of X_m of (a, b), then

$$(b - a)EU_n \leq E(X_n - a)^+ - E(X_0 - a)^+$$

119 Martingale A.S. Convergence Theorem

If X_n is a submartingale and $\sup_n EX_n^+ < \infty$ then X_n converges a.s. to some X with $E|X| < \infty$.

4.3 Examples

120 Bounded Increments Example

X_n martingale and $|X_{n+1} - X_n| \leq M < \infty$ for all n (i.e. has bounded increments) then if

$$C = \{\lim_n X_n = C < \infty\} \qquad D = \{\limsup_n X_n = \infty \text{ and } \liminf_n X_n = -\infty\}$$

then $P(C \cup D) = 1$.

121 Doob's Decomposition

X_n a submartingale, then $X_n = M_n + A_n$ **uniquely** where M_n is a martingale and A_n is an increasing predictable sequence.

Construction: $A_0 = 0$ and $A_n - A_n - 1 = E(X_n|\mathcal{F}_{n-1}) - X_{n-1} \in \mathcal{F}_{n-1}$ and then $M_n = X_n - A_n$ and check this is martingale.

122 Polya's Urn

An urn has r red balls, and g green balls. Each time, pick a ball and add c balls of chosen color.

Let X_n be the fraction of green balls after n draws and $\mathcal{F}_n$ is the information after n draws. Then X_n is martingale (because X_{n+1} independent of $\mathcal{F}_n$ and has $EX_{n+1} = X_n$ by direct computation.

4.4 Doob's Inequality, Convergence in L^p, $p > 1$

123 Inequality for expected values of X_N

If X_n is submartingale, N is a stopping time which is "bounded" meaning $P(N \geq k) = 1$ for some k, then

$$EX_0 \leq EX_N \leq EX_k$$

Pf Idea: $EX_0 = EX_{N\wedge 0} \leq EX_{N\wedge k} = EX_N$

Then take $K_n = 1_{N<n}$ predictable, so $(K \cdot X)_n$ is submartingale and

$(K \cdot X)_k = EX_k - EX_{N\wedge k} = EX_k - EX_N \geq E(K \cdot X)_0 = 0$.

Common application is $N \wedge n$ which is bounded by n.

124 Doob's Inequality

X_n submartingale, $\lambda > 0$,

$$\lambda P\left(\max_{0\leq m\leq n} X_m \geq \lambda\right) \leq EX_n^+$$

Pf Idea: Take $N = \inf\{X_m \geq \lambda \text{ or } m = n\}$. Then $X_N \geq \lambda$ if some $X_m \geq \lambda$, and $P(N \leq n) = 1$ so

$$\lambda P\left(\max_m X_m \geq \lambda\right) \leq EX_N 1_{\max_m X_m \geq \lambda} \leq EX_n 1_{\max_m X_m \geq \lambda}$$

and $EX_n 1_A \leq EX_n \leq EX_n^+$ always.

125 L^p maximal inequality

Idea: $(X_m^+)^p$ for $0 \leq m \leq n$ can be bounded in expectation above by $E(X_n^+)^p$ scaled by a constant depending only on p

Theorem: X_n submartingale and $1 < p < \infty$,

$$E\left(\max_{0\leq m\leq n} (X_m^+)^p\right) \leq \left(\tfrac{p}{p-1}\right)^p E(X_n^+)^p$$

Pf Idea: Take bounded version of variable, $X_n \wedge M$ which either matches $X_n \geq \lambda$ or trivially fails.

Express expectations in terms of integrals, apply Doob's Inequality and then perform some integral manipulations (Fubini's and regular integration) then apply Holder's Inequality.

126 L^p convergence

Theorem *X_n martingale, with $\sup E|X_n|^p < \infty$ for $p > 1$, then $X_n \to X$ a.s. and in L^p.*

Proof. Apply normal convergence to get a.s. to X. Then bound $|X_n - X|^p$ by $\sup |X_n|^p$ using this a.s. conv, which is in L^p by the L^p maximal inequality.

4.6 Uniform Integrability, Convergence in L^1

127 Uniform integrable

A collection X_n is uniformly integrable if

$$\lim_{M\to\infty}\left(\sup_n E\left(|X_i|;|X_i|>M\right)\right)=0$$

128 pre-L^1 convergence theorem

If X_n s.t. $X_n \to X$ in P and $E|X_n| < \infty$ for all n then TFAE

1. X_n are U.I.
2. $X_n \to X$ converge in L^1
3. $E|X_n| \to E|X| < \infty$

129 L^1 convergence theorem

If X_n is a submartingale, TFAE:

1. X_n are U.I.
2. X_n converge in L^1 and a.s.
3. X_n converge in L^1

4.7 Backwards Martingales

130 Backwards martingale

Backwards Martingales: X_{-n} indexed by $n = 1, 2, 3, \ldots$ and adapted to the filtration $\mathcal{F}_{-n}$, (i.e. $\cdots \subseteq \mathcal{F}_{-2} \subseteq \mathcal{F}_{-1} \subseteq \mathcal{F}_0$) such that $E(X_{-n+1} \mid \mathcal{F}_{-n}) = X_{-n}$ (i.e. $E(X_0 \mid \mathcal{F}_{-1}) = X_{-1}$)

131 Convergence Theorem for Backwards Martingales

Backwards Convergence Theorem If X_n is a backwards martingale, it converges a.s. and in L^1.

4.8 Optional Stopping Theorems

132 Optional Stopping

$EX_0 \le EX_N \le EX_\infty$

133 Exampes of Optional Stopping Theorems

1. Whenever usual integral convergence theorems hold for $EX_{N\wedge n}$.
2. If N is bounded a.s., that is $P(N \le k) = 1$ for some k. Then $EX_0 \le EX_N \le EX_k$.
3. X_n U.I. then $EX_0 \le EX_N \le EX_\infty$.
4. $E(|X_{n+1} - X_n|\mathcal{F}_n) \le B < \infty$ a.s. and $EN < \infty$

134 Wald's Identity

Thm If $X_1, X_2, \ldots$ are i.i.d. with $EX_i = 0$ and $S_n = X_1 + \cdots + X_n$ and N is a stopping time with $EN < \infty$ then $ES_N = \mu EN$.

Chapter 5 - Markov Chains

5.1 Examples

135 Markov Chain

a memoryless stochastic process. That is a sequence of random variables X_n such that $P(X_n \in A \mid X_1, \ldots, X_{n-1}) = P(X_n \in A \mid X_{n-1})$.

136 transition probability

$p(x, y) = P(X_1 = y \mid X_0 = x)$
$p^n(x, y) = P(X_n = y \mid X_0 = x)$
$p(x, A) = P(X_1 \in A \mid X_0 = x)$

5.2 Construction, Markov Properties

137 Markov Property

Let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, then $P(X_n \in A \mid \mathcal{F}_{n-1}) = P(X_n \in A \mid X_{n-1})$ for all n.

138 Strong Markov Property

Let N be a stopping time and define $\mathcal{F}_N = \{A : A \cap \{N = n\} \in \sigma(X_1, \ldots, X_n)\}$, then $P(X_N \in A \mid \mathcal{F}_{N-1}) = P(X_N \in A \mid X_{N-1})$.

5.3 Recurrence and Transience

139 ρ_{xy}, T_y^k

T_y^k is the time for the kth return y (not including X_0). $T_y = T_y^1$ is for the first visit to y.
$\rho_{xy} = P_x(T_y < \infty)$.

140 recurrence

x is recurrent if $\rho_{xx} = 1$ (that is starting at x a.s. returns to x in finite time.)

141 transience

x is transient if it is not recurrent, that is $\rho_{xx} < 1$.

142 closed

a collection of states is closed if you can never ever escape! that is if $x \in C$ and $\rho_{xy} > 0$ then $y \in C$ too.

143 irreducible

A collection of states is irreducible if all states are connected, that is $x, y \in C$ implies $\rho_{xy} > 0$ for all pairs.

5.5 Stationary Measures

144 stationary measure

A stationary measure μ satisfies $\mu(y) = \sum_x \mu(x) p(x, y)$.
This also implies for any n, $\mu(y) = \sum_x \mu(x) p^n(x, y)$.

145 reversible measure

A measure is reversible if it satisfies the detailed balance condition $\mu(y)p(y,x) = \mu(x)p(x,y)$.

146 stationary distribution

A stationary distribution is a probability measure that is stationary, so satisfies $\mu(y) = \sum_x \mu(x)p(x,y)$ and $\sum_y \mu(y) = 1$.

147 positive recurrent and null recurrent

x is positive recurrent if $E_x T_x < \infty$, if x is recurrent but not positive recurrent it is null recurrent.

148 Thm for Existence for Stationary Distribution

If p is irreducible then TFAE

- these exists one positive recurrent state
- all states are positive recurrent state
- there exists a (unique!) stationary measure

5.6 Asymptotic Behavior

149 period for a markov chain

Let $I_x = \{n : p^n(x,x) > 0\}$, then $d_x = \gcd(I_x)$ is the period of x. These are unique on irreducible sets of states.

150 aperiodic

When the period for any state (all states when irreducible) is 1.

151 Convergence Theorem for Markov Chains

If p is irreducible and aperiodic with a stationary measure π, then $p^n \to \pi$ as $n \to \infty$.

Probability Theory Quals Questions (– best questions –)

Chapter 1 - Measure Theory

1.1 Probability Spaces

1 Show that $\mathcal{R}^d$, the Borel sets on $\mathbb{R}^d$, is countably generated

countably generated: there exists a countable collection $\mathcal{C}$ such that the σ-field can be expressed as $\sigma(\mathcal{C})$

Let $S\mathbb{Q}_d$ be the empty set and all sets of the form

$$(a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_d, b_d]$$

for $-\infty \le a_i < b_i \le \infty$ and $a_i, b_i \in \mathbb{Q} \cup \{\pm\infty\}$

Claim 1: $S\mathbb{Q}_d$ is countable

there are finitely many interval endpoints and countably many options for each, so this set is countable

Claim 2: $\sigma(S\mathbb{Q}_d) = \mathcal{R}^d$ (generates the Borel sets)

First, need to show $S\mathbb{Q}_d \subseteq \mathcal{R}^d$ (then $\sigma(S\mathbb{Q}_d) \subseteq \sigma(\mathcal{R}^d) = \mathcal{R}^d$).

Emptyset is in $\mathcal{R}^d$. Take $(a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_d, b_d] \in S\mathbb{Q}_d$. Since σ-fields are closed under complements and countable union, they are closed under countable intersections.

Take $\{(a_1, b_1+1/n) \times (a_2, b_2+1/n) \times \cdots \times (a_d, b_d+1/n)\}_{n\in\mathbb{N}}$ a countable collection in $\mathcal{R}^d$, the countable intersection yields the desired set.

Second, need to show that every open set in $\mathbb{R}^d$ is in $\sigma(S\mathbb{Q}_d)$ (then $\mathcal{R}^d = \sigma(opens) \subseteq \sigma(\sigma(S\mathbb{Q}_d)) = \sigma(S\mathbb{Q}_d)$), suffices to take basic opens of the form $(a_1, b_1) \times \cdots (a_d, b_d)$ $(-\infty \le a_i < b_i \le \infty)$

Take $(a_1, b_1) \times \cdots (a_d, b_d)$ a basic open in $\mathbb{R}^d$. For each a_i, b_i there is a sequence of rationals $c_{i,n}, d_{i,n}$ such that $c_{i,n} \downarrow a_i$ and $d_{i,n} \uparrow b_i$.

Take the countable collection $\{(c_{1,n}, d_{1,n}] \times \cdots \times (c_{d,n}, d_{d,n}]\}_{n\in\mathbb{N}}$. Each set in the collection is in $S\mathbb{Q}_d$ so their countable union is in $\sigma(S\mathbb{Q}_d)$ and their union gives the basic open we started with.

2 How do σ-fields, semialgebras, and algebras relate? What are examples/non-examples of each?

$$\text{SEMIALGEBRAS} \supsetneq \text{ALGEBRAS} \supsetneq \sigma\text{-ALGEBRAS}$$

<u>σ-ALGEBRA $\Longrightarrow$ ALGEBRA:</u>

σ-algebra: closed under complements and countable unions

algebra: (i) closed under complements and (ii) *finite* unions

(i) complement closure property is the same

(ii) finite unions are countable

<u>ALGEBRA $\Longrightarrow$ SEMIALGEBRA:</u>

algebra: closed under complements and finite unions

semi-algebra: (i) closed under finite intersections and (ii) complements are finite disjoint unions of sets in the collection

(i) finite unions and complements provides finite intersections

(ii) if complement is in the collection, it also a finite union of disjoint sets (just 1)

<u>Example: Semialgebra not an algebra</u>

$\mathcal{S}_d = \{\emptyset\} \cup \{(a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_d, b_d] : -\infty \leq a_i < b_i \leq \infty\}$

finite intersections contract endpoints (possibly with empty intersection) but preserves $(\cdot, \cdot]$

complements - the complement in each dimension will be $\emptyset$, $(-\infty, a_i] \cup (b_i, \infty]$ which is a finite disjoint union of sets in that dimension, then taking products over all dimension yields the same overall.

Example: Algebra not an σ-algebra

$\Omega = \mathbb{Z}$ and $\mathcal{A}$ is the collection of integer sets that are finite or co-finite ($|A|$ or $|A^C|$ is finite).

closed under complements by construction. Finite unions are also closed, if $A, B \in \mathcal{A}$ and $|A|, |B| < \infty$ then $|A \cup B| < \infty$, if $|A^C| < \infty$ then $|(A \cup B)^C| = |A^C \cap B^C| \leq |A^C| < \infty$ (similarly for $|B^C| < \infty$.

However not closed under countably infinite unions, take $A_n = \{2n\}$ then $\cup_n A_n = 2\mathbb{Z}$ which is neither finite nor co-finite.

3 How do 'measures' extend from semi-algebras to algebras to σ-algebras?

Given $\mathcal{S}$ semi-algebra and a set function μ on $\mathcal{S}$ with $\mu(\emptyset) = 0$, we can extend to a measure on the algebra $\overline{S}$ by

$$\mu(\sqcup_i A_i) = \sum_i \mu(A_i)$$

Extending from semi-algebra to σ-algebra:

If μ is a set function on $\mathcal{S}$, a semi-algebra, with $\mu(\emptyset) = 0$ and additive on finite disjoint unions and sub-additive on infinite unions then μ extends uniquely to $\sigma(\mathcal{S})$.

4 Let $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$ be σ-algebras, what can we say about $\cup_i \mathcal{F}_i$?

This is an algebra but *not* necessarily a σ-algebra

$\cup_i \mathcal{F}_i$ is an algebra:

closed under complements - if $U \in \cup_i \mathcal{F}_i$ then $U \in \mathcal{F}_i$ for some i, so $U^C \in \mathcal{F}_i \subseteq \cup_i \mathcal{F}_i$.

closed under finite intersection/union - given $U, V \in \cup_i \mathcal{F}_i$ then $U \in \mathcal{F}_i$ and $V \in \mathcal{F}_j$ for some i, j. Let $n = \max\{i, j\}$ then $U, V \in \mathcal{F}_n$ and here we have closed finite intersection/union so $U \cup V, U \cap V \in \mathcal{F}_n \subseteq \cup_i \mathcal{F}_i$

$\cup_i \mathcal{F}_i$ may not be a σ-algebra:

What goes wrong: could have a sequence $U_1 \in \mathcal{F}_1, U_2 \in \mathcal{F}_2, \ldots, U_n \in \mathcal{F}_n, \ldots$ where the union of U_is is not contained in any one $\mathcal{F}_i$.

Example (trying to create a union that gives the algebra $\mathbb{Z}$ with finite/co-finite sets):

Let $\mathcal{F}_n$ be the σ-algebra generated by the singletons $\{0\}, \{\pm 1\}, \ldots, \{\pm n\}$ in $\mathbb{Z}$ (so complements are in $\mathbb{Z}$). The union of all of these gives the desired algebra because a set that is finite is contained in some interval $[-M, M]$ in which case it is in $\mathcal{F}_M$, so all finite or co-finite sets in $\mathbb{Z}$ lie in one of these σ-algebras.

1.2 Distributions

5 Suppose X and Y are random variables on $(\Omega, \mathcal{F}, P)$ and let $A \in \mathcal{F}$. If $Z(\omega) = X(\omega)$ for $\omega \in A$ and $Z(\omega) = Y(\omega)$ for $\omega \notin A$ then Z is a random variable.

$Z^{-1}(B) = X^{-1}(B) \cap A \cup Y^{-1}(B) \cap A^C \in \mathcal{F}$

6 Show that a distribution function has at most countably many discontinuities

If a distribution function has a discontinuity at x_0 then it is a left discontinuity (F is right continuous) so $F(x_0) > F(x_0-) = \lim_{y \uparrow x_0} F(y)$ and so $P(X = x_0) = \varepsilon_0 > 0$.

Since F is a distribution function, $\lim_{x\to\infty} F(x) = 1$, which importantly is bounded. At the same time, for each discontinuity jump, we know that F goes up by $\varepsilon > 0$. By the below claim, if there were uncountably many discontinuities then F would be unbounded in the limit, a contradiction.

Claim: The sum of uncountably many positive terms is unbounded.

Let $\sum_\alpha \varepsilon_\alpha$ be an uncountable sum of positive terms. Define $B_n = \{\varepsilon_\alpha : \varepsilon \geq 1/n\}$. If $|B_n|$ is infinite for any n then the sum is bounded below by $\sum_{\varepsilon\in B_n} 1/n$ which is unbounded. Since there are countably many B_n, an uncountable number of terms cannot be partitioned into countably many finite chunks, so the sum must be unbounded.

7 What properties characterize distribution functions?

(i) F is nondecreasing

(ii) $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$

(iii) F is right continuous

Theorem Any real function satisfying these is the distribution function of some random variable. Define $\Omega = (0,1)$ and $\mathcal{F}$ the Borel sets, with P Lebesgue measure. For $\omega \in (0,1)$,

$$X(\omega) = \sup\{y : F(y) < \omega\}$$

X here is a sort of inverse to F, sometimes denoted F^{-1}.

8 Give an example of a density function whose distribution function has no closed form.

Standard Normal Distribution

Given by the density function $f(x) = (2\pi)^{-1/2}\exp(-x^2/2)$

has no closed form but has upper and lower bounds.

9 Give an example of a distribution function with dense discontinuities.

Enumerate the rationals by $q_1, q_2, \ldots$ define the indicator functions by $1_q = 1_{[q,\infty)}$ (i.e. indicates whether past a given rational).

Choose $\alpha_i > 0$ such that $\sum_i \alpha_i = 1$ and define

$$F(x) = \sum_{i=1}^{\infty} \alpha_i 1_{q_i}$$

Has dense discontinuities because $\mathbb{Q}$ is dense in the reals, but has only countably many discontinuities.

10 Given a random variable with density function f, derive the density function for X^2.

$$\begin{aligned} F_{X^2}(x) &= P(X \leq x) = P(X \in [-\sqrt{x}, \sqrt{x}]) = P(X \leq \sqrt{x}) - P(X < -\sqrt{x}) \\ &= \int_{-\infty}^{\sqrt{x}} f(y)dy - \int_{-\infty}^{-\sqrt{x}} f(y)dy = \int_{-\sqrt{x}}^{\sqrt{x}} f(y)dy \end{aligned}$$

Differentiating,

$$f_{X^2}(x) = \tfrac{d}{dx}F_{X^2}(x) = \tfrac{d}{dx}\int_{-\sqrt{x}}^{\sqrt{x}} f(y)dy = f(\sqrt{x})\tfrac{1}{2}\sqrt{x}^{-1} + f(-\sqrt{x})\tfrac{1}{2}\sqrt{x}^{-1} = \tfrac{1}{2}\sqrt{x}^{-1}(f(\sqrt{x}) + f(-\sqrt{x}))$$

1.3 Random Variables

11 What is the extended real line? Why do we extend random variables to it?

The extended real line, $\mathbb{R}^*$, is obtained by adding $\pm\infty$ to $\mathbb{R}$. These endpoints are also added to all intervals with no lower/upper bound so $[-\infty, a)$ and similar intervals now generate the Borels on $\mathbb{R}^*$.

In cases were we look at inf or sup of random variables, we may end up with a 'random variable' that takes on values of $\pm\infty$ which lie outside the usual real line, so we extend to make sense of slightly more general random variables.

12 If X and Y are two random variables, show that $X + Y$ is one too.

Suppose $X, Y : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{R})$. Want to show on just the intervals $(-\infty, a)$ that

$$(X + Y)^{-1}((-\infty, a)) = \{X + Y < a\} \in \mathcal{F}.$$

Need to split conditions to get separate $X < \bullet$ and $Y < \bullet$ and also enumerate over something countable.

$$\{X + Y < a\} = \bigcup_{q \in \mathbb{Q}} \{X < a - q\} \cap \{Y < q\}$$

all pieces are in $\mathcal{F}$ and combined through countable intersections/unions so this is a measurable set.

13 What is the smallest σ-field that makes all continuous functions $\mathbb{R}^d \to \mathbb{R}$ measurable?

Using the Borels for $\mathbb{R}$ the smallest σ-field will be the Borels $\mathcal{R}^d$.

Take the projection map onto the kth coordinate $f_k : \mathbb{R}^d \to \mathbb{R}$. For this to be measurable, $f_k^{-1}(a, b)$ must be measurable. However

$$f_k^{-1}(a, b) = \mathbb{R}^{d-1} \times (a, b)$$

and since this holds for all k and (a, b) and these are all measurable, any σ-field containing them contains arbitrary intersections to get $(a_1, b_1) \times \cdots \times (a_d, b_d)$.

These then generate all opens and all Borels. So any σ-field making all continuous functions measurable must contain the Borels, and since this is also enough we see that this is the smallest σ-field.

1.4 Integration

14 What are integrable functions? How do we develop integration of them?

First defining integrals for simple functions in the natural way, then extending to bounded functions by approximating them with simple functions above/below.

Then using bounded functions to approximate non-negative functions. Finally splitting general (integrable) functions into two non-negative pieces and combining in the natural way.

1.5 Properties of the Integral

15 Show that $||f||_p \to ||f||_\infty$ when μ is a probability measure. What if μ is only finite?

Show inequality in both directions.

One side show $||f||_p$ bounded above by all M defining the inf and so the sequence is bounded.

On the other side, if there is a gap pick some value N less than $||f||_\infty$ and take limit to show that $||f||_p > N$ for some p

The above proof uses the fact that $\lim_{p\to\infty} \mu(E)^{1/p} \to 1$, if μ is just finite and not a probability measure, this will still hold even when $\mu(E) > 1$.

16 State Hölder's Inequality. What happens when $p = 1$ and $q = \infty$?

Hölder's Inequality: *If $p, q \in (1, \infty)$ and $1/p + 1/q = 1$, then for functions f, g,*

$$\int |fg| d\mu \leq \left(\int |f|^p d\mu \right)^p \left(\int |g|^q d\mu \right)^q = ||f||_p ||g||_q$$

Extension: This continues to hold when $p = 1$ and $q = \infty$ where we define

$$||f||_\infty = \inf\{M : \mu(\{x : |f(x)| > M\}) = 0\}$$

Proof Sketch:

Pick M that defines inf and show inequalities, then take inf

17 When does $\int \sum_n f_n d\mu = \sum_n \int f_n d\mu$?

Monotone Convergence: *When $f_n \geq 0$ and $f_n \uparrow f$ then $\int f_n d\mu \uparrow \int f d\mu$*

So if $g_m \geq 0$ and $f_n = \sum_{m=0}^n g_m$ then monotone convergence implies $\sum_m \int g_m d\mu = \int \sum_m g_m d\mu$.

Dominated Convergence: *If $f_n \to f$ (a.e.) and $|f_n| \leq g$ for integrable g, then $\int f_n d\mu \to \int f d\mu$.*

So if $\sum_n \int |f_n| d\mu < \infty$ then $F_\infty = \sum_n |f_n|$ is integrable (using Monotone convergence as above to switch sums/integral and by assumption). And $F_n = \sum_{i=1}^n f_i$ satisfies $|F_n| \leq F_\infty$ so $\int F_n \uparrow \sum_n \int f_n = \int \sum_n f_n$ (first switching finite sums with integrals and then taking the limit)

1.6 Expected Value

18 Let $f \geq 0$, how can we approximate f from below with f_n simple functions?

$f_n(x) = \min\{([2^n f(x)]/2^n), n\}$

Splits $[k, k+1]$ into 2^n pieces and flattens f in these sections, when f goes above n the function flattens out to n (this gives finitely many regions for values, hence a simple function) but at $n \to \infty$ $f_n \uparrow f$.

1.7 Product Measures, Fubini's Theorem

19 State Fubini's Theorem. What if instead we know that $\int_X \int_Y |f(x,y)| d\mu_2 d\mu_1 < \infty$, what can we conclude?

Fubini's gives conditions for switching multiple integrals using product spaces

Fubini's Theorem: *Let μ_1, μ_2 be σ-finite with $\mu = \mu_1 \times \mu_2$. If $f \geq 0$ or $\int |f| d\mu < \infty$ then*

$$\int_X \int_Y f d\mu_2 d\mu_1 = \int_{X\times Y} f d\mu = \int_Y \int_X f d\mu_1 d\mu_2$$

If instead we have $\int_X \int_Y |f(x,y)| d\mu_2 d\mu_1 < \infty$, then taking $F = |f|$ we have that $F \geq 0$, so we can apply Fubini's to $|f|$ to get $\int_{X\times Y} |f| d\mu = \int_X \int_Y |f(x,y)| d\mu_2 d\mu_1 < \infty$. Now that $\int |f| d\mu < \infty$ we can apply Fubini's to the original function f.

20 State Fubini's Theorem. What happens if we drop each of the conditions?

Fubini's gives conditions for switching multiple integrals using product spaces

Fubini's Theorem: *Let μ_1, μ_2 be σ-finite with $\mu = \mu_1 \times \mu_2$. If $f \geq 0$ or $\int |f| d\mu < \infty$ then*

$$\int_X \int_Y f d\mu_2 d\mu_1 = \int_{X \times Y} f d\mu = \int_Y \int_X f d\mu_1 d\mu_2$$

Dropped $f \geq 0$: If f is not non-negative, then Fubini's Theorem can fail. Consider the function on $\mathbb{N} \times \mathbb{N}$ that takes on 1 on the main diagonal and -1 below the main diagonal

$$\begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & 1 & \cdots \\ 0 & 0 & 1 & -1 & \cdots \\ 0 & 1 & -1 & 0 & \cdots \\ 1 & -1 & 0 & 0 & \cdots \end{bmatrix}$$

then summing with respect to the counting measure row first gives 0 and columns first gives 1 so the conclusion fails.

Dropped $\int |f| d\mu < \infty$: If f is non-negative and does not meet this condition, then f is not actually integrable in the product measure μ, hence the proof for for Fubini's Theorem falls apart, using the same example above where the $\int |f| d\mu$ diverges to ∞ also.

Dropped μ_i is σ-finite: Take $X = Y = (0, 1)$ but μ_1 is Lebesgue (Borel sets) and μ_2 is counting measure (all subsets). Then $f(x, x) = 1$ otherwise $f = 0$ gives different integrals (point mass at $x = y$ for counting measure gives a 1 if integrated first, otherwise is measure 0 so integrates to 0).

21 Use Fubini's Theorem to derive an expression for $E|X|$.

Rewrite

$$E|X| = \int_\Omega |X| dP = \int_\Omega \int_0^{|X|} x dx dP = \int_\Omega \int_0^\infty 1_{|X|>x} dx dP$$

Since $1_{|X|>x} \geq 0$, we can apply Fubini's to switch the order of integration,

$$E|X| = \int_\Omega \int_0^\infty 1_{|X|>x} dx dP = \int_0^\infty \int_\Omega 1_{|X|>x} dP dx = \int_0^\infty P(|X| > x) dx$$

Chapter 2 - Laws of Large Numbers

2.1 Independence

22 State Dynkin's π-λ Theorem. Why is it significant?

π-λ Theorem: *If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system with $\mathcal{P} \subseteq \mathcal{L}$ then $\sigma(\mathcal{P}) \subset \mathcal{L}$.*

Supporting Definitions:

π-system - closed under intersection

λ-system - $\Omega \in \mathcal{L}$, countable unions contained, and set subtraction contained ($A \subseteq B \Rightarrow B \cap A^C \in \mathcal{L}$)

Significance: This theorem allows us to lift properties from a generating set (that is a π-system) to the σ-algebra it generates.

Example: Independence
Since independence can often be formulated as independence of σ-algebras, this allows us to check only a π-system that generates instead of the entire σ-algebra.

Example: agreement of measures
If μ_1,μ_2 agree on a π-system they also agree on the σ-field generated by it ($\mathcal{L} = \{A : \mu_1(A) = \mu_2(A)\}$)

23 Give an example of four random variables where any three are independent but all four are not.

Let X_1, X_2, X_3, X_4 be independent random variables in $\{-1, 1\}$ with $P(X_i = \pm 1) = 1/2$ for all i.

Let $Y_1 = X_1X_2$, $Y_2 = X_3X_4$, $Y_3 = X_1X_3$, and $Y_4 = X_2X_4$

$P(Y_i = n) = 1/2$ for all i and $n = \pm 1$.

three Y_is are independent
If we specify 3 Y_i's we can write out 3 X_is in terms of a single one, so the probability is 1/8 (two outcomes for independent with odds of 1/16 for alignment each) which satisfies independence.

four Y_is are not independent
However $Y_1Y_2 = Y_3Y_4$ so $Y_1 = Y_2 = 1$ and $Y_3 = 1$ but $Y_4 = -1$ can never happen, so it has probability 0 but the product of independent probabilities is 1/16.

24 If two collections of sets are independent, are their generated σ-fields also independent? When can we ensure that they are?

Not true for all collections of sets.

Take $\Omega = \{1, 2, 3, 4\}$ and $P(\{n\}) = 1/4$ for all $n \in \Omega$.

$\mathcal{A}_1 = \{\{1, 2\}, \{1, 3\}\}$ and $\mathcal{A}_2 = \{\{1, 4\}\}$

These are independent as collections of sets

$$P(\{1, x\} \cap \{1, 4\}) = P(\{1\}) = 1/4 = (1/2)(1/2) = P(\{1, x\})P(\{1, 4\})$$

But $\{1, 2, 3\} \in \sigma(\mathcal{A}_1)$ and

$$P(\{1, 2, 3\} \cap \{1, 4\}) = P(\{1\}) = 1/4 \neq (3/4)(1/2) = P(\{1, 2, 3\})P(\{1, 4\})$$

so their σ-algebras are *not* independent.

If we restrict to the condition that the collections be π-systems, that is closed under intersection, then independence is preserved. (here $\mathcal{A}_1$ is not a π-system)

25 Show that the sum of two independent Poisson distributions is again Poisson.

Well Poisson means $P(X = k) = e^{-\lambda}\lambda^k/k!$ and $P(Y = k) = e^{-\mu}\mu^k/k!$ when $k = 0, 1, 2, \ldots$

$$P(X + Y = n) = \sum_m P(X = m)P(Y = n - m) = \sum_{0 \le m \le n} \tfrac{e^{-\lambda}\lambda^m}{m!}\tfrac{e^{-\mu}\mu^{n-m}}{(n-m)!}$$
$$= e^{-(\lambda+\mu)}\tfrac{1}{n!}\sum_{m=0}^{n} \tfrac{n!}{m!(n-m)!}\lambda^m\mu^{n-m} = e^{-(\lambda+\mu)}(\lambda + \mu)^n/n!$$

which is a Poisson distribution with parameter $\lambda + \mu$.

26 Given distribution functions $F_1, \ldots, F_n$, how can you construct independent random variables with these distribution functions?

Since these are distribution functions, construct a measure $\mu_i((a, b]) = F_i(b) - F_i(a)$ on $\mathbb{R}$. And extend this to the product $P = \mu_1 \times \mu_n$.

Then take X_i to be projection from $\mathbb{R}^n$ onto $\mathbb{R}$ by the ith coordinate. Then

$$P(X_i \le x) = \mu_i((-\infty, x]) \prod_{i \ne j} \mu_j(\mathbb{R}) = F_i(x)$$

and independence follows by choice of P.

2.2 Weak Laws of Large Numbers

27 State different types of convergences, how do they compare? Give distinguishing examples for each.

converges almost surely (almost everywhere): $\mu(\{x : f_n(x) \ne f(x)\}) \to 0$

converges in L^p: $E|f_n - f|^p \to 0$

converges in probability for all $\varepsilon > 0$, $P(\{x : |f_n(x) - f(x)| > \varepsilon\}) \to 0$ (equivalent to $\ge \varepsilon$)

convergences almost surely (almost everywhere) $\implies$ convergence in probability

L^p convergence $\implies$ convergence in probability

Example Ideas:

- for L^p not converging, get smaller intervals but weight higher and higher
- for a.e. not converging, make sure to shift intervals around to always return to the points again

Example: convergences in probability and L^p but *not* a.e.
Take shifting and shrinking subsets of $[0,1]$ (e.g. $[0,1]$, $[0,1/2]$, $[1/2,1]$, $[0,1/4]$, $[1/4,1/2]$, etc) and let $f_n = \mathbb{1}_{A_n}$. Then the regions where f_n and 0 differs shrinks in measure to 0 so $f_n \to 0$ in measure, but not a.e. because each point in $[0,1]$ differs from 0 for arbitrarily high n.

Converges in L^p because $E|f_n|^p = E\mathbb{1}_{A_n} = 1/2^n \to 0$ for all p.

Example: convergences in probability but *not* in L^p
Take the above example but weight each f_n by $\mu(A_n)^{-1}$ so that $E|f_n|^p = E\mu(A_n)^{-p}\mathbb{1}_{A_n} = \mu(A_n)^{1-p}$ which does not converge to 0 as $n \to \infty$ ($\mu(A_n) \to 0$) for any $p \ge 1$

Example: convergences a.e. but *not* in L^p
$f_n = n\mathbb{1}_{[0,1/n]}$ converges a.e. to 0 but not in L^p because $E|f_n|^p = \int n^p \mathbb{1}_{[0,1/n]} = n^{p-1}$ which does not go to 0 when $p \ge 1$

Example: Not converging in all 3
$f_n = (-1)^n$ on $[0,1]$. Then for any function f, if $P(\{x : |f_n - f| > 1/2\}) \to 0$ then it must be below 1 at some point, so for N even, $f \in [1/2, 3/2]$ but then for f_{N+1} $|f_{N+1} - f| \ge 1/2$ on then entire interval, so no convergence. Can't converge in L^p or a.s. since those imply in probability.

28 What is the Weak Law of Large Numbers? Sketch a proof. What if we don't have finite variance?

Weak Law of Large Numbers: *Let $X_1, X_2, \dots$ be i.i.d with finite variance (can weaken to $E|X_i| < \infty$). Let $S_n = X_1 + X_2 + \cdots + X_n$ and $\mu = EX_1$. Then $S_n/n \to \mu$ converges in probability.*

Proof: (assuming finite variance)
Let $\mathrm{var}(X_i) = \sigma^2$, then because i.i.d., $\mathrm{var}(S_n/n) = \frac{1}{n^2}\mathrm{var}(S_n) = \frac{1}{n}\sigma^2$. (and $E(S_n) = n\mu$)

Chebyshev's Inequality:

$\varphi(X) = X^2 = |X|^2$ and $A = \{x : |S_N/n - \mu| > \varepsilon\} = \{x : |S_N/n - \mu|^2 > \varepsilon^2\}$.

$$\inf_A \varphi(S_n/n - \mu) \cdot P(A) \le E\varphi(S_n/n - \mu) = E(S_n/n - \mu)^2 = \mathrm{var}(S_n/n) = \sigma^2/n$$

so then

$$P(A) = P(\{x : |S_N/n - \mu| > \varepsilon\}) \leq \sigma^2/n\varepsilon^2 \to 0 \quad \text{as} \quad n \to \infty$$

so $S_n/n \to \mu$ in probability.

Extension without finite variance

Can get the same result assuming only $E|X_i| < \infty$ (instead of $EX_i^2 < \infty$), the proof uses truncation and version of the weak law for triangular arrays.

2.3 Borel-Cantelli Lemmas

29 What are the Borel-Cantelli Lemmas? How do they relate?

Borel-Cantelli Lemma: *If* $\sum_{n=1}^{\infty} P(A_n) < \infty$ *then*

$$P(\{x : x \in A_n \text{ infinitely often}\}) = P(\limsup_{n\to\infty} A_n) = P(\lim_{n\to\infty} \cup_{m=n}^{\infty} A_m) = \boxed{P(A_n \text{ infinitely often}) = 0}$$

Second Borel-Cantelli Lemma: *If* A_n *are* *independent* *and* $\sum P(A_n) = \infty$ *then*

$$P(\{x : x \in A_n \text{ infinitely often}\}) = P(\limsup_{n\to\infty} A_n) = P(\lim_{n\to\infty} \cup_{m=n}^{\infty} A_m) = \boxed{P(A_n \text{ infinitely often}) = 1}$$

The second Borel-Cantelli Lemma is a partial converse ($\neg p + r \implies \neg q$) assuming also independence.

30 What is the second Borel-Cantelli Lemma? What happens if we remove the independence condition?

Second Borel-Cantelli Lemma: *If* A_n *are* *independent* *and* $\sum P(A_n) = \infty$ *then*

$$P(\{x : x \in A_n \text{ infinitely often}\}) = P(\limsup_{n\to\infty} A_n) = P(\lim_{n\to\infty} \cup_{m=n}^{\infty} A_m) = \boxed{P(A_n \text{ infinitely often}) = 1}$$

Fails if not Independent: $A_n = (0, 1/n)$ so $a_n \to 0$ then $\limsup_n A_n = \emptyset$ but $\sum P(A_n) = \sum a_n = \sum 1/n = \infty$. These events are not independent because $P(A_n \cap A_m) = P(A_m) \neq P(A_n)P(A_m)$ where $m > n$.

31 Assume $X_k \to X$ in probability and g is a continuous function. Is it true that $g(X_k) \to g(X)$?

Yes, in probability at least.

$X_k \to X$ in probability is equivalent to every subsequence has a subsequence that converges a.s., so $g(X_{m_k}) \to g(X)$. This means that this holds for $g(X_n)$ on these subsequences. Then using the equivalence again we have $g(X_n) \to g(X)$ in probability.

What about $g(X_n) \to g(X)$ a.s.? Well not always, for example if $g(y) = y$ then if X_k does not converge a.s. then $g(X_n) = X_n$ does not converge a.s. either.

32 Use the Borel-Cantelli Lemmas to construct a sequence of random variables that converges in probability but not almost surely.

Try indicator functions converging to 0, so take supports A_n with measure $1/n$. Could take these rotating across interval $(0, 1)$.

To apply Borel-Cantelli, we want these to be independent to have $P(A_n i.o.) = 1$. How to construct A_n with measure $1/n$ and independent?

Records! Take $X_1, X_2, \ldots$ i.i.d. and let A_n be the collection where X_n is larger than $X_1, \ldots, X_{n-1}$. These have measure $1/n$ and are independent. Intuitively, because X_n needs to be larger than all preceding ones, regardless of the order of the preceding X_i, so A_k independent from A_n. More rigorously, take an ordering, this defines a permutation, by symmetry these are uniformly distributed, so all are equally likely.

2.4 Strong Law of Large Numbers

33 State the Strong Law of Large Numbers. Sketch a proof (you may assume $EX_i^4 < \infty$). Can we weaken the assumptions? What happens if $E|X_i| = \infty$?

SLLN: *Let $X_1, X_2, \ldots$ be i.i.d. with $EX_i = \mu$ and $EX_i^4 < \infty$. Then*

$$\frac{S_n}{n} = \frac{X_1 + \cdots X_n}{n} \xrightarrow{a.s.} \mu$$

Proof (with finite 4th moment)

1. Assume $\mu = 0$ and bound ES_n^4 by n^2
2. Use Chebyshev's to bound $P(|S_n| > n\varepsilon)$
3. Apply Borel-Cantelli to get $P(|S_n| > n\varepsilon \text{ i.o.}) = 0$, meaning that $S_n/n \to \mu$

Extensions/Generalizations:

1. Can weaken $EX_i^4 < \infty$ to just $E|X_i| < \infty$.
2. Can also weaken i.i.d. to pairwise independent and identically distributed.
3. Can actually extend to anywhere EX_i exists, that is when $EX_i^+ = \infty$ and $EX_i^- < \infty$ and i.i.d. (so $EX_i = \infty$) the result holds too.

Proof Ideas: Truncate X_i by $Y_i = X_i 1_{|X_i| \leq i}$ and show the result for the Y_i's. Do more clever bounding of the variances and then apply Chebyshev to bound difference from mean in terms of that variance.

What if $E|X_i| = \infty$?

SLLN fails in this case and the S_n/n does not converge to a finite value a.e.

$$\infty = E|X_i| = \int_0^\infty P(|X_1| > x)dx \leq \sum_0^\infty P(|X_1| > n)$$

so Second Borel Cantelli $\implies$ $P(|X_n| > n \text{ i.o.}) = 1$

take $C = \{\omega : \lim S_n/n \text{ exists}\}$ and intersect with $|X_n| > n$ i.o. to show that $P(C) = 0$.

34 Let $X_1, X_2, \ldots$ be i.i.d and non-negative with $EX_i = \infty$. What can we say about S_n/n?

Well the strong law of large numbers also holds here, so $S_n/n \to \infty$ a.s.

Can you prove it?

The method uses truncation, so let B be some bound and define $Y_n = X_n 1_{|X| < B}$.

Then $T_n = Y_1 + \cdots Y_n$, and because $Y_n \leq X_n$, $T_n \leq S_n$.

Claim, $EY_i \to EX_i = \infty$. This follows because Y_i are monotonic increasing to X_i so monotone convergence theorem applies.

Then strong law of large numbers holds for Y_n and T_n/n approaches the expected value, EY_n, a.s. but as $M \to \infty$ that expected value goes to ∞ so we have a lower bound for S_n/n that grows to ∞ a.s. hence $S_n/n \to \infty$ a.s. too.

2.5 Convergence of Random Series

35 Consider a sequence of i.i.d variables $X_1, X_2, \ldots$. How can we express $X_n \to 0$ a.s. in terms of a convergence of something in probability?

Let $M_n = \sup_{i>n} |X_i|$. Then $M_n \to 0$ in probability if and only if $X_n \to 0$ a.s.

Proof

$\Rightarrow$ If $M_n \to 0$ in probability, then for all $\varepsilon > 0$ we have $P(M_n > \varepsilon) = P(\sup_{i>n} |X_i| > \varepsilon) \to 0$

Take ω such that $\lim_{n\to\infty} X_n(\omega) \neq 0$, then there exists some $\varepsilon > 0$ such that infinitely often, $|X_n(\omega)| > \varepsilon$, hence $M_n(\omega) > \omega$ for all n.

Then $P(\lim_{n\to\infty} X_n \neq 0) \leq P(M_n > \varepsilon) \to 0$ so $X_n \to 0$ a.s.

$\Leftarrow$ If $X_n \to 0$ a.s. then a.e. $X_n \to 0$ as a sequence, so $\forall \varepsilon > 0$ there exists N such that $\forall n \geq N$ $|X_n| < \varepsilon$ and so $M_N = \sup_{n>N} |X_n| < \varepsilon$. So $P(M_n < \varepsilon) \to 1$ and conversely $P(M_n > \varepsilon) \to 0$.

36 What is Kolmogorov's 0-1 Law. What is the definition of a tail σ-algebra? What about tail random variables?

Given random variables $X_1, X_2, \ldots$ define $\mathcal{F}_n = \sigma(X_n, X_{n+1}, \ldots)$ and the tail σ field is $\cap_n \mathcal{F}_n$. In words, it is events that depend only on the tail of the X_i's, so changing a finite number will not affect the tail.

Kolmogorov's 0-1 Law If $X_1, X_2, \ldots$ are independent and $\mathcal{T}$ is their tail σ field, then for any $A \in \mathcal{T}$, $P(A) \in \{0, 1\}$.

The key idea of the proof is to show that A is independent from itself, so that $P(A) = P(\cap A) = P(A)^n \to 0, 1$.

A **tail random variable** would be one that is measurable with respect to the tail field. In this case, $P(Z \in B) = 0, 1$ for all B, so it must be the case that Z is actually constant.

37 State Kolmogorov's Maximal Inequality. How does it compare to Chebyshev's Inequality?

Kolmogorov's Maximal Inequality Let $X_1, X_2, \ldots$ be independent with $EX_i = 0$ and $\operatorname{var}(X_i) < \infty$ then

$$P\left(\max_{0\leq m\leq n} |S_m| \geq x\right) \leq x^{-2} \operatorname{var}(S_n)$$

In this setting, Chebyshev's Inequality only says that $P(|S_n| \geq x) \leq x^{-2} \operatorname{var}(S_n)$ but makes no claims about the partial sums along the way.

Chapter 3 - Central Limit Theorems

3.1 The De Moivre-Laplace Theorem

38 Give a concrete example of the Central Limit Theorem. How could you prove this directly?

The De Moivre-Laplace Theorem.

$X_1, X_2, \ldots$ i.i.d with $P(X_i = 1) = P(X_i = -1) = 1/2$ and $S_n = X_1 + \cdots X_n$ (e.g. betting \$1 on a coin flip, S_n =winnings after n tosses)

$$P(S_n/\sqrt{n} \leq b) \to \int_{-\infty}^{b} (2\pi)^{-1/2} e^{-x^2/2} dx$$

So th distribution functions for $S_n/\sqrt{n}$ converge to the distribution function for χ with normal distribution. So $S_n/\sqrt{n}$ converges weakly to χ.

Direct Proof Sketch.

Express $P(S_{2n} = 2k)$ in terms of factorials, then use Stirling's Formula to rewrite these. Careful choice of k with growth in terms of n, allows us to determine the asymptotic behavior which we show gives the integral desired.

3.2 Weak Convergence

39 What is weak convergence? How does it relate to convergence in probability and a.s. convergence?

Weak and a.s. Convergence

$X_n \Rightarrow X_\infty$ if and only if there exists $Y_i \overset{d}{=} X_i$ for $i \in \mathbb{N} \cup \{\infty\}$ such that $Y_n \to Y_\infty$ a.s.

Pf Sketch: Take $F_n \Rightarrow F_\infty$ dist from X and construct random variables directly from them $Y_n(x) = \sup\{y : F_n(y) < x\}$. Then show that, except at countable places, $Y_n(x) \to Y_\infty(x)$ so $Y_n \to Y_\infty$ a.s.

Weak and Convergence in P

$X_n \to X_\infty$ in P implies $X_n \Rightarrow X_\infty$

Pf Sketch: (slightly tricky!)

Step 1: show $F(a - \varepsilon) \le \lim_n F_n(a) \le F(a + \varepsilon)$

Express $F_X(a) \le F_Y(a + \varepsilon) + P(|X - Y| \ge \varepsilon)$ and apply twice to $F_n(a)$ and $F(a - \varepsilon)$

Step 2: For continuity point a, take $\varepsilon \to 0$ so then $\lim_n F_n(a) \to F(a)$.

$X_n \Rightarrow C$ for a constant C, then $X_n \to C$ in probability

Pf Sketch: Find $F_C(x) = 1$ when $x \ge c$ and $F_C(x) = 0$ when $x < C$ (so continuous outside $x = C$). Since $F_n(x) \to F_C(x)$,

$$P(|X_n - C| \ge \varepsilon) = F_n(C - \varepsilon) + 1 - F_n(C + \varepsilon) \to F_C(C - \varepsilon) + 1 - F_C(C + \varepsilon) = 0 + 1 - 1 = 0$$

40 What is an example of a r.v. that converges weakly but not in probability?

Take $X_n : (0,1) \to (0,1)$ defined by $X_n(w) = \begin{cases} w & n = 2m \\ 1 - w & n = 2m + 1 \end{cases}$

Then $F_n(x) = x$ for all n, and so $X_n \Rightarrow X_1$ however $X_n \not\to Y$ for any Y because on $(0, 1/3)$, $X_{2m} \in (0, 1/3)$ but $X_{2m+1} \in (2/3, 1)$ and so for any r.v. Y, $|X_n - Y| > 1/6$ infinitely often, so $\liminf P(|X_n - Y| \ge \varepsilon) \ge 1/6$ and so does not converge to 0.

41 Why do we only get convergence at continuity points for weak convergence?

Take $X_n = X + 1/n$ then $F_n(y) = F(y - 1/n)$ so only converges to $F(y)$ if F is left continuous (i.e. continuous) at y.

3.3 Characteristic Functions

42 What is the significance of the inversion formula for characteristic functions?

The inversion formula takes a characteristic function and produces the distribution function of the corresponding random variable. It's importance is that there is a 1-1 mapping between $\varphi(t)$ and $F(y)$.

43 Give an example where the characteristic functions φ_n of X_n converges to $\varphi(t)$ discontinuous at $t = 0$. What is the limit of the distribution function of the X_n?

$\varphi_n(t) = e^{-nt^2/2}$. Then φ_n converge to $\varphi(t) = \begin{cases} 0 & t \neq 0 \\ 1 & t = 0 \end{cases}$ which is discontinuous at $t = 0$.

The distribution function limit $F(y) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi n}} e^{-y^2/2n} \to 1/2$ for all x but hard to show.

44 Suppose you have $X_1, X_2, \ldots$ and corresponding characteristic functions $\varphi_1, \varphi_2, \ldots$ converging point-wise to $\varphi(t)$. What can you say? What is tightness? How do continuity and tightness relate?

Initially, only that $\varphi(0) = 1$ since all the φ_n are char. fun.

If φ is continuous at 0 (for example φ is char fun for some Y), then the φ_n are tight, and φ is a characteristic function for some X.

This is because the decay of the measure (captured by tightness condition) is related to the behavior of φ at 0, specifically bounded by an integral that gets arbitrarily small if φ is continuous.

3.4 Central Limit Theorems

45 State the i.i.d. central limit theorem. Prove it (possibly with added assumptions)

Central Limit Theorem $X_1, X_2, \ldots$ i.i.d. R.V.s with $EX_i = \mu$ and $\text{var}(X_i) = \sigma^2 \in (0, \infty)$ then

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \Rightarrow \chi = \mathcal{N}(0,1)$$

Proof. Use characteristic functions. Assume $\mu = 0$ and let X_i have characteristic function $\varphi(t)$. Then S_n has characteristic function $\varphi(t)^n$ and dividing by $\sigma\sqrt{n}$ gives $\varphi(\frac{t}{\sigma\sqrt{n}})^n$.

Using a Taylor series approximation, $\varphi(t) = Ee^{itX} = 1 + itEX - \frac{t^2 EX^2}{2} + o(t^2) = 1 - \frac{t^2\sigma^2}{2} + o(t^2)$. So then

$$\varphi_{S_n/\sigma\sqrt{n}}(t) = \left(1 - \frac{t^2\sigma^2}{2n\sigma^2} + o(\frac{t^2}{n})\right)^n = \left(1 - \frac{t^2}{2n} + o(\frac{t^2}{n})\right)^n \to (1 + \frac{-t^2/2}{n})^n \to e^{-t^2/2} = \varphi_\chi(t).$$

46 The type of convergence in the Central Limit Theorem is the convergence in distribution. Why isn't the convergence almost sure?

Claim: $\frac{S_n}{\sqrt{n}}$ cannot converge in probability (and thus not almost surely).

Suppose that it did, then $\left|\frac{S_{2n}}{\sqrt{2n}} - \frac{S_n}{\sqrt{n}}\right| \to 0$ in probability. Let $Y_n = \frac{S_{2n}}{\sqrt{2n}} - \frac{S_n}{\sqrt{n}}$.

Then $Y_n \to 0$ in probability, but can split up Y_n as

$$Y_n = \frac{S_{2n}}{\sqrt{2n}} - \frac{S_n}{\sqrt{n}} = \frac{S_{2n} - S_n}{\sqrt{2n}} + \frac{S_n}{\sqrt{n}}(\frac{1}{\sqrt{2}} - 1)$$

and since $\frac{S_n}{\sqrt{n}} \Rightarrow \sigma\chi$, both pieces are independent and converge to a normal distribution with finite nonzero variance, so $Y_n \Rightarrow C\chi$ for some $C \neq 0$. This contradicts $Y_n \to 0$ in P (which implies $Y_n \Rightarrow 0$).

47 Suppose $X_1, X_2, \ldots$ are bounded but $\sum_n \text{var}(X_n) = \infty$, what can you say about the limiting behavior of S_n?

Up to suitable scaling this will still converge to $\chi = \mathcal{N}(0,1)$.

Lindeberg-Feller CLT:
Triangular array $X_{n,m}$ independent variables for $1 \leq m \leq n$ where $EX_{n,m} = 0$ and

(i) $\sum_{m=1}^n EX_{n,m}^2 \to \sigma^2 > 0$ as $n \to \infty$

(ii) $\forall \varepsilon > 0 \sum_{m=1}^n E(|X_{n,m}|^2; |X_{n,m}| > \varepsilon) \to 0$ as $n \to \infty$

$$S_n = X_{n,1} + \cdots + X_{n,n} \Rightarrow \sigma\chi$$

So center the X_i's so that they have mean 0. Then we want to scale by something so that $\sum_{m=1}^n EX_{n,m}^2$ converges to a positive finite value. If $X_{n,m} = X_m/c_n$ then this sum will be $\frac{1}{c_n^2}\sum_{m=1}^n var(X_i)$ so letting $c_n = \sqrt{\text{var}(S_n)}$ this is exactly 1 for all n.

Now for (ii), since the X_m are bounded, and the denominator goes to ∞, at some point $|X_{n,m} < \varepsilon$ so the sum is 0.

Applying L-F CLT,

$$X_{n,1} + \cdots + X_{n,m} = \frac{S_n}{c_n} = \frac{S_n}{\sqrt{\text{var}(S_n)}} \Rightarrow \chi$$

and if we add back in the means, we have $\frac{S_n - ES_n}{\sqrt{\text{var}(S_n)}} \Rightarrow \chi$.

48 What is the Lindeberg-Feller Central Limit Theorem? How does it relate to the i.i.d. Central Limit Theorem?

Lindeberg-Feller CLT:

Triangular array $X_{n,m}$ independent variables for $1 \le m \le n$ where $EX_{n,m} = 0$ and

(i) $\sum_{m=1}^n EX_{n,m}^2 \to \sigma^2 > 0$ as $n \to \infty$

(ii) $\forall \varepsilon > 0 \sum_{m=1}^n E(|X_{n,m}|^2; |X_{n,m}| > \varepsilon) \to 0$ as $n \to \infty$

$$S_n = X_{n,1} + \cdots + X_{n,n} \Rightarrow \sigma\chi$$

This is a generalization of the i.i.d. CLT, because we can take $X_{n,m} = \frac{X_m}{\sqrt{n}}$ then $X_{n,1} + \cdots + X_{n,n} = \frac{S_n}{\sqrt{n}} \Rightarrow \chi$.

Condition (i) follows directly from computation, and (ii) converges to zero by dominated convergence theorem and the fact that $P(|X_m| > \varepsilon\sqrt{n}) \to 0$ by applying Chebyshev's Inequality and using the finite variance of X_m.

3.6 Poisson Convergence

49 What is Poisson Convergence and why is it called the "law of rare events"?

Poisson Convergence (or Poisson Limit Theorem) gives an analogous limit theorem to CLT but for a triangular array of Bernoulli variables with a Poisson distribution as the limiting behavior.

This is called the law of rare events because the probability of the events ($X_{n,m} = 1$) must get small as n grows, so the sum is counting the number of rare occurrences.

3.10 Limit Theorems in $\mathbb{R}^d$

50 State and prove the Central Limit Theorem. Can you state a version of the Central Limit Theorem for random vectors?

Central Limit Theorem $X_1, X_2, \ldots$ i.i.d. R.V.s with $EX_i = \mu$ and $\text{var}(X_i) = \sigma^2 \in (0, \infty)$ then

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \Rightarrow \chi = \mathcal{N}(0,1)$$

Proof. Use characteristic functions. Assume $\mu = 0$ and let X_i have characteristic function $\varphi(t)$. Then S_n has characteristic function $\varphi(t)^n$ and dividing by $\sigma\sqrt{n}$ gives $\varphi(\frac{t}{\sigma\sqrt{n}})^n$.

Using a Taylor series approximation, $\varphi(t) = Ee^{itX} = 1 + itEX - \frac{t^2EX^2}{2} + o(t^2) = 1 - \frac{t^2\sigma^2}{2} + o(t^2)$. So then

$$\varphi_{S_n/\sigma\sqrt{n}}(t) = \left(1 - \tfrac{t^2\sigma^2}{2n\sigma^2} + o(\tfrac{t^2}{n})\right)^n = \left(1 - \tfrac{t^2}{2n} + o(\tfrac{t^2}{n})\right)^n \to (1 + \tfrac{-t^2/2}{n})^n \to e^{-t^2/2} = \varphi_{\chi}(t).$$

Central Limit Theorem in $\mathbb{R}^d$ $X_1, X_2, \ldots$ i.i.d random vectors with $EX_i = \mu \in \mathbb{R}^d$ with finite covariance $\Gamma_{i,j} = E(X_i - \mu_i)(X_j - \mu_j)$,

$$\frac{S_n - n\mu}{\sqrt{n}} \Rightarrow \mathcal{N}_d(0, \Gamma)$$

where $\mathcal{N}_d(0, \Gamma)$ is the multivariate Gaussian with mean 0 and covariance Γ.

Chapter 4 - Martingales

4.1 Conditional Expectation

51 What are regular condition probabilities? Why are they useful?

A regular conditional distribution of X with respect to $\mathcal{F}$ where $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ is a function $\mu : \Omega \times \mathcal{S} \to [0, 1]$ such that

i. Fixing $\omega \in \Omega$, $A \in \mathcal{S} \mapsto \mu(\omega, A)$ is a probability measure on $\mathcal{S}$

ii. Fixing $A \in \S$, $\omega \in \Omega \mapsto \mu(\omega, A)$ is a version for $E(X \in A \mid \mathcal{F})$.

If X is the identity on Ω then $\mu(\omega, A)$ is a version for $E(1_A \mid \mathcal{F}) = P(A \mid \mathcal{F})$ and μ is a regular conditional probability.

These are useful because they give a way of computing conditional expectations on $f(X)$, via $E(f(X) \mid \mathcal{F}) = \int f(x)\mu(\omega, dx)$.

4.2 Martingales, Almost Sure Convergence

52 What are some examples of martingales? Submartingales?

symmetric random walk on $\mathbb{Z}$ is martingale, if non-symmetric could be sub/super martingale.

Betting on a fair casino game ... earnings at time n is martingale. Unfair game gives a supermartingale.

53 What is the upcrossing inequality? Why is it useful? How is it proved?

Upcrossing Inequality If X_n is submartingale, $a < b$, and U_n denotes the number of complete upcrossings (going from $X_n \leq a$ to $x_n \geq b$) by time n, then

$$(b - a)EU_n \leq E(X_n - a)^+ - E(X_0 - a)^+.$$

Proof. First we switch to $Y_n = (X_n - a)^+$ so that we do not incur losses from incomplete upcrossings but maintain all the upcrossing behavior of X_n. Define the predictable sequence (betting strategy) H_n to be 1 when between occurences of $Y_n \leq a$ and the next time $Y_n \geq b$ (i.e. only bets during upcrossing phases). Then $(b - a)U_n \leq (H \cdot Y)_n$.

Now take the complement strategy $(1-H)_n$ and since X_n is submartingale, so is Y_n and so $(1-H\cdot Y)_n$ is submartingale too, meaning $E(1-H\cdot Y)_n \geq E(1-h\cdot Y)_0 = 0$. It also satisfies $(1-H\cdot Y)_n + (H\cdot Y)_n = Y_n - Y_0$.

$$(b-a)EU_n \leq E(H\cdot Y)_n \leq E(H\cdot Y)_n + E(1-H\cdot Y)_n = E(Y_n - Y_0) = E(X_n-a)^+ - E(X_0-a)^+.$$

Application: The upcrossing inequality is primarily useful in proving the Martingale A.S. convergence theorem.

54 Give the definition of martingales and state the a.s. convergence theorem for them. How can you prove this convergence theorem?

Martingale A sequence of random variables X_n adapted to a filtration $\mathcal{F}_n$ such that $E|X_n| < \infty$ for all n and $E(X_{n+1} \mid \mathcal{F}_n) = X_n$ for all $n \geq 0$.

If $E(X_{n+1} \mid \mathcal{F}_n) \geq X_n$ then X_n is submartingale. If $E(X_{n+1} \mid \mathcal{F}_n) \leq X_n$ then X_n is supermartingale.

A.S. Convergence Theorem
If X_n is submartingale and $\sup_n EX_n^+ < \infty$ then $X_n \to X$ a.s. and $E|X| < \infty$.

Proof. First, we use apply the upcrossing inequality, noting that $E(X_n-a)^+ \leq EX_n^+ + |a|$, so then

$$(b-a)EU_n \leq E(X_n-a)^+ - E(X_0-a)^+ \leq EX_n^+ + |a| \leq \sup_n EX_n^+ + |a| < \infty$$

so we see that $EU_n < \infty$ for all $a < b$. As $n \to \infty$, $EU_n \uparrow EU$ so by monotone convergence we have that $EU < \infty$ also. This means that $U < \infty$ a.s..

On the set where $U < \infty$ we will show that $\lim_n X_n$ exists. Suppose that $\liminf_n X_n < \limsup_n X_n$ then choose $a < b$ between them. Since $\liminf_n X_n < a < b < \limsup_n X_n$ there would be infinitely many upcrossings to maintain support for $\liminf, \limsup$ so $U \not< \infty$ but this happens on a measure zero set, so $\liminf_n X_n = \limsup_n X_n = \lim_n X_n$ a.s. and let $X = \lim_n X_n$.

Now to conclude that $E|X| < \infty$ we apply Fatou's lemma to each piece, namely $E|X| = EX^+ + EX^-$. By Fatou's, $\liminf_n EX_n^+ \geq EX^+$ and

$$\liminf_n EX_n^+ < \sup_n EX_n^+ < \infty$$

so $EX^+ < \infty$. Now for the negation, $X_n^- = X_n^+ - X_n$, and applying Fatou's

$$EX^- \leq \liminf_n EX_n^- \leq \liminf_n EX_n^+ < \sup_n EX_n^+ < \infty.$$

55 Can you give an example of a martingale that converges a.s. but not in L^1? What condition can we put on the sequence to prevent this?

Let S_n be a symmetric random walk on $\mathbb{Z}$ starting at 1. Let N be the stopping time $N = \inf\{n : S_n = 0\}$. Define $X_n = S_{N\wedge n}$. This will converge a.s. to 0. First, taking $-X_n \leq 0$ which is also martingale, we have $\sup_n(-X_n)^+ = 0 < \infty$ so $X_n \to X_\infty$ a.s.. If $X_\infty \neq 0$ then taking $\varepsilon = 1/2$, we see that if $|X_n - X_\infty| < \varepsilon$ then $|X_{n+1} - X_\infty| > \varepsilon$ so it does not converge, implying that $X_\infty = 0$.

However $E|X_n - 0| = E|S_{N\wedge n}| = E|S_{N\wedge 0}| = E|S_0| = 1$ for all n, so convergence does not happen in L^1.

If UI then a.s. convergence gives L^1 convergence and vice versa.

56 Let X_n be a martingale with respect to $\mathcal{F}_n$, and suppose $E(X_n^2) \leq B < \infty$ for all n. What can you conclude about X_n?

Hint: can we apply a convergence theorem to it?

$(EX_n^+)^2 \leq E(X_n^+)^2 \leq EX_n^2 \leq B$. So $\sup_n EX_n^+ \leq \sqrt{B} < \infty$ so we can apply a.s. convergence to X_n.

57 Let X_n be a submartingale, H_n be a predictable sequence, and N a stopping time. Show that $(H \cdot X)_n$ and $X_{N\wedge n}$ are both submartingale as well.

$$\begin{aligned} E((H\cdot X)_n \mid \mathcal{F}_{n-1}) &= E\left(\sum_{m=1}^{n} H_m(X_m - X_{m-1}) \mid \mathcal{F}_{n-1}\right) = \sum_{m=1}^{n} E\left(H_m(X_m - X_{m-1}) \mid \mathcal{F}_{n-1}\right) \\ &= E(H_n(X_n - X_{n-1}) \mid \mathcal{F}_{n-1}) + \sum_{m=1}^{n-1} H_m(X_m - X_{m-1}) \\ &= H_n E(X_n \mid \mathcal{F}_{n-1}) - H_n X_{n-1} + \sum_{m=1}^{n-1} H_m(X_m - X_{m-1}) \\ &= \sum_{m=1}^{n-1} H_m(X_m - X_{m-1}) = (H\cdot X)_{n-1} \end{aligned}$$

Now define the predictable sequence $H_n = 1_{N>n}$. Then $(H\cdot X)_n$ is submartingale, and evaluates to $X_{N\wedge n} - X_0$. Adding back X_0 gives that $X_{N\wedge n}$ is submartigale as well.

4.3 Examples

58 Second Borel-Cantelli Lemma Version II

Thm.$\mathcal{F}_n$ *a filtration (with $\mathcal{F}_0 = \{\emptyset, \Omega\}$), and B_n, $n \geq 1$ sequence of events with $B_n \in \mathcal{F}_n$.*

$$\{B_n \text{ i.o }\} = \left\{\sum_{n=1}^{\infty} P(B_n|\mathcal{F}_{n-1}) = \infty\right\}$$

Pf Sketch: Take $X_n = \sum_{m=1}^{n} 1_{B_n}$, then this is submartingale. Apply Doob's decomposition to get $X_n = M_n + A_n$ where the martingale will be bounded. Using bounded increments, consider how the limit of X_n (sum of indicators) relates to the limit in M_n, in each of the cases given by bounded increments.

4.4 Doob's Inequality, Convergence in L^p, $p > 1$

59 When can we conclude L^p convergence for martingales? How is it proved? Is there a similar proof for L^1?

L^p **Convergence Theorem** If X_n is submartingale and $\sup_n E|X_n|^p < \infty$ (for some $p > 1$) then X_n converges a.s. and in L^p to some X.

Proof. First we show a.s. convergence. Observe that $(X_n^+)^p \leq |X_n|^p$ so $\sup EX_n^+ \leq \sqrt[p]{\sup_n E|X_n|^p} < \infty$ so the a.s. convergence lemma gives that $X_n \to X$ a.s. and $E|X| < \infty$.

Now to show that this convergence happens in L^p we need to apply an integral convergence theorem to $E|X_n - X|^p$ since $|X_n - X|^p \to |X - X|^p = 0$ since $X_n \to X$ a.s.. First we apply L^p maximal inequality to get

$$E\left(\max_{0\leq m\leq n} |X_m|^p\right) \leq \left(\frac{p}{p-1}\right)^p E|X_n|^p$$

and taking $n \to \infty$ we have

$$E\left((\sup_n |X_m|)^p\right) = E\left(\sup_n |X_m|^p\right) \leq \left(\frac{p}{p-1}\right)^p \sup_n E|X_n|^p < \infty.$$

Now we can use $(\sup_n |X_m|)^p$ to apply dominated convergence since

$$|X_n - X| \leq |X_n| + |X| \leq \sup_n |X_n| + \lim |X_n| \leq 2 \sup_n |X_n|$$

so $|X_n - X|^p \leq 2^p (\sup_n |X_n|)^p$ and since p is fixed, 2^p is a fixed finite constant, making $2^p(\sup_n |X_n|)^p$ integrable by our inequality above. Then dominated convergence applies to $E|X_n - X|^p \to 0$ so our convergence happens in L^p.

L^1 **convergence** There is no similar proof for L^1 convergence because there is no L^1 maximal inequality. Instead L^1 convergence is proved using uniform integrability.

60 How does Doob's Inequality relate to Kolmogorov's Maximal Inequality?

Doob's Inequality For X_n submartingale and λ some constant, then

$$P\left(\max_{0 \leq m \leq n} X_n^+ \geq \lambda\right) \leq \lambda^{-1} E(X_n^+).$$

Kolmogorov's Inequality X_n independent random variables with $EX_n = 0$ and $\text{var}(X_i) < \infty$. Then $S_n = X_1 + \cdots + X_n$ and taking x, satisfies

$$P\left(\max_{0 \leq m \leq n} |S_m| \geq x\right) \leq x^{-1} \text{var}(S_n).$$

Claim: Doob's Inequality $\implies$ Kolomogorov's Inequality
Since $EX_n = 0$, S_n from Kolmo. is a submartingale, and since $x \mapsto x^2$ is convex and $E|S_n^2| = \sum \text{var}(X_i) < \infty$, S_n^2 is submartingale as well. Applying Doob's to this with $\lambda = x^2$ gives the desired result.

4.6 Uniform Integrability, Convergence in L^1

61 Show that if X_n is uniformly integrable and N is a stopping time that $X_{N \wedge n}$ is also uniformly integrable.

First we break up $E(|X_{N\wedge n}|; |X_{N \wedge n}| > M) = E(|X_N|; |X_N| > M, N \leq n) + E(|X_n|; |X_n| > M; n < N)$.
We will show that $E|X_N| < \infty$ to show that the first piece goes to 0 as $M \to \infty$. For this, we want to show that X_n u.i. implies $X_{N \wedge n} \to X_N$ a.s. by the a.s. convergence theorem which also tells us that $E|X_N| < \infty$. To see this, choose $M >> 0$ such that $\sup_n E(|X_n|; |X_n| > M) \leq 1$. Then $\sup E|X_n| \leq M + 1 < \infty$. Since $X_{N\wedge n}$ is submartingale, $E|X_{N \wedge n}| \leq E|X_n|$ for all n, so

$$\sup_n EX_{N \wedge n}^+ \leq \sup_n E|X_{N \wedge n}| \leq \sup_n E|X_n| < \infty$$

so a.s. convergence tells us that $X_{N \wedge n} \to X$ a.s.. And since $N \wedge n \uparrow N$ as $n \to \infty$, we have that $X = X_N$, thus $E|X_N| < \infty$.

Now the second piece goes to 0 as $M \to \infty$ because X_n is U.I. and the added $n < N$ condition only shrinks the expectation. Hence the combined sum goes to 0 making $X_{N \wedge n}$ U.I.

4.7 Backwards Martingales

62 What are backwards martingales? What can you say about their convergence?

Backwards Martingales: X_{-n} indexed by $n = 1, 2, 3, \ldots$ and adapted to the filtration $\mathcal{F}_{-n}$, (i.e. $\cdots \subseteq \mathcal{F}_{-2} \subseteq \mathcal{F}_{-1} \subseteq \mathcal{F}_0$) such that $E(X_{-n+1} \mid \mathcal{F}_{-n}) = X_{-n}$ (i.e. $E(X_0 \mid \mathcal{F}_{-1}) = X_{-1}$)

Given the direction of the filtration, these have nicer convergence theorems.

Backwards Convergence Theorem If X_n is a backwards martingale, it converges a.s. and in L^1.

Proof (using U.I. convergence). $X_n = E(X_0 \mid \mathcal{F}_{-n})$ and since $E|X_0| < \infty$, this is a UI collection, and since X_n are submartingale and U.I. they converge a.s. and in L^1.

Proof (using A.S. conv). Applying the upcrossing inequality to $X_{-n}, \ldots, X_0$ we have $(b-a)EU_n \leq E(X_0 - a)^+ - E(X_{-n} - a)^+ \leq E(X_0 - a)^+$ so this is finite meaning $EU < \infty$ so $U < \infty$ a.s. which just as for submartingales implies that the limit exists a.s..

Now use U.I to lift this to L^1 convergence. Define the cut-off function $\varphi_M(x)$ to be x where $x \in [-M, M]$ but stay at $\pm M$ when it goes beyond, then

$$E|X_n - X| \leq E|X_n - \varphi_M(X_n)| + E|\varphi_M(X_n) - \varphi(X)| + E|\varphi_M(X) - X|$$

We show each piece goes to 0 as $M \to \infty$.

First, $E|X_n - \varphi_M(X_n)| = E(|X_n - \varphi_M(X_n)|; |X_n| > M) \leq E(|X_n; |X_n| > M) \to 0$ since X_n is U.I.

Second, $E|\varphi_M(X_n) - \varphi(X)| \to 0$ since $X_n \to X$ a.s. and φ_M is a bounded and continuous function.

Third, $E|\varphi_M(X) - X| = E(|X - \varphi_M(X)|; |X| > M) \leq E(|X|; |X| > M) \to 0$ since $E|X| < \infty$.

63 Use reverse martingales to derive the SLLN. Hint: let $\mathcal{F}_{-n} = \sigma(S_n, X_{n+1}, X_{n+2}, \ldots)$.

SLLN $X_1, X_2, \ldots$ iid with $EX_i = \mu$ and $E|X_i| < \infty$. Then $S_n/n \to \mu$ a.s.

Let $Y_{-n} = \frac{S_n}{n}$. First we want to show that this gives a backwards martingale with respect to the filtration given. Then we show that $Y_{-\infty} = \mu$.

Since $Y_{-n} = S_n/n \in \sigma(S_n) \subseteq \mathcal{F}_{-n}$ so it is adapted to $\mathcal{F}_{-n}$. Since $E|X_i| < \infty$, we have $E|Y_{-n}| < \infty$. Finally,

$$E(Y_{-n+1} \mid \mathcal{F}_{-n}) = E\left(\frac{X_1 + \cdots X_{n-1}}{n-1} \mid \sigma(S_n, X_{n+1}, \ldots)\right)$$

With respect to $\sigma(S_n)$ all the $X_1, \ldots, X_{n-1}$ are interchangeable and have expectation $EX_i = S_n/n$. So this gives $\frac{(n-1)\frac{S_n}{n}}{n-1} = \frac{S_n}{n} = Y_{-n}$ so Y_{-n} is a backwards martingale.

By the backwards martingale convergence we have $\frac{S_n}{n} = Y_{-n} \to Y_{-\infty}$ a.s.. And $Y_{-\infty} = E(Y_{-1} \mid \mathcal{F}_{-\infty}) = E(X_1 \mid \cap_n \mathcal{F}_{-n})$. Since $\cap_n \mathcal{F}_{-n}$ is the exchangable algebra for $X_1, \ldots$ $Y_{-\infty}$ is trivial and thus constant with expectation $E(X_1) = \mu$ so $S_n/n \to \mu$ a.s.

4.8 Optional Stopping Theorems

64 What is optional stopping? What are some conditions under which it holds?

If X_n is a submartingale and N is a stopping time then $X_{N \wedge n}$ is submartingale so $EX_{N \wedge 0} \leq EX_{N \wedge n}$ for all n, but this can only be strengthened to $EX_0 \leq EX_N$ when *optional stopping holds*. Rephrasing, we want to know when $EX_{N \wedge n} \to EX_N$, for example when we could apply integral convergence theorems to it.

Some cases when this holds:

1. Whenever usual integral convergence theorems hold for $EX_{N \wedge n}$.

2. If N is bounded a.s., that is $P(N \leq k) = 1$ for some k. Then $EX_0 \leq EX_N \leq EX_k$.

   [Pf.] $X_{N \wedge n}$ submartingale, so $EX_0 = EX_{N \wedge 0} \leq EX_{N \wedge k} = EX_N$. Then $K_n = 1_{N > n}$ predictable, so $EX_k - EX_N = (K \cdot X)_k \geq 0$

3. X_n U.I. then $EX_0 \le EX_N \le EX_\infty$.

   X_n U.I. $\implies X_{N\wedge n}$ U.I. too. $\implies X_{N\wedge n} \to X_N$ a.s. and in L^1.
   $EX_N - EX_0 \le E|X_N - X_{N\wedge n}| + EX_{N\wedge n} - E_X0 \to EX_{N\wedge n} - E_X0 \ge 0$

4. $E(|X_{n+1} - X_n|\mathcal{F}_n) \le B < \infty$ a.s. and $EN < \infty$

   Show that $X_{N\wedge n}$ is U.I. by bounding by integrable r.v. in terms of sum of increments.

65 Let $\zeta_1, \zeta_2, \ldots$, be i.i.d. with $E\zeta_i = \mu$, N a stopping time with finite expectation. Show that $ES_N = \mu EN$.

Here is an example where we can get optional stopping by appealing to traditional methods.

Let $X_n = S_{N\wedge n} - \mu(N \wedge n)$. Since this is martingale, $EX_0 = EX_n$ for all n meaning $ES_{N\wedge n} = \mu E(N \wedge n)$. On the right hand side, $0 \le N \wedge n \uparrow N$ and N has finite expectation so $E(N \wedge n) \to EN$.

The left hand side is a little more careful. First, we argue a reduction to the case that $\zeta_i \ge 0$. If not, let $\zeta_i = \zeta_i^+ - \zeta_i^-$. Since the ζ_i are iid then so too are the $\zeta_i^\pm$. Now $ES_{N\wedge n} = ES^+_{N\wedge n} - ES^-_{N\wedge n}$ so if can show each piece converges in an optional stopping way this suffices. Now if $\zeta_i \ge 0$ then $S_{N\wedge n} \ge 0$ and is martingale, so we can apply the a.s. convergence (take $-S \le 0$ to bound sup) and $S_{N\wedge n} \to S_N$ a.s.. Now with a.s. convergence we have convergence of the expectations so $ES_{N\wedge n} \to ES_N$ as desired.

66 Apply optional stopping to get a formula for the probability that the simple symmetric random walk on $\mathbb{Z}$, started at 0, hits some $-a$ before b? What is the expected time it takes for either of these to happen?

Let S_n be our simple symmetric random walk. Define $N = \inf\{n : S_n = -a \text{ or } S_n = b\}$. If we can apply optional stopping to $S_{N\wedge n} - N \wedge n$ then $ES_N = ES_0 = 0$ and we can directly compute ES_N in terms of $P(S_N = -a)$.

We obtain optional stopping by showing that $EN < \infty$ and the conditional expectation of the increments is bounded. The latter is easy since $E(|S_{n+1} - S_n| \mid \mathcal{F}_n) = E(1 \mid \mathcal{F}_n) \le 1 < \infty$. Now to bound N, observe that in $a+b$ steps if every step is in the same direction we will trigger N, so if $P(N > a+b)$ then we can't take all steps in the same direction so $P(N > a + b) < 1 - 2^{a+b}$. If we take $m(a + b)$ steps we can repeat this calculation m times to get $P(N > m(a + b)) < (1 - 2^{a+b})^m$. Since this is a gemetric series that has a finite sum, $EN < \infty$.

Now applying optional stopping

$$0 = ES_N = -aP(S_N = -a) + b(1 - P(S_N = -a)) \implies P(S_N = -a) = \frac{b}{a+b}.$$

Now to compute EN we want to look at $X_n = S_n^2 - n$ which is martingale ($\sigma = 1$). Applying bounded optional stopping, $E(S^2_{N\wedge n} - N \wedge n) = EX_0 = 0$ so $ES^2_{N\wedge n} = E(N \wedge n)$. The right hand side converges to EN by monotone convergence. The left hand side is a martingale that is bounded below by 0 so taking $-S^2_{N\wedge n}$ gives a.s. convergence to S^2_N. Hence $EN = ES^2_N = a^2P(a) + b^2P(b) = \frac{a^2b}{a+b} + \frac{b^2a}{a+b} = ab$.

Chapter 5 - Markov Chains

5.1 Examples

67 Give an example of a Markov chain. What about a markov chain that is also martingale?

A markov chain is "memory-less" sequence of random variables, so a random walk is an example, betting in a casino is an example. Branching processes are another.

Random walks that are symmetric are also martingale since the probabilities balance out. Betting in a casino on a fair game would be martingale.

68 Give the transition probabilities for Ehrenfest chain.

The Ehrenfest chain is the markov chain described by partitioning a box into two sides and distributing m marbles between the two sides. At each step, randomly select a marble from the box as a whole (with equal prob for all marbles) and move that marble to the other side. Let X_n be the fraction of marbles on the left side (or right by symmetry but fix one).

Then if $X_n = r$ the probability that $X_n = r+1$ is $p(r, r+1) = \frac{m-r}{m}$ and $X_n = r-1$ is $p(r, r-1) = \frac{r}{n}$ and $p(k, l) = 0$ for all other pairs.

5.2 Construction, Markov Properties

69 What is the Markov Property? The Strong Markov Property?

The markov property is that the probability of X_n depends only on X_{n-1}. Formally, for any event A and $\mathcal{F}_{n-1} = \sigma(X_1, \ldots, X_{n-1})$, $P(X_n \in A \mid \mathcal{F}_{n-1}) = P(X_n \in A \mid X_{n-1})$.

The strong markov property generalizes this to stopping times. So if N is a stopping time and $\mathcal{F}_N = \{A : A \cap \{N = n\} \in \mathcal{F}_n\}$ then $P(X_{N+1} \in A \mid \mathcal{F}_N) = P(X_{N+1} \mid X_N)$.

Both properties extend beyond the 'next' state so that $P(X_{n+t} \in A \mid \mathcal{F}_{n-1}) = P(X_{n+t} \in A \mid X_{n-1})$.

5.3 Recurrence and Transience

70 Define recurrence and transience. What does "recurrence is contagious" mean? Prove it.

Recurrence is a property of a state the $\rho_{xx} = P_x(T_x < \infty) = 1$, i.e. the probability of starting at x and returning to x in finite time is 1.

Claim: If x, y are in an irreducible set (or just $\rho_{xy} > 0$) and x is recurrent, then y is recurrent also.

Proof. Let $N(y)$ be the number of visits to y, we show that $E_y N(y) = \infty$ and use this to show that $\rho_{yy} = 1$.

First, $E_y N(y) = \sum_{n=0}^{\infty} p^n(y, y) \geq \sum_{k=0}^{\infty} p^a(y, x) p^k(x, x) p^b(x, y) = p^a(y, x) p^b(x, y) \sum_{k=0}^{\infty} p^k(x, x) = p^a(y, x) p^b(x, y) E_x N(x)$.

Next we show that $E_z N(z) = \infty$ iff $\rho_{zz} = 1$ which follows immediately from the following:

$$E_z N(z) = \sum_{n=0}^{\infty} P_z(N(z) \geq n) = \sum_{n=0}^{\infty} \rho_{zz}^n = \frac{1}{1 - \rho_{zz}}.$$

Now if $\rho_{xx} = 1$, then $E_x N(x) = \infty$ and since x, y are in an irreducible set we can find a, b such that $p^a(y, x), p^b(x, y) > 0$ so then $E_y N(y) = \infty$ meaning $\rho_{yy} = 1$ and y is reccurent as well.

71 Give a decomposition for the set of recurrent states in a markov chain.

Let R be the collection of all reccurent states in a markov chain. We will show that R can be decomposed into disjoint closed and irreducible collections.

Let $x \in R$ and define $C_x = \{y : \rho_{xy} > 0\}$. We will show that $\rho_{xy} > 0$ is an equivalence relation on reccurent states. First, $\rho_{xx} = 1 > 0$ so it is reflexive. Second, if $\rho_{xy} > 0$ then since x is recurrent it must be able to get back so that $\rho_{yx} > 0$ too. Finally, if $\rho_{xy}, \rho_{yz} > 0$ then there exists some n, m such that $p^n(x, y) > 0$ and $p^m(y, z) > 0$ so then $p^{n+m}(x, z) \geq p^n(x, y) p^m(y, z) > 0$ making $\rho_{xz} > 0$. Thus we can partition R into these sets.

Now we show that C_x is irreducible and closed. If $y, z \in C_x$ then $\rho_{xy}, \rho_{xz} > 0$ but by symmetry and transitivity, this means $\rho_{yz} > 0$ so this is irreducible. Now if $y \in C_x$ and $\rho_{yz} > 0$ then by transitivity, $\rho_{xz} > 0$ so $z \in C_x$ meaning it is closed.

5.5 Stationary Measures

72 Let p have a stationary measure, can you say anything recurrent states? What can you add to say something about recurrent states?

Initially not much, might not even have one for example symmetric random walk of $\mathbb{Z}$ all states are transient but the uniform measure is a stationary measure.

If the stationary measure can be scaled to be a distribution then on an irreducible subset taking positive probabilities all states are positive recurrent.

5.6 Asymptotic Behavior

73 Give an example of a periodic markov chain and explain why it cannot converge.

The Ehrenfest chain has period 2, since the parity of X_n changes at each step, meaning that $p^n(x,x) = 0$ when n is odd. This chain cannot converge for exactly this reason, if π were a stationary measure, and x is any state with positive mass (which will be recurrent), $|p^n(x,x) - \pi(x)|$ will infinitely often take the value $|\pi(x)| > 0$ so cannot converge to zero.

74 State and prove a convergence theorem for markov chains.

Convergence: If p is a markov chain that is irreducible and aperiodic with a stationary measure π, then $p^n \to \pi$ (convergence of measures).

Proof. We proceed by constructing a paired markov chain $X \times Y$ both with the same transition probability p but X starting at some state x and Y starting with the initial stationary distribution π.

Define the transition probability $\bar{p}((a,b),(c,d)) = p(a,c)p(b,d)$. We first show that this is irreducible. Take any two states (a,b) and (c,d). We know that p is irreducible so for some n, m we have $p^n(a,c) > 0$ and $p^m(b,d) > 0$. Furthermore, since p is aperiodic, for some M for every $\ell \geq M$, $p^\ell(y,y) > 0$ for $y \in \{a,b,c,d\}$. Then

$$\bar{p}^{M+n+m}((a,b),(c,d)) = p^{M+n+m}(a,c)p^{M+n+m}(b,d) \geq p^n(a,c)p^{M+m}(c,c)p^m(b,d)p^{M+n}(d,d) > 0$$

so $\bar{p}$ is irreducible.

Using the definitions of stationary distributions, we see that $\bar{\pi} = \pi\pi$ is a stationary on $\bar{p}$. Since $\bar{p}$ is irreducible, this means every state is positive recurrent, and thus recurrent. Let (y,y) be some state in the diagonal, we will show that $T_{(y,y)}$ the stopping time when $X \times Y = (y,y)$ is a.s. finite. Positive recurrent means that $E_{(y,y)}T_{(y,y)} < \infty$ and since we are irreducible the expected value from any starting position is still finite, meaning $T_{(y,y)}$ is a.s. finite too.

Now let T be the stopping time of hitting the diagonal where $X_n = Y_n$. Then $T < T_{(y,y)} < \infty$ a.s.. So in particular $P(T > n) \to 0$ as $n \to \infty$.

Now using the strong markov property,

$$P(X_n = y) = P(X_n = y, T \leq n) + P(X_n = y, T > n) \leq P(Y_n = y) + P(X_n = y, T > n)$$

and the same holds switching X_n and Y_n so that $|P(X_n = y) - P(Y_n = y)| \leq P(X_n = y, T > n) + P(Y_n = y, T > n)$. Summing over all states y we have

$$\sum_y |P(X_n = y) - P(Y_n = y)| = \sum_y |p^n(x,y) - \pi(y)| \leq 2P(T > n) \to 0 \text{ as } n \to \infty.$$