Probability Theory

Marcel Nutz

Transcribed by Ron Wu

This is an advanced undergraduate course. Offered in Spring 2014 at Columbia University. Recommended Texts: Jacod and Protter, *Probability Essentials*; Resnick, *A Probability Path*; Stirzaker, *Elementary Probability.* Office hours: TuTh 9:55-10:45

Contents

1 Measure Theory Part I

Lecture 1 (1/21/14)

We shall introduce a probability space $(\Omega, \mathcal{F}, P)$. Let $\Omega \neq \varnothing$, called the sample space or possible outcomes of an experiment. E.g. 2 coins tosses

$$\Omega = \{0,1\}^2 = \{(0,0),(0,1),(1,0),(1,1)\}$$

This is a finite set. For Ω infinite, it was not so clear how to make sense of it, in particular it got to be additive and normalizable. This was the struggle for the early development of the probability theory. It was Kolmogorov who formally used the measure theory to make the subject rigorous.

1.1 σ Field and Measure

Definition 1. A collection $\mathcal{F}$ of subsets and Ω is a σ field (or σ algebra) if

1) $\varnothing \in \mathcal{F}$

2) $A \in \mathcal{F} \implies A^c \in \mathcal{F}$

3) $A_1, A_2 ... \in \mathcal{F} \implies \bigcup_{n\geq 1} A_n \in \mathcal{F}$

$(\Omega, \mathcal{F})$ is called a measurable space. An $A \in \mathcal{F}$ is called an event. For now assume $\mathcal{F}$ is all subsets of Ω, but soon we need to restrict it to make it measurable.

Definition 2. A function $P : \mathcal{F} \to [0,1]$ is a probability measure if

1) $P(\varnothing) = 0$

2) $P(\Omega) = 1$

3) $A_1, A_2 ...$ pairwise disjoint $P(\cup_{n\geq 1} A_n) = \sum_{n\geq 1} P(A_n)$

Since typically $P(\text{a point}) = 0$, P of countable union of points is 0. If we had make definition 1 3) to include uncountable union like for a topology, then we would have

$$P(\text{everything}) = 0$$

for everything is made of points. Bad

Remark 3. Let $A, B \in \mathcal{F}$

1) $P(A^c) = 1 - P(A)$

2) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

3) $A \subset B \implies P(A) \leq P(B)$

4) $A_1, A_2, ... \in \mathcal{F} \implies P(\cup_{n\geq 1} A_n) \leq \sum_{n\geq 1} P(A_n)$ ($\sigma-$subadditivity)

Example 4. 1) $\mathcal{F} = \{\varnothing, \Omega\}$ is a σ field, "the trivial σ field"

2) $\mathcal{F} = \mathcal{P}(\Omega)$ power set. If Ω is discrete. $\mathcal{F}$ is σ. If Ω is not discrete, bad thing can happen.

1.2 Borel Sets, Measurable functions

This is recurring scheme. In topology, one studies open sets, (but no compliment for compliment of open is not usually open.) And one studies the continuous function, because contentious functions are compatible with the structure, i.e. per-image of continuous function of open set is open. Here the compatible function is measurable function, because we will show the per-image of event of measurable function is an event.

Exercise 5. Let $\mathcal{C} \in \mathcal{P}(\Omega)$ be any collection of subset of Ω, then there exists a minimal σ field containing $\mathcal{C}$, denote $\sigma(\mathcal{C})$ and called the σ field generated by $\mathcal{C}$.

Definition 6. Let (S, d) be a metric space and let $O = \{\text{all open sets in } S\}$ then $B(S) := \sigma(O)$ is called the Borel σ field of (S, d). The element of $B(S)$ are called Borel set or Borel measurable set.

One can show the number of Borel set of the real is equal the cardinality of real, so they are continuous.

Example 7. For $S = \mathbb{R}$ interval

$$(a, b)\ [a, b)\ (a, b]\ [a, b]$$

are Borel. Any of these types generates $B(\mathbb{R})$.

Definition 8. Let $(\Omega, \mathcal{F})$ and $(E, \mathcal{E})$ be measurable space, a function $X : \Omega \to E$ is called $\mathcal{F}/\mathcal{E}$ measurable or just measurable (mbl) if

$$X^{-1}(B) \in \mathcal{F} \text{ for all } B \in \mathcal{E}$$

or short hand

$$X^{-1}(\mathcal{E}) \subset \mathcal{F}$$

In the definition $X^{-1}(B)$ means $\{w \in \Omega : X(w) \in B\}$ or short hand $\{X \in B\}$.

If E is metric, we take $\mathcal{E} = B(E)$ by default. i.e. $X : \Omega \to \mathbb{R}$ is mbl, then it's $\mathcal{F}/B(\mathbb{R})$ mbl.

How to check a function is mbl?

Theorem 9. *(Measurability Lemma) Let $X : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ be a function. Assume $\mathcal{E} = \sigma(\mathcal{C})$ for some $\mathcal{C} \in \mathcal{P}(E)$. If*

$$X^{-1}(\mathcal{C}) \subset \mathcal{F} \implies X^{-1}(\mathcal{E}) \subset \mathcal{F}$$

i.e. X is mbl.

Proof. Consider $\mathcal{D} = \{A \subset E : X^{-1}(A) \in \mathcal{F}\}$ is a σ field. By assumption $\mathcal{C} \subset \mathcal{D}$ then by minimality $\mathcal{E} \subset \mathcal{D}$, so $X^{-1}(\mathcal{E}) \subset \mathcal{F}$. □

Example 10. 1) Let $X : \Omega \to \mathbb{R}$. $\mathcal{C} = \{(-\infty, a] : a \in \mathbb{R}\}$ generates $B(\mathbb{R})$, it suffices to check that $X^{-1}((-\infty, a]) \in \mathcal{F}\ \forall a \in \mathbb{R}$.

2) Let Ω be metric space $\Omega = B(S)$ any continuous function, then $X : \Omega \to \mathbb{R}$ is measurable.

Examples of Probability

Lecture 2 (1/23/14)

1) n coin tosses

$$\Omega = \{1, 0\}^n$$

$w \in \Omega$ is of the form $w = \{w_1, ..., w_n\}$, each w_k is the outcome of one coin toss. $\mathcal{F} = \mathcal{P}(\Omega)$. Fair coin

$$P(A) = \frac{|A|}{|\Omega|} = 2^{-n}|A|$$

2) general discrete case

Ω finite or countably infinite set, $\mathcal{F} = \mathcal{P}(\Omega)$. Let

$$(P_w)_{w \in \Omega} \subset [0, 1]$$

satisfies $\sum_{w \in \Omega} P_w = 1$ and put $P(A) = \sum_{w \in A} P_w$. This defines a general form of a probability on Ω.

3) $\Omega \neq \varnothing$ any set. $\mathcal{F} = \mathcal{P}(\Omega)$, $P = \delta_x$ Dirac measure of $x \in \Omega$

$$\delta_x = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

4) Lebesgue measure

By the following theorem which we will prove later.

Theorem 11. *There exists a unique measure* λ *on* $(\mathbb{R}, B(\mathbb{R}))$ *such that*

$$\lambda((a, b]) = b - a$$

for all $a < b$*.*

We let $\Omega = [0, 1]$, $\mathcal{F} = B([0, 1])$, $P = \lambda|_{\mathcal{F}} \implies (\Omega, \mathcal{F}, P)$ is a probability space, P is called uniform distribution.

5) Distribution function

Definition 12. A cumulative probability distribution function is a non decreasing right continuous function

$$F : \mathbb{R} \cup \{\pm\infty\} \to \mathbb{R} \cup \{\infty\}$$

such that $F(-\infty) = 0$, $F(\infty) = 1$.

Likewise define Lebesgue-Stielties measure

Definition 13. Given such F there exists a unique (we will prove uniqueness later) probability measure μ_F on $(\mathbb{R}, B(\mathbb{R}))$ such that $\mu_F((a, b]) = F(b) - F(a)$ for $a < b$.

1.3 Random Variables

Definition 14. A random variable (rv) is a mbl function

$$X : \Omega \to \mathbb{R}$$

Think of an RV as a function depending on the outcome of an experiment.

Definition 15. 1) Let $X : \Omega \to \mathbb{R}$ be a RV then

$$\mu_F(B) := P(X \in B)$$

[Recall $\{X \in B\}$ is short hand for $X^{-1}\{B\} = \{w \in \Omega, X(w) \in B\}$.] defines a probability measure on $(\mathbb{R}, B(\mathbb{R}))$, the distribution of X. Denote

$$\mu_F = P \circ X^{-1}$$

2) The function $F_x : \mathbb{R} \to [0, 1]$

$$F_x(y) = P(y \leq x) = \mu_x((-\infty, x])\ x \in \mathbb{R}$$

is the cumulative distribution function (cdf) of x.

Definition 16. Let $X : (\Omega, \mathcal{F}, P) \to \mathbb{R}$ and $Y : (\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P}) \to \mathbb{R}$ be rv's. We say that they are identical distributed if

$$\mu_X = \mu_Y$$

Remark 17. If $X : \Omega \to (\mathbb{R}^d, B(\mathbb{R}^d))$ $d > 1$ is mbl. X is called random vector and the cdf is a function

$$F_X : \mathbb{R}^d \to \mathbb{R}$$

$$F_x(X_1, X_2, ..., X_d) = P(X_1 \leq x_1, ..., X_d \leq x_d) = P(\cap_{i=1}^d \{X_i \leq x_i\})$$

Definition 18. Discrete distribution, μ_x is discrete if it is of the form

$$\mu_x = \sum_{x \geq 1} p_n \delta_{x_n} \quad p_n \in [0, 1]$$

Definition 19. The distribution is absolutely continuous if the cdf F_x is of the form

$$F_x(X) = \int_{-\infty}^{x} f_X(y) dy$$

for some Borel mbl function f_X called the probability density function (pdf) of X.

Many famous distributions belong to either one of above

Example 20. of absolutely continuous distribution. Gaussian (or normal) distribution with parameter $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}$ has density

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-x_0)^2}{2\sigma^2}}$$

Later we'll do central limit theorem, showing this is the limit of independent experiments.

Example 21. of discrete distribution: Poisson with parameter $\lambda > 0$

$$F(x) = \sum_{\substack{0 \leq n \leq x \\ n \in \mathbb{N}}} e^{-\lambda} \frac{\lambda^n}{n!}$$

1.4 Independence

Lecture 3 (1/28/14)

Definition 22. Let $A, B \in F$ are independent (denotes $A \perp B$) if $P(A \cap B) = P(A)P(B)$.

Suppose $P(A) > 0$,

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

then $A \perp B \iff P(B|A) = P(A)$. Also note that A is independent of itself iff

$$P(A) = P(A)^2 \iff P(A) = 0 \text{ or } P(A) = 1$$

Definition 23. RV X, Y are independent if

$$X^{-1}(A),\ Y^{-1}(B) \text{ are independent, } \forall A, B \in \mathcal{B}(\mathbb{R})$$

Definition 24. Let I be any index set,

1) for each $\alpha \in I$, let $A_\alpha \in \mathcal{F}$ be an event, then $(A_\alpha)_{\alpha \in I}$ is independent if

$$P(A_{\alpha_1} \cap A_{\alpha_2} \cap ... \cap A_{\alpha_n}) = \prod_{i=1}^{n} P(A_{\alpha_i})$$

for any finite family $\alpha_1, ..., \alpha_n \in I$ of distinct indices.

2) for each $\alpha \in I$, let $\mathcal{C}_\alpha \subseteq \mathcal{F}$ be a collection of events, then

$$(\mathcal{C}_\alpha)_{\alpha \in I} \text{ is independent if}$$

$$A_{\alpha_1} \in \mathcal{C}_{\alpha_1}, ..., A_{\alpha_n} \in \mathcal{C}_{\alpha_n} \implies P(A_{\alpha_1} \cap ... \cap A_{\alpha_n}) = \prod_{i=1}^{n} P(A_{\alpha_i})$$

for any finite family $\alpha_1, ..., \alpha_n \in I$ of distinct indices.

3) for each $\alpha \in I$, let X_α be a RV, then

$$(X_\alpha)_{\alpha \in I} \text{ is independent if}$$

$$(\sigma(X_\alpha))_{\alpha \in I} \text{ is independent in the sense of 2)}$$

Hence $\sigma(X) := X^{-1}(\mathcal{B}(\mathbb{R})) = \{X^{-1}(B) : B \in \mathcal{B}\}$.

Exercise 25. Show a collection and pairwise independent event is not independent in general.

Exercise 26. If $(A_\alpha)_{\alpha \in I}$ is a family of independent events and $B_\alpha \in \{A_\alpha, A_\alpha^c\}$, then $(B_\alpha)_{\alpha \in I}$ is independent. So we can forget about definition 23.

Corollary 27. *A family $(A_\alpha)_{\alpha \in I}$ of event is independent iff $(\sigma(A_\alpha))_{\alpha \in I}$ is independent.*

Note 28. $\sigma(A) = \{\varnothing,\ \Omega,\ A,\ A^c\}$.

Exercise 29. If $(\mathcal{C}_\alpha)_{\alpha \in I}$ are independent collection of events, and $\mathcal{D}_\alpha \subseteq \mathcal{C}_\alpha$, then $(\mathcal{D}_\alpha)_{\alpha \in I}$ are independent.

Exercise 30. If $(X_\alpha)_{\alpha \in I}$ are independent RV and $f_\alpha : \mathbb{R} \to \mathbb{R}$ are mbl functions then

$$(f_\alpha(X_\alpha)_{\alpha \in I} \text{ are independent}$$

Hint: because $\sigma(f_\alpha(X_\alpha)) \subseteq \sigma(X_\alpha)$.

2 Integration

We all know that Riemann integral is not suitable for advanced analysis, because limit of Riemann integral functions are not always Riemann integral. So we use Lebesgue.

Consider functions $X : \Omega \to \bar{\mathbb{R}} = [-\infty, \infty]$ on $\bar{\mathbb{R}}$, we use the σ field $\{\mathcal{B}(\mathbb{R}), \{-\infty\}, \{+\infty\}\}$, hence

$$X : \Omega \to \bar{\mathbb{R}} \text{ is mbl iff}$$
$$X^{-1}(\mathcal{B}(\mathbb{R})) \subseteq \mathcal{F} \text{ and } X^{-1}(\pm\infty) \in \mathcal{F} \text{ iff}$$
$$\{X \leq a\} \subset \mathcal{F}\ \forall a \in \mathbb{R}$$

recall $\{X = -\infty\} = \cap_{n \geq 1}\{X \leq -n\}$.

Lemma 31. *If $X, Y : \Omega \to \bar{\mathbb{R}}$ are mbl, then $\{X > Y\} \in \mathcal{F}$.*

Proof. $\{X > Y\} = \cup_{r \in Q}\{X > r\} \cap \{r > y\}$, hence countable union of measurable set is measurable. ☐

Proposition 32. *Let $X_n : \Omega \to \bar{\mathbb{R}}$, $n \geq 1$ be mbl then*

$$sup_n X_n,\ inf_n X_n,\ \limsup_n X_n,\ \liminf_n X \text{ are mbl.}$$

Proof. $\{\sup_x X_n \leq a\} = \cap_n\{X_n \leq a\}$, $\limsup X_n = \inf_{n \geq 1} \sup_{m \geq n} X_m$. The same way to prove $\inf$. ☐

2.1 Simple Functions and Approximation

Definition 33. The indicator function of $A \subseteq \Omega$ is

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

Since $\{1_A > a\} \in \{\varnothing, \Omega, A\}$, A mbl $\iff 1_A$ is mbl.

Definition 34. $X : \Omega \to \mathbb{R}$ is simple if it has finitely many values i.e.

$$\{X(\omega) : \omega \in \Omega\}$$

is a finite set.

If X is simple and $\alpha_1, ..., \alpha_N$ are the distinct values of X, then

$$X = \sum_{i=1}^{N} \alpha_i 1_{\{X=\alpha_i\}}$$

moreover X is amble off

$$\{X = \alpha_i\} \in \mathcal{F} \text{ for } i = 1, ..., N$$

Theorem 35. *(Approximation theorem) Let $X : \Omega \to \bar{\mathbb{R}}$ be any function. There exists simple functions $X_n : \Omega \to \mathbb{R}$ such that*

$$\lim_n X_n(\omega) = X(\omega) \ \forall \omega \in \Omega$$

and 1) if $X \geq 0$, then $(X_n)_{n\geq 1}$ is monotone increasing;

2) if X is ml, then X_n are mbl;

3) if X is bounded, i.e. $\sup_\omega |X(\omega)| < \infty$, then $X_n \to X$ uniformly, i.e.

$$\sup_\omega |X_n(\omega) - X(\omega)| \to 0$$

Lecture 4 (1/30/14)

Proof. Assume $X \geq 0$, define

$$X_n = \sum_{j=1}^{n2^n} (j-1)2^{-n} 1_{\{(j-1)2^{-n} \leq X \leq j2^{-n}\}} + n 1_{\{X \geq n\}}$$

then on $\{X = +\infty\}$, $X_n = n \uparrow X = \infty$; on $\{X < \infty\}$, $0 \leq X(\omega) - X_n(\omega) \leq 2^{-n}$ $\forall n \geq X(\omega)$. So X_n is monotone. If X is bounded, then

$$\sup_{\omega\in\Omega} |X(\omega) - X_n(\omega)| \leq 2^{-n}$$

for n large enough, so $X_n \to X$ uniformly. For general X, we have

$$X = X^+ - X^-$$

where $X^+ = \max\{0, X\}$, $X^- = (-X)^+$. Applying previous observation to $X^\pm$, we have

$$Y_n \to X^+ \text{ and } Z^n \to X^-$$

with $X_n = Y_n - Z^n$, $|X| = X^+ + X^-$. □

2.2 Integration and Non-negative Functions

Fix $(\Omega, \mathcal{F}, \mathcal{P})$, the idea is the following:

1)

$$\int 1_A dp = P(A)$$

2) By linearity

$$\int (\sum \alpha_n 1_{A_n}) dp = \sum \alpha_n P(A_b)$$

3) For $X \geq 0$ let X_n as in the approximation therm and

$$\int X dp := \lim \int X_n dp$$

Note in probability, we write $E[X]$ for $\int X dp$. And we want all results in this section apply to general measures.

Notation 36.

$$M = M(\Omega, \mathcal{F}) = \{X : \Omega \to \bar{\mathbb{R}} \text{ mbl}\}$$

$$M_+ = \{X \in M,\ X \geq 0\}$$

$$S = \{X \in M \mid X \text{ simple}\}$$

$$S_1 = \{X \in S \mid X \geq 0\}$$

Definition 37. Let $X \in S_+$ i.e.

$$X = \sum_{k=1}^{n} \alpha_k 1_{A_k} \text{ with } A_k = \{X = \alpha_k\}$$

$$E[X] := \int_\Omega X dp = \sum_{k=1}^{n} \alpha_k P(A_k)$$

called weighted average.

Lemma 38. *Let $X, Y \in S_+$, $A \in \mathcal{F}$, $\alpha \geq 0$*

1) $E[1_A] = P(A)$

2) $E[\alpha X] = \alpha E[X]$

3) $E[X + Y] = E[X] + E[Y]$

4) $X \leq Y \implies E[X] \leq E[Y]$

Proof. As exercise. □

Remark 39. The integral is taken over the whole space Ω

$$\int_A X dp := \int (X 1_A) dp$$

Exercise 40. Let $Z \in S_+$, then $\mu(A) := E[1_A, Z]$ defines a measure μ. Note $\mu(\Omega) = E[Z]$ and Z is the density of μ wrt P.

Definition 41. Let $X \in M_+$ then

$$E[X] = \sup\{E[Z] \ : \ Z \in S_+, Z \in X\}$$

Proposition 42. *Let $X \leq Y \in M_+$ then $E[X] \leq E[Y]$.*

Proof. The sup in the definition of $E[Y]$ is taken over a larger set. □

Proposition 43. *(Chebyshev's inequality) Let $X \in M_+$ and $a > 0$, then*

$$P(X \geq a) \leq \frac{E[X]}{a}$$

Proof. Note that $a 1_{\{X \geq a\}} \leq X 1_{\{X \geq a\}} \leq X$, then take E, then use monotonicity, we get

$$aP(X \geq a) \leq E[X]$$

□

Corollary 44. *Let $X \in M_+$, then $E[X] < \infty \implies P(X = \infty) = 0$.*

Proof. $\forall\, n \in \mathbb{N}$,

$$P(X = \infty) \leq P(X \geq 0) \overset{\text{cheb.}}{\leq} \frac{E[X]}{n} \to 0.$$

$\square$

Next we want to examine $X_n \to X$ in some sense $\implies E[X_n] \to E[X]$ in some corresponding sense.

Theorem 45. *(Monotone Convergence Theorem or called Beppo Levi's Theorem) Let $X_n \in M_+$, $n \geq 1$ be a monotone increasing sequence and*

$$X := \lim_n X_n = \sup_n X_n$$

then $X \in M_+$ and

$$E[X] = \lim_n E[X_n] = \sup_n E[X_n]$$

Proof. Recall that X is mbl as a limit of mbl functions. By monotonicity

$$E[X_n] \leq E[X_{n+1}] \leq ... \leq E[X]\ \forall\, n$$

hence $\lim E[X_n] \leq E[X]$. It remains to show $\lim E[X_n] \geq E[X]$

Next by definition of $E[X]$, we need to show $\lim E[X_n] \geq E[Z]\ \forall\, Z \in S_+$ with $Z \leq X$ so let $Z \in S_+$ and let $\alpha \in (0, 1)$, define

$$A_n = \{X_n \geq \alpha Z\}$$

then $A_n \uparrow \Omega$, moreover by monotonicity

$$\alpha \underbrace{E[Z 1_{A_n}]}_{\mu(A_n)} \leq E[X_n 1_{A_n}] \leq E[X] \tag{2.1}$$

Consider the finite measure $\mu(A) := E[Z 1_A]$. By continuity of measure

$$\lim \mu(A_n) < \mu(\Omega) = E[Z]$$

With (2.1), we get

$$\alpha E[Z] \leq \lim E[X_n]$$

$\alpha \in (0, 1)$ was arbitrary $\implies E[Z] \leq \lim E[X_n]$. $\square$

Note that in the proof we used continuity of measure

$$A_n \uparrow \text{ or } \downarrow A \implies \mu(A_n) \uparrow \text{ or } \downarrow \mu(A)$$

it only holds for $\mu(A) < \infty$, because its proof involves complement.

Proposition 46. *Let $X, Y \in M_+ \alpha \geq 0$ then 1)*

$$E[\alpha X] = \alpha E[X]$$

2)

$$E[X + Y] = E[X] + E[Y]$$

Proof. 1) Let $Z_n \in S_+$ with $Z_n \uparrow X$ by approximation theorem

$$E[\alpha X] = E[\lim \alpha Z_n] \overset{\text{MCT}}{=} \lim E[\alpha Z_n] \underset{S_+}{\overset{\text{linearity}}{=}} \alpha \lim E[Z_n] \overset{\text{MCT}}{=} \alpha E[\alpha].$$

2) take also $\tilde{Z}_n \uparrow Y$, $\tilde{Z}_n \in S_+$

$$\begin{aligned} E[X+Y] &= E(\lim(Z_n + \tilde{Z}_n)) \overset{\text{MCT}}{=} \lim E(Z_n + \tilde{Z}_n) \\ &\overset{\text{linearity}}{=} \lim(E(Z_n) + E(\tilde{Z}_n)) = \lim E[Z_n] + \lim E[\tilde{Z}_n] \\ &\overset{\text{MCT}}{=} E[X] + E[Y] \end{aligned}$$

□

Lecture 5 (2/4/14)

Theorem 47. *(Faton's lemma) Let $X_n \in M_+$, $n \geq 1$ then*

$$E[\liminf_{n\to\infty} X_n] \leq \liminf_{n\to\infty} E[X_n]$$

Proof. Let $Y_n = \inf_{k \geq n} X_k$, then

$$Y_n \uparrow \liminf_{n\to\infty} X_n \text{ and } Y_n \leq X_n \text{ for } k \geq n$$

By monotonicity $E[Y_n] \leq \inf_{k \geq n} X_n \to \liminf_{k\to\infty} E[X_k]$ as $n \to \infty$. Hence

$$E[\liminf_{n\to\infty} X_n] = E[\lim Y_n] \overset{\text{MCT}}{=} \lim E[Y_n] \leq \liminf E[X_n]$$

□

Proposition 48. *Let $X \in M_+$ and define*

$$\mu(A) = E[X1_A]$$

then μ is a measure.

Definition 49. $A \subset \Omega$ is a null set if $A \subseteq B \in \mathcal{F}$ and $P(B) = 0$

Definition 50. A property is said to hold almost surely (a.s) or almost everywhere (a.e) if it holds outside a null set. E.g.

$$X = 0 \text{ a.s means that } \{X \neq 0\} \text{ is a null set.}$$

Null set is not a big deal, since we can always redefine measure, i.e. completion.

Lemma 51. *Let $X \in M_+$ then $E[X] = 0 \iff X = 0$ a.s*

Proof. $\Longrightarrow$

$$P(X > 0) \overset{\text{MCT}}{=} \lim_{n\to\infty} P(X > \frac{1}{n}) \overset{\text{cheb}}{\leq} \lim_{n} nE[X] = 0$$

$\Longleftarrow$ Let $X_n = n1_{\{X>0\}}$ then

$$E[X_n] = nP(X > 0) = 0$$

Note

$$X \leq \infty 1_{\{X>0\}} = \lim X_n \overset{\text{MCT}}{\Longrightarrow} E[X] \leq E[\lim X_n] = \lim E[X_n] = 0$$

□

A more general result

Proposition 52. *Let $X, Y \in M_+$ then*

$$X = Y \text{ } a.s. \implies E[X] = E[Y]$$

Proof. Let $A = \{X = Y\}$, $B = \{X \neq Y\}$. By previous lemma

$$E[, X1_B] = E[Y1_B] = 0$$

By definition of A, $E[X1_A] = E[Y1_A]$, hence

$$\begin{aligned} E[X] &= E[X[1_A + 1_B]] = E[X1_A] + E[X1_B] \\ &= E[Y1_A] + E[Y1_B] = E[Y[1_A + 1_B]] = E[Y] \end{aligned}$$

□

2.3 Integrable Functions

Definition 53. $\mathcal{L}^0 = \{X : \Omega \to \mathbb{R},\ \text{mbl}\}$

Definition 54. $X : \Omega \to \bar{\mathbb{R}}$ is integrable if it's measurable and $E[X^+] < \infty$ and $E[X^-] < \infty$

Definition 55. $\mathcal{L}^1 = \{X \in \mathcal{L}^0 :\ X \text{ integrable}\}$.

Lemma 56. *Let $X : \Omega \to \mathbb{R}$ be mbl*

1) $X \in \mathcal{L}^1 \iff E[|X|] < \infty$

2) $|E[X]| \leq E[|X|]$ for $X \in \mathcal{L}^1$

The reason to restrict X real-valued function is to use subtraction without worrying $\infty - \infty$.

Proof. 1) $E[|X|] = E[X^+ + X^-] = E[X^+] + E[X^-]$

2)
$$|E[X]| = \left|E[X^+ - X^-]\right| \leq E[X^+] + E[X^-] = E[|X|]$$

□

Lemma 57. *Let $X \in \mathcal{L}^0$, $Y \in \mathcal{L}^1$ and $|X| \leq |Y|$ then $X \in \mathcal{L}^1$.*

Proof. By previous lemma,

$$Y \in \mathcal{L}^1 \implies E[|Y|] < \infty \implies E[X] < \infty \implies X \in \mathcal{L}^1$$

□

Theorem 58. *$\mathcal{L}^1$ is a vector space and $E : \mathcal{L}^1 \to \mathbb{R}$ is linear.*

Proof. As an exercise. □

Theorem 59. *(Dominated Convergence theorem or called Lebesgue Theorem) Let $X_n, X \in \mathcal{L}^0$ and*

1) $X_n \to X$ a.s. 2) $\exists Y \in \mathcal{L}^1$ s.t. $|X_n| \leq |Y|$ a.s. for all n.

then $X_n, X \in \mathcal{L}^1$ and

i) $E[X_n] \to E[X]$ ii) $E[|X_n - X|] \to 0$

Proof. Since $|E[X_n] - E[X] \leq E[|X_n - X|]|$, it suffices to check ii) first. We may assume 1), 2) hold everywhere by setting X_n, X to 0 on the null set and $Y > 0$, so by Faton,

$$E[|X|] = E[\left|\liminf_x X_n\right|] \leq \liminf_n E[|X_n|] \leq E[|Y|] < \infty$$

Next let $Y_n = |X| + Y - |X_n - X|$, then $Y_n \geq 0$, $Y_n \in \mathcal{L}^1$, by Faton

$$\begin{aligned} E[|X| + Y] &= E[\lim Y_n] \leq \liminf_{n\to\infty} E[Y_n] \\ &= \liminf_{n\to\infty} E[|X| + Y - |X_n - X|] \\ &= \underbrace{E[|X| + Y]}_{<\infty \to 0} - \limsup_{n\to\infty} E[|X_n - X|] \end{aligned}$$

so

$$\lim E[|X_n - X|] = 0$$

□

3 Convergence

3.1 L^p Spaces

For this discussion, unless mentioned, everything should work for general measure.

Definition 60. Let V be a vector space (over $\mathbb{R}$) Let $\|\cdot\| : V \to \mathbb{R}$ be used that

1) $\|v\| \geq 0 \; \forall \, v \in V$

2) $\|\alpha v\| = |\alpha| \, \|v\| \; \forall \, \alpha \in \mathbb{R} \; v \in V$

3) $\|v + w\| \leq \|v\| + \|w\| \; \forall \, v, w \in V$ then $\|\cdot\|$ is called a semi-norm

if also

4) $\|v\| = 0 \iff v = 0$, then $\|\cdot\|$ is a norm.

Proposition 61. $\|\cdot\|_1 : \mathcal{L}^1 \to \mathbb{R}$, $\|X\|_1 = E[X]$ *is a semi-norm.*

Proof. It follows directly from monotonicity and linearity of E. □

We have to get rid of null sets to get a proper norm.

Lecture 6 (2/6/14)

Definition 62. $L^2 = \mathcal{L}^1(\Omega, \tilde{\mathbb{R}}, P)$ is defined as the set of equivalence class

$$L^1 = \mathcal{L}^1 / \sim$$

where $X \sim Y \iff X = Y$ a.s. We could denote $[X]$ the equivalence class of X, but it is standard to identity X as $[X]$.

Theorem 63. *$(L^1, \|\cdot\|_1)$ is a normed vector space.*

Proof. $\|X - Y\| = 0 \iff |X - Y| = 0$ a.s. $\iff [X] = [Y]$. □

Definition 64. Let $1 \le p < \infty$ then

$$\begin{aligned} \mathcal{L}^p &= \{X : \mathbb{R} \to \mathbb{R} \text{ mbl } E[|X|^p] < \infty\} \\ L^p &= \mathcal{L}^p / \sim \\ \|X\|_p &= (E[|x|^p])^{1/p} \end{aligned}$$

Theorem 65. *(Holder inequality) Let $1 < q, p < \infty$ be s.t.*

$$\frac{1}{p} + \frac{1}{q} = 1$$

(p, q) are called conjugate. Let $X \in L^p$, $Y \in L^q$ then

$$\|XY\|_1 \le \|X\|_p \|Y\|_q$$

Proof. WOLG assume $\|X\|_p, \ \|Y\|_q \ne 0$, applying Young's inequality with $\alpha = |X| / \|X\|_p$, $\beta = |Y| / \|Y\|_q$, then

$$\frac{|XY|}{\|X\|_p \|Y\|_q} \le \frac{|X|^p}{p \|X\|_p^p} \frac{|Y|^q}{q \|Y\|_q^q}$$

taking $E[]$, we get

$$\frac{\|XY\|_1}{\|X\|_p \|Y\|_q} \le \frac{\|X\|_p^p}{p \|X\|_p^p} \frac{\|Y\|_q^q}{q \|Y\|_q^q} = \frac{1}{p} + \frac{1}{q} = 1$$

□

Corollary 66. *(Cauchy-Schwartz) For $X, Y \in L^2$*

$$\|XY\|_1 \le \|X\|_2 \|Y\|_2$$

i.e.

$$E[|XY|] \le \sqrt{E[X^2] E[Y^2]}$$

Theorem 67. *(Generalized Triangle Inequality or Minkowski Inequality) $X, Y \in L^p$, $p \geq 1$ then*

$$X + Y \in L^p$$

and

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$$

Proof. We are done with $p = 1$. Let $p > 1$ and put q such that $\frac{1}{p} + \frac{1}{q} = 1$, we have

$$|X + Y|^p \leq 2^p(\max\{|X|, |Y|\})^p \leq 2^p(|X|^p + |Y|^p)$$

that is

$$E[|X + Y|^p] \leq 2^p(E[|X|^p] + E[|Y|^p])$$

moreover

$$|X + Y|^p = |X + Y| |X + Y|^{p-1} \leq |X| |X - Y|^{p-1} + |Y| |X + Y|^{p-1} \tag{3.1}$$

Since $|X + Y|^p \in L^1$ and $(p - 1)q = p$, we have $|X + Y|^{p-1} \in L^q$. Then apply Holder

$$E[|X||X + Y|^{p-1}] \leq \|X\|_p \left\| |X + Y|^{p-1} \right\|_q = \|X\|_p \|X + Y\|_p^{p/q}$$

Then by (3.1)

$$\|X + Y\|_p^p \leq (\|X\|_p + \|Y\|_p) \|X + Y\|_p^{p/q}$$

□

Corollary 68. $(L^p, \|\cdot\|_p)$ *is a normed vector space.*

Definition 69. Let X_n, $X \in \mathcal{L}^p$, we say that $X_n \to X$ in L^p if

$$\|X_n - X\| \to 0$$

i.e. $E[|X_n - X|^p] \to 0$.

Definition 70. A normed vector space is a Banach space, if it is complete, i.e. all Cauchy sequence converge.

Theorem 71. *(Riesz Fischer) Let $1 \leq p \leq \infty$ and let $(X_n)_{n\geq 1} \in \mathcal{L}^n$ be Cauchy, then*

1) $\exists X \in \mathcal{L}^p$ s.t.

$$X_n \to X \text{ in } L^p$$

2) $\exists$ a sub sequence $(X_{n_k})_{k\geq 1}$ such that $X_{n_k}(\omega) \to X(\omega)$ for almost every ω.

Proof. Since (X_n) is Cauchy, we can take a "fast" sub sequence $Y_k = X_{n_k}$ st.t.

$$\|Y_{k+1} - Y_k\|_p \leq 2^{-k},\ k \geq 1$$

Let $H_n = \sum_{k=1}^n |Y_{k+1} - Y_k|$ by Minkowski

$$\|H_k\| \leq \sum_{k=1}^n \|Y_{k+1} - Y_k\| \leq 1$$

Let $H = \lim_n H_n = \sum_{k=1}^\infty |Y_{k+1} - Y_k|$

$E[H^p] = E[\lim H_n^p] \overset{\text{MCT}}{=} \lim E[H^p] \leq 1 \implies N = \{H = \infty\}$ is a nullset.

For $\omega \in \Omega / \sim H_n(\omega) \uparrow H(\omega) < \infty$ and $(Y_n(\omega))_{n\geq 1}$ is Cauchy in $\mathbb{R}$. For $n > m > 1$

$$\begin{aligned} |Y_m(\omega) - Y_n(\omega)| &\leq |Y_m(\omega) - Y_{m-1}(\omega)| + ... |Y_n(\omega) - Y_{n-1}(\omega)| \\ &= H_m(\omega) - H_n(\omega) \text{ is Cauchy} \end{aligned}$$

$\mathbb{R}$ is complete $\implies (Y_n(\omega))_n$ has a limit, call it $Y(\omega)$. For $\omega \in N$, set $Y(\omega) = 0$.

TO see that $Y_n \to Y$ in L^p

$$\lim_n \|Y - Y_n\|_p^p = \lim E[|Y - Y_n|^p] = E[\lim |Y - Y_n|] = 0$$

The second equality above is given by DCT with $|Y - Y_n|^p \leq H^p \in \mathcal{L}^1$.

We have proved any sub sequence of (X_n) has a sub sequence which converges in L^p and the limit is always Y.

This implies that (X_n) itself converges to Y. □

Remark 72. We have proved something more. X_n Cauchy $\sum_n \|X_n - X_{n-1}\|_p < \infty \implies X_n$ converges a.s.

3.2 L^∞ Space

Definition 73. $\mathcal{L}^\infty$ is the space of mbl function that are bounded

$$\mathcal{L}^\infty = \{X : \Omega \to \mathbb{R} \text{ mbl } \exists\, c \in \mathbb{R} \text{ s.t. } |X| \le c \text{ a.s.}\}$$

$$\|X\|_\infty = \{c \in \mathbb{R} : |X| < c \text{ a.s.}\}$$

Definition 74. $L^\infty = \mathcal{L}^\infty / \sim$

Theorem 75. *$(L^\infty, \|\cdot\|_\infty)$ is Banach.*

Remark 76. Work only for finite measure

1) If $X \in \mathcal{L}^\infty$, then $X \in \mathcal{L}^p\ \forall\, 1 \le p < \infty$ and $\|X\|_\infty = \lim_\infty \|X\|_p$

2) Conversely if $\sup_{p\to\infty} \|X\|_p < \infty$, then $X \in \mathcal{L}^\infty$.

Proof. As exercise □

Lecture 7 (2/11/14)

Theorem 77. *(Jensen)Let $X \in \mathcal{L}^1$ and let $\psi : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ be a convex function, then*

$$E[\psi(X)] \ge \psi(E[X])$$

Proof. Let $X_0 \in \psi(X)$, Since ψ is convex, $\exists\, a, b \in \mathbb{R}$ s.t.

$$aX + b \le \psi(X)\ \forall\, X \in \mathbb{R}$$

and

$$aX_0 + b = \psi(X_0)$$

so

$$E[\psi(X)] \ge E[aX + b] = \psi(X) + b = aX_0 + b = \psi(E[X])$$

□

Corollary 78. *$\forall\, 1 \le p \le q \le \infty$,*

$$L^1 \supseteq L^p \supseteq L^q \supseteq L^\infty \text{ and } \|\cdot\|_p \le \|\cdot\|_q$$

Proof. Let $p \le q < \infty$ then $X \mapsto X^{p/q}$ is convex. For $X \in \mathcal{L}^p$

$$E[|X|^p] = E[|X|^p]^{\epsilon/p} \ge E[|X|^\epsilon]^{q/p}$$

□

Definition 79. (Convergence in measure/probability) Let X_n, X be RV on $(\Omega, \mathcal{F}, \mathcal{P})$, we say that $X_n \to X$ in measure (or $X_n \overset{P}{\to} X$) if

$$P(|X_n - X_m| > \alpha) \to 0 \text{ as } m, n \to \infty$$

Theorem 80. *Let (X_n) be Cauchy in measure P*

1) $\exists$ a RV X s.t. $X_n \overset{P}{\to} X$

2) $\exists$ a sub sequence $X_{n_k} \to X$ a.s.

Proof. Use the fast sub sequence (Y_k) as before, then

$$A_k = \{\omega : |Y_{k+1}(\omega) - Y_k(\omega)| \geq 2^{-k}\}$$

we have

$$P(A_k) \leq 2^{-k}$$

$$B_k = \bigcup_{j \geq k} A_j,\, B = \bigcap_{k \in \mathbb{N}} B_k$$

then

$$P(B_k) \leq \sum_{j \leq k} P(A_j) \leq \sum_{j \geq k} 2^{-k} \leq 2^{-(k-1)} \;\forall\, k \geq 1$$

Hence $P(B) = 0$. Let $\omega \notin B_{k_0}^C$ for some $k_0 \in \mathbb{N}$

$$B_{k_0}^C = \bigcap_{j \geq k_0} A_j^C = \bigcap_{j \geq k_0} \{|Y_{k+1} - Y_j| < 2^{-j}\}$$

For $i \geq j \geq k_0$

$$\begin{aligned} |Y_i(\omega) - Y_j(\omega)| &\leq |Y_i(\omega) - Y_{i-1}(\omega)| + ... + |Y_{j+1}(\omega) - Y_j(\omega)| \\ &\leq \frac{1}{2^{i-1}} + ... + \frac{1}{2^j} \leq \frac{1}{2^{j-1}} \end{aligned} \tag{3.2}$$

then $(Y_i(\omega))_{i \geq 1}$ is Cauchy in $\mathbb{R}$, let

$$X(\omega) = \begin{cases} \lim Y_i(\omega) & \omega \notin B \\ 0 & \omega \in B \end{cases}$$

Since $\omega \notin B_k \implies$ by (3.2),

$$|Y_k(\omega) - X(\omega)| \leq \frac{1}{2^{k-1}} \tag{3.3}$$

for $j \geq k$. We have $Y_k \to X$ uniformly on B_k^C, in particularly $Y_k \to X$ a.s.

We have $Y_k \xrightarrow{P} X$, and by (3.3)

$$P(|X_\alpha - X| > \alpha) \leq P(|X_n - X_k| \geq \frac{\alpha}{2}) + P(|Y_k - X| \geq \frac{\alpha}{2}) \to 0 + 0 = 0$$

for $n, k \to \infty$, because they are Cauchy. □

Definition 81. $L^0 = \mathcal{L}^0 / \sim$

Notation 82. $x \wedge y = \min\{x, y\}$

Theorem 83. *(this only works for finite measure) Let*

$$d(X, Y) = E[|X - Y| \wedge 1]$$

$X, Y \in \mathcal{L}^0$, then (L^0, d) is a complete metric space. Moreover $d-$convergence is equivalent to convergence in measure and $d-$Cauchy is equivalent to Cauchy in measure.

Corollary 84. *(true for general measure) A sequence can have at most one limit in measure and any Cauchy sequence in measure converges in measure.*

Corollary 85. *(for probability measure) Let $X_n, X \in \mathcal{L}^0$ then $X_n \to X$ a.s. implies*

$$X_n \xrightarrow{P} X$$

Proof. Use DCT. □

Theorem 86. $1 \leq p \leq \infty$

$$X_n \xrightarrow{L^p} X \implies X_n \xrightarrow{P} X$$

or called L^0 convergence.

Proof. (first proof, using probability measure)

$$X_n \to X \text{ in } L^p \implies X_n \to X \text{ in } L^1$$

so

$$|X_n - X| \wedge 1 \to 0 \text{ in } L^1$$

i.e. $d(X_n, X) \to 0$.

(second proof, general measure)

a) $p < \infty$, let

$$A_n(\alpha) = \{|X_n - X| > \alpha\},$$

then

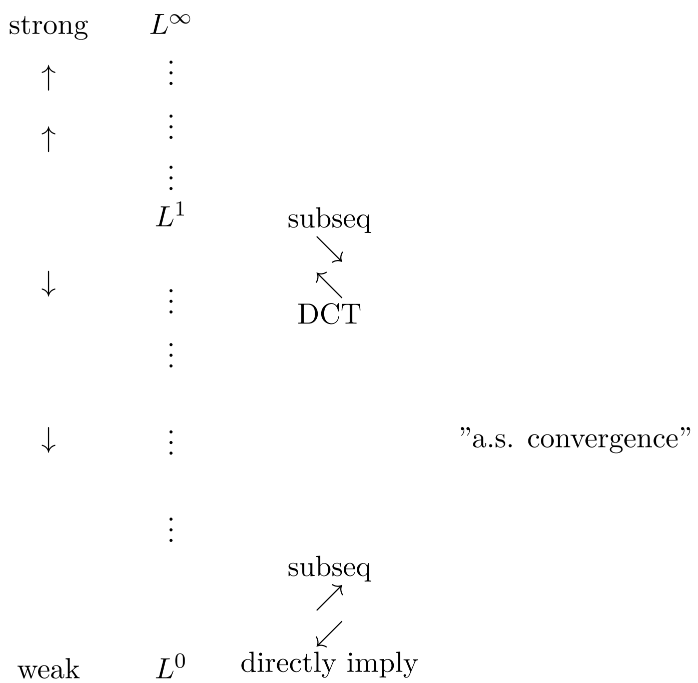
$$E[|X - X_n|^p] \geq E[1_{A_n} |X - X_n|^p] \geq E[\alpha^p 1_{A_n}] = \alpha^p P(A_n)$$

Hence $P(A_n) \to 0 \; \forall \, \alpha > 0$.

b) $p = \infty$, $\forall \, \alpha > 0 \; \exists \, N \in \mathbb{N}$ s.t. $\|X_n - X\|_\infty < \infty \; \forall \, n \geq N$ hence

$$\forall \, n \geq N, \; P(|X_n - X| > \alpha) = 0 \; \forall \, n \geq N$$

Convergence Summary for probability measure:



□

3.3 Uniform Integrability

In this section, P has to be finite measure.

Definition 87. A family $(X_\lambda)_{\lambda\in I}$ of RV is uniformly integrable (UI) if

$$\lim_{c\to\infty}\sup_{\lambda\in\Lambda} E[|X_\lambda|\,1_{|X_\lambda|\geq c}] = 0$$

Remark 88. If $(X_\lambda)_{\lambda\in I}$ is UI then (X_λ) is bounded in L^1 i.e.

$$\sup_{\lambda\in\Lambda}\|X_\lambda\|_1 < \infty$$

Proof.

$$\|X_\lambda\|_1 = E[|X_\lambda|] = \underbrace{E[|X_\lambda|\,1_{|X_\lambda|<c}]}_{\leq c} + \underbrace{E[|X_\lambda|\,1_{|X_\lambda|\geq c}]}_{\to 0 \text{ uniform in } \lambda}$$

hence it's bounded. □

Theorem 89. *($\epsilon-\delta$ Criterion) Let $(X_\lambda)_{\lambda\in I}$ be rv's. The following are equivalent*

1) (X_λ) is UI

2) a) (X_λ) is bounded in L^1 b) $\forall\,\epsilon>0\;\exists\,\delta>0$ s.t.

$$A\in\mathcal{F}, P(A)\leq\delta \implies \sup_\lambda E[|X_\lambda|\,1_A]\leq\epsilon$$

Proof. Assume 1) with remark 88 $\implies$ 2)a). To see b) Let $\epsilon>0$, $\forall\,c$

$$E[|X_\lambda|\,1_A]\leq cP(A)+E[|X_\lambda|\,1_{|X_\lambda|>c}]$$

$cP(A)\leq\frac{\epsilon}{2}$ for $\delta=\frac{\epsilon}{2c}$, and $E[|X_\lambda|\,1_{|X_\lambda|>c}]\leq\frac{\epsilon}{2}$ for large c, so

$$E[|X_\lambda|\,1_A]\leq\epsilon$$

Now assume 2) let

$$\alpha=\sup_\lambda E[|X_\lambda|]$$

By Chebyshev

$$P(|X_\lambda|>c)\leq\frac{E[X_\lambda]}{c}\leq\frac{\alpha}{c}$$

By assumption,

$$E[|X_\lambda|\,1_{|X_\lambda|>c}]\leq\epsilon$$

for c large. □

Example 90. 1) finite subset

$$\{X_1, ..., X_N\} \subseteq \mathcal{L}^2$$

are UI

2) $Y \in L^1$, $|X_n| \leq Y \implies (X_n)_{n\geq 1}$ is UI

3) (X_λ) UI, $Y \in \mathcal{L}^1 \implies Y + X_\lambda$ is UI.

Theorem 91. *Let X_n,X be rv's s.t.*

$$X_n \xrightarrow{\mathcal{L}^0} X$$

then

$$X_n \xrightarrow{\mathcal{L}^1} X \iff (X_n) \text{ is UI}$$

Most of time people use $\Longleftarrow$ part of the theorem.

Lecture 8 (2/13/14)

Proof. WOLG $X \equiv 0$

($\Longleftarrow$) Let (X_n) be UI we have

$$E[|X_n|] \leq \epsilon + E[|X_n| \, 1_{|X_n|>c}] \; \forall \, \epsilon > 0 \tag{3.4}$$

As $X_n \to 0$, given $\delta > 0$, $\epsilon > 0 \; \exists \, k_0$ s.t.

$$P(|X_k| > \epsilon) \leq \delta \; \forall \, k \geq k_0$$

using the $\delta - \epsilon$ criterion with $A = \{|X_n| > \epsilon\}$ we get

$$E[|X_n| \, 1_{|X_n|>c}] \leq \epsilon$$

with (3.4) we have the conclusion.

($\implies$) Let $X_n \xrightarrow{\mathcal{L}^1} X$ i.e.

$$E[|X_n|] \to 0$$

Let $\epsilon > 0$,

$$E[|X_n| \, 1_{|X_n|>c}] \leq E[|X_n|] \leq \epsilon$$

□

for $n \geq N(\epsilon)$.

Moreover for c large, $E[|X_n| \, 1_{|X_n|>c}] \leq \epsilon$ for $n = 1, ..., N(\epsilon)$.

Theorem 92. *(delia-wallace-pansino) TFAE*

1) (X_λ) *is UI*

2) $\exists\, \psi : \mathbb{R}_+ \to \mathbb{R}_+$ *mbl, increasing s.t.*

$$\lim_{X\to\infty} \frac{\psi(X)}{X} = +\infty$$

and

$$\sup_\lambda E[\psi(X_\lambda)] < \infty$$

Proof. We only show 2) $\implies$ 1) Let $\alpha > 0$ for C large enough $\psi(X) \geq \alpha X \;\forall\, X \geq c$ Hence

$$\psi(|X_\lambda|) \geq \alpha\, |X_\lambda|$$

on

$$\{|X_\lambda| \geq c\} \implies E[|X_\lambda|\, 1_{|X_\lambda|>c}] \leq \frac{E[\psi(|X_\lambda|)]}{\alpha} \leq \frac{\text{const}}{\alpha}$$

□

Corollary 93. *If* (X_λ) *is bounded in* $L^{1+\epsilon}$ *for some* $\epsilon > 0$ *i.e.*

$$\sup_\lambda E[|X_\lambda|^{1+\epsilon}] < \infty$$

then (X_λ) *is UI.*

Proof. Take $\psi(X) = X^{1+\epsilon}$. □

Recap what we did

$$\begin{aligned}
&E[X]^2 \\
E[1_A] &= P(A) \\
E[\textstyle\sum \alpha_n 1_{A_n}] &= \textstyle\sum a_n P(A_n) \\
E[X] &= \sup\{E[Z] : Z \leq X \;Z \text{ simple}\} \text{ with } X \geq 0 \\
E[\cdot] &\quad \text{linear on } \mathcal{L}^1 \\
\text{MCT} &\quad X_n \uparrow X,\; X^- \in \mathcal{L}^1 \implies E[X_n] \to E[X] \\
&\quad E[|X_n - X|] \to 0 \text{ if } X \in \mathcal{L}^1 \\
\text{Faton} &\quad X_n \geq Y \text{ s.t. } Y^- \in \mathcal{L}^1 \\
&\quad E[\liminf X_n] \leq \liminf E[X_n] \\
\text{DTC} &\quad |X_n| \leq Y \in \mathcal{L}^1,\; X_n \to X \text{ a.s. (or in } P) \Longrightarrow X_n \xrightarrow{L^1} X
\end{aligned}$$

4 Simple Random Walks

4.1 Binomial Walk

Definition 94. Let $I \subseteq \mathbb{R}$ “time”, a stochastic process is a function

$$\begin{aligned} S : \Omega \times I &\rightarrow \mathbb{R} \text{ or something else} \\ (\omega, t) &\mapsto S_t(\omega) \text{ or } S(t, \omega) \end{aligned}$$

E.g $I = [0, \infty)$ continuous time process, $I = \mathbb{N}$ discrete time.

Two equivalent ways of thinking about it

1) “a family of r.v.” indexed by t

2) a family of function $t \mapsto S_t(\omega)$ indexed by ω, and $t \mapsto S_t(\omega)$ is called sample path.

Definition 95. A random walk on $\mathbb{Z}$ (starting at 0) is a process

$$S_k = \sum_{j=1}^{k} X_j,\ k = 0, 1, ...$$

where $(X_j)_{j \geq 1}$ are iid with values in $\mathbb{Z}$.

- S is simple r.v. if $P(X_1 = 1) = p$, $P(X_1 = -1) = 1 - p = q$
- S is symmetric if $p = q = 1/2$

Note existence of a random walk with given distribution of X_1 is true but not trivial, which we will show later.

A model for simple random walk is N steps: $\Omega = \{w_1, ..., w_N\} = \{\pm 1\}^N$
$w_i \in \{\pm 1\}$

$$\mathcal{F} = \mathcal{P}(\Omega),\ P(A) = \frac{|A|}{|\Omega|}$$

Example 96. Let S be a simple random walk,

$$S_0 = 0,\ S_k = \sum_{j=1}^{k} X_j, (X_j)_{j \geq 1} \text{ iid}$$

with $P(X_j = 1) = p \in [0, 1]$, and $P(X_j = -1) = 1 - p = q$. Later we will show in 1 or 2D, all destinations within the range are reachable, but in 3D it is not the case.

Definition 97. A r.v. Z has Binomial (n, p) distribution $Z \sim B$ in (n, p) if

$$P(Z = k) = \begin{pmatrix} n \\ k \end{pmatrix} p^k (1-p)^{n-k}, k = 0, ..., n$$

where $\begin{pmatrix} n \\ k \end{pmatrix} = \frac{n!}{k!(n-k)!}$

This is the number of successes out of n independent trials if each success has probability p.

Proposition 98.

$$P(S_n = k) = \begin{cases} \begin{pmatrix} n \\ \frac{n+k}{2} \end{pmatrix} p^{\frac{n+k}{2}} q^{\frac{n-k}{2}} & k \in \{-n, ..., n\} \\ 0 & else \end{cases}$$

Proof. Let

$$\begin{aligned} A_n &= \#\{k \leq n : X_k = 1\} \\ B_n &= \#\{k \leq n : X_k = -1\} \end{aligned}$$

then

$$S_n = A_n - B_n \implies 2A_n = S_n + A_n + B_n = S_n + n \implies A_n = \frac{S_n + n}{2}$$

so

$$P(S_n = k) = P(A_n = \frac{k+n}{2})$$

□

The following theorem used beyond random walk

Proposition 99. *(Reflection Principle) For* $k > 0$, $l \geq 0$

$$\# \text{ paths of length } n \text{ from } 0 \text{ to } k - l \text{ visiting } k \text{ at laest one}$$
$$=$$
$$\# \text{ paths of length } n \text{ from } 0 \text{ to } k + l \text{ visiting } k \text{ at laest one}$$

$$= \begin{cases} \begin{pmatrix} n \\ \frac{n+k+l}{2} \end{pmatrix} & if \ \frac{n+k+l}{2} \in \{0, 1, ..., n\} \\ 0 & else \end{cases}$$

Proof. By picture. □

Lecture 9 (2/18/14)

Definition 100. The passage of time is defined as for $k \in \mathbb{Z}$, let

$$T_k = \begin{cases} \text{first time walk visiting } k & \text{if } k \neq 0 \\ \text{first time return to } 0 & k = 0 \end{cases} = \min\{n : n > 0 \text{ and } S_n = k\}$$

Corollary 101. *(corollary of reflection principle) If $p = q = 1/2$,*

$$P(T_k \leq n, S_n = k - l) = P(S_n = k + l)$$

for $k > 0$, $l \geq 0$ (similar result for $k < 0, l \leq 0$)

Corollary 102. *(second corollary of reflection) If $p = q = 1/2$, $k \neq 0$*

$$P(T_k \leq n) = 2P(S_n < k) + P(S_n = k) = P\{S_n \notin [-k, k)\}$$

Proof.

$$\begin{aligned} P(T_k \leq n) &= \sum_{b \in \mathbb{Z}} P(T_k \leq n,\ S_k = b) = \sum_{b \geq k} P(S_n = b) + \sum_{b < k} P(S_n = 2k - b) \\ &= P(S_n \geq k) + P(S_n > k) \\ &= P(S_n \geq k) + P(S_n < k) = P\{S_n \notin [-k, k)\} \end{aligned}$$

□

Theorem 103. *(Passage Time Theorem) If $k \neq 0$*

$$P(T_k = n) = \frac{|k|}{n} P(S_n = k) = \frac{|k|}{n} \binom{n}{\frac{n+k}{2}} p^{\frac{n+k}{2}} q^{\frac{n-k}{2}}$$

Proof. WLOG $k > 0$, how many paths with $T_k = n$ are there?

Claim: it is

$$\# \text{ paths of length } n-1 \text{ from } 0 \text{ to } (k-1)$$

$$-$$

$$\# \text{ subsets of paths of those paths visiting } k \text{ before } (n-1)$$

$$= \binom{n-1}{\frac{n-1+k-1}{2}} - \binom{n-1}{\frac{n-1+k+1}{2}} = \frac{k}{n} \binom{n}{\frac{n+k}{2}}$$

also each such path has $\frac{n+k}{2}$ up steps and $\frac{n-k}{2}$ down steps $\implies$ each such path has weight $p^{\frac{n+k}{2}} q^{\frac{n-k}{2}}$. □

Theorem 104. *(First Return to 0) For $q = p = 1/2$*

1)

$$P(T_0 = 2n) = P(T_1 = 2n-1) = \frac{1}{2n-1}\begin{pmatrix} 2n-1 \\ n \end{pmatrix} 2^{-(2n-1)}$$

2)

$$P(T_0 > 2n) = P(S_{2n} = 0) = \begin{pmatrix} 2n \\ n \end{pmatrix} 2^{-2n}$$

Proof. 1)

$$\begin{aligned} P(T_0 = 2n) &= P(S_1 \neq 0, ..., S_{2n-1} \neq 0, S_{2n} = 0) \\ &= 2P(S_1 = -1, S_2 - S_1 \leq 0, ..., S_{2n-1} - S_1 \leq 0, S_{2n} - S_1 = 1) \\ &= 2\underbrace{P(S_1 = -1)}_{1/2} P(S_2 - S_1 \leq 0, ..., S_{2n-1} - S_1 \leq 0, S_{2n} - S_1 = 1) \end{aligned}$$

Note $\hat{S}_n = S_{n-1} - S_1$, $n \geq 0$ is another simple random walk

$$\begin{aligned} \hat{S}_n &= P(\hat{S}_1 \leq 0, ...\hat{S}_{2n-2} \leq 0, \hat{S}_{2n-1} \leq 1) \\ &= P(S_1 \leq 0, ...S_{2n-2} \leq 0, S_{2n-1} = 1) \\ &= P(T_1 = 2n-1) \\ &= \frac{1}{2n-1}\begin{pmatrix} 2n-1 \\ \frac{2n-1+1}{2} \end{pmatrix} \left(\frac{1}{2}\right)^{\frac{2n-1+2n-1}{2}} \left(\frac{1}{2}\right)^{\frac{2n-1-(2n-1)}{2}} = \frac{1}{2n-1}\begin{pmatrix} 2n-1 \\ \frac{2n-1+1}{2} \end{pmatrix} \left(\frac{1}{2}\right)^{2n-1} \end{aligned}$$

2) By 1)

$$\begin{aligned} P(T_0 > 2n) &= P(T_1 > 2n-1) \\ &= 1 - P(T_1 \leq 2n-1) \\ &= 1 - P(S_{2n-1} \notin [-1, 1)) \qquad (4.1) \\ &= P(S_{2n-1} \notin \{-1, 0\}) \\ &= P(S_{2n-1} = -1) = P(S_{2n} = 0) \qquad (4.2) \end{aligned}$$

Equality (4.1) is from corollary 102. □

Corollary 105. *If $p = q = 1/2$ then*

$$P(T_k < \infty) = 1$$

for all $k \in \mathbb{Z}$. That is every state k is recurred at the end a.s.

Proof. Use Stirling formula $n! \sim (n/e)^n\sqrt{2\pi n}$, so

$$\binom{2n}{n} 2^{-2n} \sim \frac{1}{\sqrt{\pi n}}$$

Case 1: Let $k \neq 0$ we know by corollary 101

$$P(T_k > 2n) = P(S_{2n} \in [-k, k))$$

note also $P(S_{2n} = k) \leq P(S_{2n} = 0)\ \forall\, k$.

$$\begin{aligned} P(T_k > 2n) &\leq 2kP(S_{2N} = 0) \\ &= 2k\binom{2N}{N}2^{-2N} \\ &\sim 2k\frac{1}{\sqrt{\pi N}} = \frac{\text{const}}{\sqrt{N}} \to 0 \end{aligned}$$

Case 2: $k = 0$

$$P(T_0 > 2N) = P(S_{2N} = 0) \to 0$$

by the same argument. □

Corollary 106. $E[T_0] = \infty$, *that is it will take long time to return.*

Proof.

$$\begin{aligned} E[T_0] &= \sum_{k\geq 1} kP(T_0 = k) = \sum_{k\geq 1} P(T_0 \geq k) = \sum_{k\geq 0} P(T_0 > k) \\ &\geq \sum_{k\geq 0} \underbrace{P(S_{2k} = 0)}_{\geq\frac{\text{const}}{\sqrt{k}}} = \infty \end{aligned}$$

□

Proposition 107. *For* $0 < k \leq N$, $k + N$ *even*

$$P(S_1 > 0, ..., S_{N-1} > 0 | S_N = k) = \frac{k}{N}$$

i.e. independent of q, p.

Lecture 10 (2/20/14)

Theorem 108. *(Last visit to 0) Let* $p = q = 1/2$ *Fix* $N \in \mathbb{N}$ *and consider the least visit to 0 before 2N*

$$L_{2N} = \max\{m \leq 2N, S_m = 0\}$$

then

$$P(L_{2N} = 2n) = \binom{2n}{n}\binom{2N-2n}{N-n}\left(\frac{1}{2}\right)^{2N}$$

i.e. the $L_{2N}/2N$ has the so-called discrete arcsine distribution.

Proof.

$$\begin{aligned}
P(L_{2N} = 2n) &= P(S_{2n} = 0, S_{2n+1} \neq 0, ..., S_{2N} \neq 0) \\
&= P(S_{2n} = 0, S_{2n+1} - S_{2n} \neq 0, ..., S_{2N} - S_{2n} \neq 0) \\
&= P(S_{2n} = 0)P(S_{2n+1} - S_{2n} \neq 0, ..., S_{2N} - S_{2n} \neq 0)
\end{aligned}$$

So

$$\begin{aligned}
\hat{S}_j = S_{2n+j} - S_{2n} &= \sum_{i=1}^{j} X_{2n+i} \text{ is another random walk} \\
&= P(S_{2n} = 0)P(\hat{S}_1 \neq 0, ..., \hat{S}_{2N-2n} \neq 0) \\
&= P(S_{2n} = 0)P(S_1 \neq 0, ..., S_{2N-2n} \neq 0) \\
&= P(S_{2n} = 0)P(T_0 > 2N - 2n) \\
&= P(S_{2n} = 0)P(S_{2N-2n} = 0) \\
&= \binom{2n}{n}\left(\frac{1}{2}\right)^{2n}\binom{2N-2n}{N-n}\left(\frac{1}{2}\right)^{2N-2n}
\end{aligned}$$

□

Remark 109. 1) L_{2N} is typically close to 0 or $2N$.

2) Arcsine is motivated as follows

$$\text{theorem 108} + \text{ stirling} \implies P(L_{2N} = 2n) \sim \frac{1}{\pi\sqrt{n(N-n)}}$$

so for N large and $X \in [0, 1]$

$$\approx \sum$$

$$\begin{aligned}
P(\frac{L_{2N}}{2N} \leq X) = \sum_{n \leq XN} P(L_{2N} = 2n) &\approx \sum_{n \leq XN} \frac{1}{\pi\sqrt{n(N-n)}} \\
&\approx \sum_{n \frac{n}{N} \leq X} \frac{\frac{1}{N}}{\pi\sqrt{\frac{n}{N}(1 - \frac{n}{N})}} \approx \int_0^X \frac{dx}{\pi\sqrt{x(1-x)}} \\
&= \frac{2}{\pi} \arcsin\sqrt{X}
\end{aligned}$$

4.2 Gambler's Ruin

In this section $T_0 \equiv 0$.

we study $P(T_{a_1} < T_{a_2})$ for a_1, $a_2 \in \mathbb{Z}$, $a_1 < 0 < a_2$, so this is probability of hitting a_1 before a_2 (hitting a_2 means broken)

Notation 110. $b \in \{1, 2, ...\}$ total capital, $S \in \mathbb{N}$ $S \leq b$ gambler's initial capital, $b - S$ casino's initial capital.

In each iid round, one player pays \$1 to the other. P =probability of gambler winning one round.

$$\begin{aligned}\{T_{-S} < T_{b-S}\} &= \{\text{gambler is ruined before casino}\}\\ \{T_{-S} > T_{b-S}\} &= \{\text{casino is ruined before gambler }\}\end{aligned}$$

Proposition 111. *1) if $p \neq 1/2$*

$$P(T_{-S} < T_{b-S}) = \frac{\left(\frac{q}{p}\right)^S - \left(\frac{q}{p}\right)^b}{1 - \left(\frac{q}{p}\right)^b}$$

$$P(T_{-S} > T_{b-S}) = \frac{1 - \left(\frac{q}{p}\right)^S}{1 - \left(\frac{q}{p}\right)^b}$$

2) if $p = 1/2$, fair game

$$P(T_{-S} < T_{b-S}) = 1 - \frac{S}{b}$$

$$P(T_{-S} > T_{b-S}) = \frac{S}{b}$$

In particular

$$P(T_{-S} = T_{b-S}) = 0$$

The proof is more important than the result, because it illustrates an important trick.

Proof. For $S = 1, ..., b - 1$

$$\begin{aligned}h(S) &= P(T_{-S} < T_{b-S})\\ &= P(T_{-S} < T_{b-S}|X_1 = 1)P(X_1 = 1) + P(T_{-S} < T_{b-S}|X_1 = -1)P(X_1 = -1)\\ &= P(T_{-(S+1)} < T_{b-(S+1)})p + P(T_{-(S-1)} < T_{b-(S-1)})q\\ &= h(S+1)p + h(S-1)q\end{aligned}$$

To solve above difference equation, we need boundary condition

$$h(1) = 1, h(b) = 0$$

so

$$h(S) = \begin{cases} A + B\left(\frac{q}{p}\right)^S & p \neq q \\ A + BS & p = q \end{cases}$$

the formula for $P(T_{-S} > T_{b-S})$ are found similarly. □

What is the duration $T_{-S} \wedge T_{b-S}$?

Proposition 112.

$$E[T_{-S} \wedge T_{b-S}] = \begin{cases} \frac{S}{q-p} - \frac{b}{q-p}\frac{1-\left(\frac{q}{p}\right)^S}{1-\left(\frac{q}{p}\right)^b} & p \neq q \\ S(b-S) & p = q \end{cases}$$

Proof. Let $D_S = T_{-S} \wedge T_{b-S}$,

$$\begin{aligned} f(S) &= E[D_S] = \sum_{k\geq 0} kP(D_S = k) \\ &= \sum_k k\left(P(D_S = k|X_1 = 1)p + P(D_S = k|X_1 = -1)q\right) \\ &= \sum_k k\left(P(D_{S+1} = k)p + P(D_{S-1} = k)q\right) \\ &= E[1 + D_{S+1}]p + E[1 + D_{S-1}]q \\ &= 1 + f(S+1)p + f(S-1)q \end{aligned}$$

difference equation boundary condition $f(0) = f(b) = 0$ □

What if the casino has ∞ capital?

If $b = \infty$, then $T_{b-S} = \infty$ gambler cannot win, but maybe he can survive $T_{-S} = \infty$

$$P(T_{-S} < \infty) = \lim_{b\to\infty} P(T_{-S} < T_{b-S}) = \begin{cases} 1 & p \leq 1/2 \\ \left(\frac{q}{p}\right)^S & p > 1/2 \end{cases}$$

Example 113. Roulette $p = \frac{18}{37}$ for casino. A capital \$128 with \$1 round is enough for probability $> .999$.

The duration of the game for $b \to \infty$

$$E[T_{-S}] = \lim_{b\to\infty} E[T_{-S} \wedge T_{b-S}] = \begin{cases} \infty & p \geq 1/2 \\ S/(q-p) & p < 1/2 \end{cases}$$

Hence if the game is fair, the gamble will be ruined eventually but it may take very long.

5 Conditional Expectation

Lecture 11 (2/25/14)

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a given probability space, put $\mathcal{A} \subseteq \mathcal{F}$ sub $\sigma-$field. $\mathcal{A} = \{A_i, i \geq 1\}$

$$E[X|\mathcal{A}]^{\omega} = \sum_i \frac{E[X 1_{A_i}]}{P(A_i)} 1_{A_i(\omega)}$$

Best L^2 approximation of X is $\mathcal{A}$ mbl. Recall that

$$H := L^2(\Omega, \mathcal{F}, \mathcal{P}) = \mathcal{L}^2(\mathcal{F})$$

is a Banach space. $< X, Y >= E[XY]$ $X, Y \in \mathcal{L}^2$ defines an inner product, which induces the L^2 norm, that is $(H, < \cdot >)$ is a Hilbert space (a complete inner product space).

Theorem 114. *Let $\{0\} \neq U$ be a closed linear subspace of a Hilbert space. Then there exists a linear mapping $\Pi : H \to U$ s.t.*

1) $\Pi^2 = \Pi$ i.e. $\Pi^2(x) = \Pi(x)\ \forall\, x \in H$

2) $< x - \Pi(x), y >= 0\ \forall\, x \in H,\ y \in U$

moreover $\|x - \Pi(x)\| = \min_{y\in U} \|x - y\|$ i.e. $\Pi(x)$ is the best approximation to x in U.

We will show the proof later using completeness.

Proposition 115. *Let $X \in L^2(\mathcal{F})$ then $\exists Y \in \mathcal{L}^2(\mathcal{A})$ s.t.*

1) $E[X 1_A] = E[Y 1_A]\ \forall\, A \in \mathcal{A}$

2) $E[XZ] = E[YZ]\ \forall\, Z \in \mathcal{L}^2(\mathcal{A})$

Proof. Recall that $\mathcal{L}^2(\mathcal{A}) \subseteq \mathcal{L}^2(\mathcal{F})$ is complete $\implies$ it is closed $\implies$ $\exists\, \Pi : \mathcal{L}^2(\mathcal{F}) \to \mathcal{L}^2(\mathcal{A})$ orthogonal projection. Set $Y = \Pi(X)$ then

$$< X - Y, Z >= 0\ \forall\, Z \in \mathcal{L}^2(\mathcal{A}) \iff E[XZ] = E[YZ]$$

i.e. 1), 2) are a special case of property of Hilbert space. □

Remark 116. $X \geq 0$ a.s. $\implies Y \geq 0$ a.s.

Proof. take $A = \{Y < 0\}$ in 1) above. □

Lemma 117. *Let $Y_{1,2} \in L^1(\mathcal{A})$ if $E[Y_1 1_A] = E[Y_2 1_A]$ for all $A \in \mathcal{A}$,*

$$Y_1 = Y_2 \ a.s.$$

Proof. Let $A = \{Y_1 < Y_2\}$, then

$$E[(Y_1 - Y_2)1_A] = 0 \implies P(A) = 0$$

□

This is even true for $Y_{1,2} \in L^0_+$.

Corollary 118. *Let $X \in L^1(\mathcal{F})$, $\exists$ at most one $Y \in L^1(A)$ s.t.*

$$E[X1_A] = E[Y1_A] \qquad \forall A \in \mathcal{A}$$

Proposition 119. *Let $X \geq 0$ there exists an (a.s.) unique $\mathcal{A}$-mbl*

$$Y : \Omega \to \bar{\mathbb{R}}_+ \ s.t. \ E[X1_A] = E[Y1_A] \ \forall A \in \mathcal{A}$$

Proof. For $n \geq 0$ let $X_n = X \wedge n$ then $X_n \uparrow X$ a.s. and $X_n \in \mathcal{L}^2(\mathcal{F})$. Let $Y_n = \Pi(X_n)$ then

$$Y_0 \leq Y_1 \leq ... \leq .... \text{ a.s.}$$

define $Y = \lim_n Y_n$ then Y is a $\mathcal{A}-$mbl and for all $A \in \mathcal{A}$

$$\begin{aligned} E[y1_A] &= E[\lim_n Y_n 1_A] = \lim E[Y_n 1_A] \\ &= \lim E[X_n 1_A] = E[\lim X_n 1_A] = E[X1_A] \end{aligned}$$

□

5.1 Definition

Definition 120. If $X \geq 0$ is any RV then (a.s.) $\exists$ unique $\mathcal{A}-$mbl $Y \geq 0$ s.t. $E[X1_A] = E[Y1_A]\ \forall A \in \mathcal{A}$ is called the conditional expectation of X given $\mathcal{A}$ and denoted by

$$Y = E[X|\mathcal{A}]$$

Definition 121. For $X \in \mathcal{L}^1(\mathcal{F})$ define

$$E[X|\mathcal{A}] = E[X^+|\mathcal{A}] - E[X^-|\mathcal{A}]$$

Proposition 122. $E[\cdot|\mathcal{A}] : \mathcal{L}^1(\mathcal{F}) \to \mathcal{L}^1(\mathcal{A})$

1) is linear and continuous

2) is monotone i.e. $X > 0$ *a.s.* $\implies E[X|\mathcal{A}] \geq 0$ *a.s.*

3) $E[XZ|\mathcal{A}] = ZE[X|\mathcal{A}]\ \forall Z \in \mathcal{L}^\infty(\mathcal{A})$

Proof. 1) , 2)

$$\begin{aligned} X \in \mathcal{L}^1 \implies X^+, X^- \in \mathcal{L}^1 &\implies E[E[X^+|\mathcal{A}]] = E[X^+] \\ &\implies E[X^\pm|\mathcal{A}] \in \mathcal{L}^1 \implies E[X|\mathcal{A}] \in \mathcal{L}^1 \end{aligned}$$

Linearity follows from definition. Next show continuous

$$\begin{aligned} \|E[X|\mathcal{A}]\|_1 &= E[|E[X|\mathcal{A}]|] = E[E[X|\mathcal{A}]^+] + E[E[X|\mathcal{A}]^-] \\ &= E[E[X^+|\mathcal{A}]] + E[E[X^-|\mathcal{A}]] \\ &= E[X^+] + E[X^-] \\ &= E[|X|] = \|X\|_1 \end{aligned}$$

hence

$$\|E[X_1|\mathcal{A}] - E[X_2|U]\|_1 = \|X_1 - X_2\|_1$$

we showed it is Lipschitz continuous

3) For $Z = 1_B$, $B \in \mathcal{A}$. Claim

$$1_B E[X|\mathcal{A}] = E[X1_B|\mathcal{A}]$$

proof of claim: $\forall A \in \mathcal{A}$

$$E[(X1_B)1_A] = E[X1_{\underbrace{A \cap B}_{\in \mathcal{A}}}] = E[1_B E[X|\mathcal{A}]1_A]$$

By linearity 3) is okay for Z simple.

Let $Z \in \mathcal{L}^\infty(\mathcal{A})$ by approximation theorem $\exists$simple functions $Z_n \in \mathcal{L}^\infty(\mathcal{A})$ s.t. $Z_n \to Z$ a.s., we have

$$Z_n E[X|\mathcal{A}] \to ZE[X|\mathcal{A}] \text{ in } L^1 \text{ by DCT}$$

or

$$XZ_n \to XZ \text{ in } L^1$$

by linearity

$$E[XZ_n|\mathcal{A}] \to E[XZ|\mathcal{A}] \text{ in } L^1$$

by uniqueness of limit

$$E[XZ|\mathcal{A}] = ZE[X|\mathcal{A}]$$

□

Lecture 12 (2/27/14)

Definition 123. If Y is a rv we define

$$E[X|Y] = E[X|\sigma(Y)]$$

Proposition 124. *Let $(A_n)_{n\geq 1}$ be a $\mathcal{F}$-mbl, partition of Ω and let $\mathcal{A} = \sigma(A_{n\geq 1})$ then*

$$E[X|\mathcal{A}] = \sum_{n\geq 1} \frac{E[X1_{A_n}]}{P(A_n)} 1_{A_n}$$

Proof. Exercise. □

Definition 125. Conditional Probability is defined as follows

$$P(B|A) := E[1_B|\mathcal{A}], B \in \mathcal{F}$$

Note 126. If $\mathcal{A} = \sigma(A) = \{\varnothing, \Omega, A, A^C\}$

$$P(B|\mathcal{A}) = \frac{P(B\cap A)}{P(A)} 1_A + \frac{P(B\cap A^C)}{P(A^C)} 1_{A^C}$$

Notices the common notion to express above, although not consistent with measure theory is

$$P(B|A) = \frac{P(B\cap A)}{P(A)}$$

Next we discuss various convergence theorems in conditional expectation settings:

Theorem 127. *(MCT)*

$$0 \leq X_n \nearrow X \ a.s. \implies E[X_n|\mathcal{A}] \nearrow E[X|A] \ a.s.$$

Theorem 128. *(Faton)*

$$X_n \geq 0 \ a.s. \implies E[\liminf X_n|\mathcal{A}] \leq \liminf E[X_n|\mathcal{A}] \ a.s.$$

Theorem 129. *(DCT)*

$$X_n \to X \ a.s. \ |X_n| \leq Y \in L^1 \implies E[X_n|\mathcal{A}] \to E[X|\mathcal{A}] \ a.s. \ in \ L^1$$

also 1) continuous, 2) $E[X_n|\mathcal{A}] \leq E[|X_n|\,|\mathcal{A}] \leq E[|Y|\,|\mathcal{A}]$

Remark 130. The statement "$X_n \to X$ a.s. (X_n) UI $\implies E(X_n|\mathcal{A}) \to E[X|\mathcal{A}]$ a.s." is false, but it is true that

$$X_n \to X \text{ in } L^1 \implies \exists\, (X_{n_k}), E[X_{n_k}|\mathcal{A}] \to E[X|\mathcal{A}] \text{ a.s.}$$

5.2 Filtration & Martingale

Theorem 131. *(Conditional Jensen's Inequality) If* $X \in \mathcal{L}^1$, ψ *convex*

$$E[\psi(X)|\mathcal{A}] \geq \psi(E[X|\mathcal{A}]) \ a.s.$$

Proof. Similar as for usual Jensen. □

Corollary 132. $E[\cdot|\mathcal{A}]$ *is a contraction of* L^p, $p \in [1, \infty]$

$$\|E[X|\mathcal{A}]\|_p \leq \|X\|_p$$

Proof. For $p < \infty$ apply Jensen to $X \to |X|^p$, for $p = \infty$ $-\|X\|_\infty \leq X \leq \|X\|_\infty$ a.s.

$$\implies -\|X\|_\infty \leq E[X|\mathcal{A}] \leq \|X\|_\infty \text{ a.s.}$$

□

Example 133. Let $X \in \mathcal{L}^1(\mathcal{F})$

a) $\mathcal{A}$ is trivial i.e. $\mathcal{A} = \{\varnothing, \Omega\} \implies E[X|\mathcal{A}] = E[X]$ a.s.

Proof. $E[X]$ is $\mathcal{A}-$mbl,

$$E[E[X]1_A] = E[X1_A] \;\forall\, A \in \mathcal{A}$$

□

b) if $\sigma(X)$ and $\mathcal{A}$ are independent

$$E[X|\mathcal{A}] = E[X] \text{ a.s.}$$

Proof. Assume first that X is mbl

$$E[E[X]1_A] = E[X]E[1_A] = E[X1_A]\forall\, A \in \mathcal{A}$$

The second equality is by Independence. This is true for $E[X]$, $E[Y] \in L^1 \implies E[X]E[Y] = E[XY] \in L^1$. □

c) Tower property. Let $\mathcal{A}_0 \subseteq \mathcal{A} \subseteq \mathcal{F}$

$$E[E[X|\mathcal{A}]\mathcal{A}_0] = E[X|\mathcal{A}_0] \text{ a.s.}$$

Proof. $E[X|\mathcal{A}_0]$ is $\mathcal{A}_0-$mbl, $\forall\, A \in \mathcal{A}$

$$\begin{aligned} E[E[X|\mathcal{A}_0]1_A] &= E[E[X1_A|\mathcal{A}_0]] \\ &= E[X1_A] = E[E[X|\mathcal{A}]1_A] \end{aligned}$$

□

One can generalize $A_{n\geq 1} \searrow A_\infty$

$$...E[...E[E[X|\mathcal{A}_1]|\mathcal{A}_2]] = E[X|\mathcal{A}_\infty]$$

called backward most convergence

d) Let $(X_k)_{k\geq 1}$ iid $X_k \in \mathcal{L}^1$ $E[X_k] = 0$

$$S_n = X_1 + ... + X_n$$

$$\mathcal{F}_n = \sigma(X_1...X_n) = \sigma(S_1...S_n)$$

then $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq ....$ (called a **filtration**), and for all n

$$E[S_{n+1}|\mathcal{F}_n] = S_n$$

called a **martingale**.

Proof.

$$\begin{aligned} E[S_{n+1}|\mathcal{F}_n] &= E[S_n + X_{n+1}|\mathcal{F}] \\ &= E[S_n|\mathcal{F}_n] + E[X_{n+1}|\mathcal{F}_n] \\ &= S_n + E[X_{n+1}] = S_n \end{aligned}$$

the third equality is due to the facts that $\mathcal{F}_n-$mbl and X_{n+1}, $\mathcal{F}_n$ independent. □

5.3 Hilbert Projection

Lecture 13 (3/6/14)

Theorem 134. *If $U \subseteq H$ is a closed linear subspace in a Hilbert space, $\exists!$ orthogonal projection*

$$\pi : H \to H$$

Theorem 135. *Let $\phi \neq M \subseteq H$ be a closed, convex subset, for each $X \in H$ $\exists!$ $Y_* \in M$ s.t.*

$$\|X - Y_*\| = \inf_{Y \in M} \|X - Y_*\|$$

i.e. Y_* is the closest point in M.

Proof. (of theorem 135) Uniqueness: Let $Y_{1,2} \in M$ s.t.

$$\|Y_1\| = \|Y_2\| = \inf_{Y| \in M} \|Y\| = D$$

M convex $\implies \frac{Y_1+Y_2}{2} \in M \implies \|Y_1 + Y_2\| \geq 2D$. We have the parallelogram law

$$\begin{aligned} \|Y_1 - Y_2\|^2 + \|Y_1 + Y_2\|^2 &= 2(\|Y_1\|^2 + \|Y_2\|^2) \\ &= < Y_1 - Y_2, Y_1 - Y_2 > + < Y_1 + Y_2, Y_1 + Y_2 > \end{aligned}$$

Hence

$$\|Y_1 - Y_2\|^2 = 2\|Y_1\|^2 + 2\|Y_2\|^2 - \|Y_1 + Y_2\|^2 \leq 2D^2 + 2D^2 - (2D)^2 = 0 \implies Y_1 = Y_2$$

Existence: $D := \inf_{Y| \in M} \|Y\| \implies \exists Y_n \in M$ s.t.

$$\|Y_n\| \to D$$

using convexity + parallelogram inequality

$$\begin{aligned} \|Y_m - Y_n\|^2 &= 2(\|Y_m\|^2 + \|Y_n\|^2) - \|Y_m + Y_n\|^2 \\ &\leq 2(\|Y_m\|^2 + \|Y_n\|^2) - 4D^2 \to 0 \end{aligned}$$

$\implies (Y_n)$ is Cauchy, H is complete, there is a limit Y_* i.e.

$$\|Y_m - Y_n\| \to 0 \implies D = \lim \|Y_n\| = \|Y_*\|$$

□

Now prove theorem 134.

Proof. Let $M = U$ be a closed linear space, by theorem 135, given $X \in H$ define

$$\pi(X) = Y_*$$

as the closest point in H

1) it is clear that $\pi(\pi(X)) = \pi(X)$

2) it remains to show that Y_* is characterized by

$$< X - Y_*, Y >= 0 \quad \forall Y \in U$$

Let $Y = 0 \implies < X - Y_*, Y >= 0$, let $Y \in U/\{0\}$ for any $\lambda \in \mathbb{R}$

$$\begin{aligned} \|X - Y_*\|^2 &\leq \|X - (Y_* + \lambda Y)\|^2 =< X - (Y_* + \lambda Y), X - (Y_* + \lambda Y) > \\ &= \|X - Y_*\|^2 - \lambda < X - Y_*, Y > -\lambda^2 \|Y\|^2 \end{aligned}$$

choose

$$\lambda = \frac{< X - Y_*, Y >}{\|Y\|^2} \implies < X - Y_*, Y >^2 \leq 0 \implies < X - Y_*, Y >= 0$$

Conversely let $\tilde{Y} \in U$ satisfy $< X - \tilde{Y}, Y >= 0$, $\forall Y \in U$. Given $Y \in U$

$$\begin{aligned} \left\|X - \tilde{Y}\right\|^2 &\leq \left\|X - \tilde{Y}\right\|^2 + \left\|\tilde{Y} - Y\right\|^2 \\ &= \left\|(X - \tilde{Y}) + (\tilde{Y} - Y)\right\|^2 \qquad \because Y - \tilde{Y} \in U \\ &= \|X - Y\|^2 \end{aligned}$$

i.e.

$$\left\|X - \tilde{Y}\right\| = \inf_{Y \in U} \|X - Y\|$$

□

6 Measure Theory Part II

6.1 Change of Variable Formula

Theorem 136. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. (S, ρ) another measure space and let $X : \Omega \to S$ be mbl. Let $\mu = P \circ X^{+}$ be the distribution of X i.e.*

$$\mu(A) = P(X \in A) \quad A \in \rho$$

and let $h : (S, \rho) \to \mathbb{R}$ be mbl

1) if h is non-negative, then

$$\forall\, A \in \rho, \quad \int_{X^{-1}(\Lambda)} h(X(\omega))d\rho(\omega) = \int_A h(s)\mu ds \tag{6.1}$$

2) h is $\mu-$integrable iff $h \circ X$ is $P-$integrable, and in this cases (6.1) *holds.*

Note 137. In part

$$E[h(X)] = \int h d\mu$$

Proof. (Sketch) Let $h = 1_B$, $B \in \rho$ then

$$\int_{X^{-1}(A)} h(X)dp = P(X \in A \cap B) = \mu(A \cap B) = \int_A 1_B d\mu = \int_A h d\mu$$

By linearity, the theorem holds for simple functions. Then the general conclusion follows by approximation. □

Example 138. Let $X : \Omega \to \mathbb{R}$, $h : \mathbb{R} \to \mathbb{R}$, assume that X has a pdf, f i.e.

$$\mu(A) = \int_A f(x)dx$$

then

$$E[h(X)] = \int_{\mathbb{R}} h(x)f(x)dx$$

if $h \geq 0$ or integrable.

Example 139. Let $X : \Omega \to \mathbb{Z}$, $h : \mathbb{Z} \to \mathbb{R}$, let $P_n = P(X = n)$, $n \in \mathbb{Z}$ then

$$E[h(X)] = \sum_{n \in \mathbb{Z}} h(n)P_n$$

if $h \geq 0$ or integrable.

Example 140. Let $\lambda > 0$ $X : \Omega \to \{0, 1, 2, ...\}$ is Poisson $(\lambda)-$distribution if

$$p_k = P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \; k = 0, 1, 2, ...$$

then

$$\begin{aligned} E[X] &= \sum_{k\geq 0} kp_k = \sum_{k\geq 0} ke^{-\lambda}\frac{\lambda^k}{k!} = \lambda e^{-\lambda}\sum_{k\geq 1}\frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda e^{-\lambda}\sum_{k=0}\frac{\lambda^k}{k!} = \lambda \end{aligned}$$

also

$$\begin{aligned} E[X(X-1)] &= \sum_{k\geq 2} k(k-1)e^{-\lambda}\frac{\lambda^k}{k!} \\ &= \lambda^2 e^{-\lambda}\sum_{k\geq 0}\frac{\lambda^{k-2}}{(k-2)!} = \lambda^2 \end{aligned}$$

that is

$$E[X^2] = \lambda^2 + E[X] = \lambda^2 + \lambda$$

Definition 141. Variant is defined as

$$\mathrm{var}(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2$$

This is a way to remember Jensen. Apply Poisson

$$\mathrm{var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$$

6.2 Borel Cantelli Lemma

Lecture 14 (3/13/14)

Standard application of this lemma: Monkey types binary text handout, will complete the text if we wait long enough.

Definition 142. Consider a sequence $(A_{n\geq 1}) \subseteq \mathcal{F}$ in a probability space $(\Omega, \mathcal{F}, P)$,

$$\begin{aligned} \{A_n \text{ infinitely open (i.o.)}\} &= \limsup_n A_n := \bigcap_{k\geq 1}(\cup_{n\geq k} A_n) \\ &= \{\omega \in \Omega : \omega \in A_n \text{ for } \infty \text{ many } n\} \end{aligned}$$

Remark 143. $\limsup_n 1_A = 1_{\{A_n \text{ i.o.}\}} = 1_{\{\limsup A_n\}}$

Lemma 144. *(1st Borel Cantelli Lemma) Let* $(A_{n\geq 1}) \subseteq \mathcal{F}$ *then*

$$\sum_{n\geq 1} P(A_n) < \infty \implies P(A_n,\ i.o.) = 0$$

Proof.

$$\begin{aligned} P(A_n \text{ i.o.}) &= P(\bigcap_{k\geq 1} \cup_{n\geq k} A_n) \\ &= \lim_k P(\cup_{n\geq k} A_n) \leq \lim_k \sum_{n\geq k} P(A_n) \to 0 \end{aligned}$$

the second equality is given by the continuity of measure. □

The 2nd Borel Cantelli Lemma is partial converse of the 1st.

Lemma 145. *(2nd Borel Cantelli Lemma) Let* $(A_n) \subseteq \mathcal{F}$ *be independent then*

$$\sum P(A_n) = \infty \implies P(A_n\ i.o.) = 1$$

Proof. Let $1 \leq k \leq m$ recall that $(A^c_{n\geq 1})$ are again independent, hence

$$0 \leq 1 - P(\bigcup_{n=k}^{m} A_n) = P(\bigcap_{n=k}^{m} A_n^c) = \prod_{n=k}^{m} P(A_n^c) = \prod_{n=k}^{m} (1 - P(A_n))$$

Since $1 - X \leq e^{-X}$ for $0 \leq X \leq 1$, we get

$$1 \geq P(\bigcup_{n=k}^{m} A_n) \geq 1 - \prod_{n=k}^{m} e^{-P(A_n)} = 1 - e^{\overbrace{-\sum_{n=k}^{m} P(A_n)}_{\to\infty}} \quad \text{as } m \to \infty$$

Use continuity of measure

$$P(\bigcup_{n=k}^{\infty} A_n) = \lim_m P(\bigcup_{n=k}^{\infty} A_n) \geq 1 \text{ for every } k$$

$\implies$

$$P(A_n \text{ i.o.}) = P(\bigcap_k \bigcup_{n=k}^{\infty} A_n) = 1$$

□

Example 146. Let $S = (a_0, a_1, ..., a_n)$ be a binary sequence of length n. A monkey is given a 0/1 keyboard and he produces an iid sequence $(X_{k\geq 1})$ with $P(X_k = 1) = P(X_k = 0) = 1/2$, then let

$$A_k = \{X_k = a_0, ..., X_{k+n} = a_n\}$$

$$P(A_k \text{ i.o.}) = 1$$

Proof. $P(A_k) = 2^{-n+1} \implies \sum P(A_k) = \infty$. The A_k are not independent but $(A_{(n+1)k})_{k\geq 1}$ are independent and

$$\sum_k P(A_{(n+1)k}) = \infty$$

by the 2nd Borel Cantelli Lemma

$$P(A_k \text{ i.o.}) = 1$$

□

6.3 Monotone Classes and Dynkin System

Definition 147. Let $\mathcal{C}$ be a non-empty collection of subsets of Ω.

1) $\mathcal{C}$ is a monotone class if it's a closed under countable increasing unions and countable decreasing interaction

2) $\mathcal{C}$ is a $\Pi-$system if it is stable under intersection

3) $\mathcal{C}$ is a $\lambda-$system if $\Omega \in \mathcal{C}$, $\mathcal{C}$ is closed under countable increasing unions, and $A, B \in \mathcal{C}$ and if $B \subseteq A$, then $A/B \subseteq \mathcal{C}$.

Note if $(\mathcal{C}_{i\in I})$ are monotone classes, so is $\bigcap_{i\in I} \mathcal{C}_i$, same for $\lambda-$system. Hence given any family $(A_{j\in J}) \in 2^{\Omega}$, $\exists$ a minimal monotone class $m(A_j)_{j\in J}$ and a minimal $\lambda-$system $\lambda(A_j)_{j\in J}$ containing (A_j)

Lemma 148. *If $\mathcal{C}$ is a monotone class, it is a field.*

Proof. Exercise. □

Theorem 149. *(Monotone Class theorem for Sets) Let $\mathcal{C}$ be a field, then*

$$m(\mathcal{C}) = \sigma(\mathcal{C})$$

Proof. Let $M = m(\mathcal{C})$, since any field is a monotone class.

$$M \subseteq \sigma(\mathcal{C})$$

in light of the lemma, it remains to show that M is a field.

Fix $G \in \mathcal{C}$ and let

$$\mathcal{C}_G = \{\mathcal{F} \in M : \mathcal{F}/G, G/\mathcal{F}, \mathcal{F} \cap G \in M\}$$

then $\phi \in \mathcal{C}_G$, $G \in \mathcal{C}_G$ and $\mathcal{C}_G$ is a monotone class.

Moreover $\mathcal{C} \subseteq \mathcal{C}_G$ since $\mathcal{C}$ is a field $\implies M \subseteq \mathcal{C}_G \implies M = \mathcal{C}_G$. Now let $G \in M$, the above shows $G \in \mathcal{C}_F \ \forall F \in \mathcal{C}$ which implies $F \in \mathcal{C}_G$ by symmetry. Again $\mathcal{C}_G$ is a monotone class $\implies M \subseteq \mathcal{C}_G \implies M = \mathcal{C}_G$. This holds for every $G \in M$ meaning that M is a field. □

Lemma 150. *If $\mathcal{C}$ is a $\Pi-$system and a $\lambda-$system, then $\mathcal{C}$ is a $\sigma-$field.*

Proof. Exercise. □

Theorem 151. *(Dynkin) Let $\mathcal{C}$ be a $\Pi-$ system, then $\lambda(\mathcal{C}) = \sigma(\mathcal{C})$.*

Proof. Clearly $\lambda(\mathcal{C}) \subseteq \sigma(\mathcal{C})$, in view of the lemma it suffices to show that

$$\mathcal{L} := \lambda(\mathcal{C}) \text{ is a } \Pi-\text{system.}$$

Let $\mathcal{L}_1 = \{A \in \mathcal{L} : A \cap B \in \mathcal{L} \text{ for all } B \in \mathcal{C}\}$, then $\mathcal{C} \subseteq \mathcal{L}_1$, since $\mathcal{C}$ is a $\Pi-$system. Using that $\mathcal{L}$ is a $\lambda-$system, $\mathcal{L}_1$ is a $\lambda-$system $\implies \mathcal{L} = \mathcal{L}_1$ by minimality.

Let $\mathcal{L}_2 = \{A \in \mathcal{L} : A \cap B \in \mathcal{L} \text{ for all } B \in \mathcal{L}\}$, then $\mathcal{L} = \mathcal{L}_1 \implies \mathcal{C} \subseteq \mathcal{L}_2$. Again $\mathcal{L}_2$ is a $\lambda-$system $\implies \mathcal{L}_2 = \mathcal{L}$ i.e. $\mathcal{L}$ is a $\Pi-$system. □

Theorem 152. *(Determination of Measures) Let $\mathcal{C}$ be a $\Pi-$system s.t. $\sigma(\mathcal{C}) = \mathcal{F}$. If P, Q are finite measure on $\mathcal{F}$ and $P = Q$ on $\mathcal{C}$, then $P = Q$ on $\mathcal{F}$.*

Corollary 153. *If P, Q are finite measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ s.t. $P([a,b]) = Q([a,b]) \ \forall a < b$, then $P = Q$ on $\mathcal{B}(\mathbb{R})$.*

Example 154. (counterexamples)

$$\begin{aligned}
\Omega &= \{a, b, c, d\} \\
\mathcal{C} &= \{\{a,b\}, \{b,c\}, \{c,d\}, \Omega\} \\
P(a) &= P(b) = P(c) = P(d) = \frac{1}{4} \\
Q(a) &= Q(c) = \frac{1}{3} Q(b) = Q(d) = \frac{1}{6}
\end{aligned}$$

$\implies P = Q$ on $\mathcal{C}$, $\mathcal{C}$ generates 2^Ω but $P \neq Q$.

Lecture 15 (3/25/14)

Proof. (of theorem 152) Recall if $\mathcal{C}$ is a $\Pi-$system, then $\lambda(\mathcal{C}) = \sigma(\mathcal{C})$. Consider $\mathcal{L} = \{A \in \mathcal{F}, P(A) = Q(A)\}$, then $\mathcal{C} \subseteq \mathcal{L}$. Moreover $\mathcal{L}$ is a $\lambda-$system

1) $P(\Omega) = 1 = Q(\Omega)$ so $\Omega \in \mathcal{L}$

2) Let $A, B \in \mathcal{L}$,$A \subseteq B$

$$P(B/A) = P(B) - P(A) = Q(B) - Q(A) = Q(B/A)$$

hence $B/A \in \mathcal{L}$

3) Let $(A_n)_{n\in\mathbb{N}} \in \mathcal{L}$ s.t. $A_n \subseteq A_{n+1}$

$$\begin{aligned} P(\bigcup_{n=1}^{\infty} A_n) &= \lim_{N\to\infty} P(\bigcup_{n=1}^{N} A_n) = \lim_{N\to\infty} P(A_N) \\ &= \lim_{N\to\infty} Q(A_N) = Q(\bigcup_{n=1}^{N} A_n) \end{aligned}$$

So $\bigcup_{n=1}^{N} A_n \in \mathcal{L}$ and $\mathcal{L}$ is a $\lambda-$system so $\lambda(\mathcal{C}) \subseteq \mathcal{L}$ but

$$\lambda(\mathcal{C}) = \sigma(\mathcal{C}) = \mathcal{F}$$

so $\mathcal{F} = \mathcal{L}$ so $\forall\, A \in \mathcal{F}$, $P(A) = Q(A)$. ☐

Theorem 155. *Let I, $J \in \mathcal{F}$ be independent $\Pi-$system, then $\sigma(I)$ and $\sigma(J)$ are independent.*

Proof. Exercise. ☐

Theorem 156. *(Monotone Class Theorem for Functions) Let M be a class of bounded function $\Omega \to \mathbb{R}$ which is closed under multiplication (if $f, g \in M$, so is fg) Let M be a vector space of function $\Omega \to \mathbb{R}$ containing M. If $(f_n) \in M$ with $0 \leq f_1 \leq f_2 \leq ...$ s.t. $\mathcal{F} = \lim_{n\to\infty} f_n$ is bounded, then $f \in M$.*

Put this theorem in words: M contains all bounded $\sigma(M)$ - mbl functions.

Proof. Exercise. ☐

6.4 Product Measure

Kernels and Fubinis Theorem

Definition 157. Let (S_1, ρ_1) and (S_2, ρ_2) be measurable spaces. A (stochastic) kernel $K(X_1, dX_2)$ from (S_1, ρ_1) and (S_2, ρ_2) is a mapping

$$K : S_1 \times S_2 \to [0, 1]$$

s.t. $\forall X_1 \in S_1$ $K(X_1, \cdot)$ is a probability measure, and $\forall A_2 \in S_2$, the function $K(X_1, A_2)$ is ρ_1-measurable.

Probability interpretation: S_1 and S_2 are states of a random system at time 1 and 2. $K(X_1, \cdot)$ is the probability of the state of the system at time 2 given that the system was in state X_1 at time 1.

Example 158. $\forall X_1 \in S_1$, let $K(X_1, \cdot) = P_2$ fixed probability measure on S_2, i.e. no dependence on X_1.

Example 159. Suppose $S_1 = S_2 = S$ which is a finite set $\{o_1, ..., o_m\}$ and $\rho_1 = \rho_2 = P(S)$. Any probability measure on S is determined by $P(\{o_1\}), ..., P(\{o_m\})$. Then K is given by a matrix

$$\tilde{K} = (\tilde{K}_{ij})_{1 \leq i,j \leq n}$$

where $\tilde{K}_{ij} = K(o_i, \{o_j\})$, $\tilde{K}$ is stochastic matrix, namely

$$\tilde{K}_{ij} \geq 0 \text{ and } \sum_{j=1}^{n} K(o_i, \{o_j\}) = 1 \, \forall \, o_i \in S$$

So $\tilde{K}_{ij}$ is the probability to jump from o_i to o_j.

Let $\Omega = S_1 \times S_2$ with $\sigma-$algebra $\sigma(\rho_1 \times \rho_2)$. Let P be a probability on (S_1, ρ_1) and let K be a Kernel from (S_1, ρ_1) to (S_2, ρ_2), we want to construct a “product” measure $P = P_1 \times K$.

Discrete Case

Let S_i be a countable sets and $\rho_i = P(S_i)$ then Ω is countable and P is given by the weights

$$P(\{X_1, X_2\}) = P(\{X_1\})K(X_1, \{X_2\})$$

then $\forall f : \Omega \to \mathbb{R}_+$

$$\begin{aligned} E^p(f) &= \int_\Omega f dp = \sum_{X_1, X_2} f(X_1, X_2) P(\{(X_1, X_2)\}) \\ &= \sum_{X_1} \sum_{X_2} f(X_1, X_2) K(X_1, \{X_2\}) P_1(\{X_1\}) \\ &= \int_{S_1} \left(\int_{S_2} f(X_1, X_2) K(X_1, X_2) \right) P_1(dX_1) \end{aligned}$$

General Case

On $\Omega = S_1 \times S_2$ we use the product $\sigma-$field $F : \rho_1 \times \rho_2 = \sigma(\{A_1 \times A_2 | A_i \in \rho_i\})$. This is the $\sigma-$algebra generated by "rectangle" $A_1 \times A_2$.

Theorem 160. *(Fubini)* $\exists!$ *probability measure* P *on* $(\Omega, \mathcal{F}) = (S_1 \times S_2, \rho_1 \times \rho_2)$ *s.t.*

$$\int_\Omega f dP = \int_{S_1} \left(\int_{S_2} f(X_1, X_2) K(X_1, dX_2) \right) P(dX_1) \tag{6.2}$$

$\forall f \in \mathcal{L}_0^+(\Omega, \mathcal{F})$. *In particular*

$$P(A) = \int K(X_1, AX_1) P_1(dX_1) \ \forall A \in \mathcal{F}$$

where $AX_1 = \{X_2 \in S_2 | (X_1, X_2) \in A\}$ *is the* X_1-*section of* A. *In particular if* $A = A_1 \times A_2$,

$$P(A) = \int_{A_1} K(X, A_2) P_1(dX_1)$$

Remark 161. 1) (6.2) also holds for $f \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$

2) this can be extended to $\sigma-$finite- measure.

3) Special case if $K(X_1, \cdot) = p_2 \ \forall X_1$, we have

$$P = P_1 \times P_2$$

a product of two measures. Then

$$(P_1 \times P_2)(A_1 \times A_2) = P_1(A_1) P_2(A_2)$$

and we get the classical Fubini theorem

$$\int_\Omega f d(P_1 \times P_2) = \int \left(\int f dP_1 \right) dP_2 = \int \left(\int f dP_2 \right) dP_1$$

$\forall f \in \mathcal{L}_+^0$ or $f \in L^1(P)$.

Lecture 16
(3/27/14)

Lemma 162. *If $f:\Omega \to \bar{\mathbb{R}}$ is $\rho_1 \times \rho_2-$mbl and $X_1 \in S_1$ then the section*

$$f_{X_1}(\cdot) := f(X_1, \cdot) : (S_2, \rho_2) \to \bar{\mathbb{R}}$$

is mbl.

Proof. The mapping

$$f_{X_1}(\cdot) : X_2 \mapsto (X_1, X_2)$$

from (S_2, ρ_2) to $(S_1 \times S_2, \rho_1 \times \rho_2)$ is mbl.

$$\psi_{X_1}^{-1}(A_1 \times A_2) = \begin{cases} \varnothing & X_1 \notin A_1 \\ A_2 & X_1 \in A_1 \end{cases}$$

then the composite $F_{X_1} = f \circ \psi_{X_1}$ is mbl. □

Corollary 163. $A \in \rho_1 \times \rho_2 \implies A_{X_1} \in \rho_2 \;\forall\, X_1 \in S_1$.

Proof. (of theorem 160 Fubini) 1) Uniqueness: P is unique because the rectangle $A_1 \times A_2 \in \rho_1 \times \rho_2$ generates $\rho_1 \times \rho_2$ and form a $\Pi-$system.

2) $P(A)$ is well-defined. By the corollary AX_1 is mbl. So

$$P(A = \int K(X_1, AX_1) P_1(dX_1)$$

make sense. P is a probability measure, so

i)

$$P(\Omega) = \int K(X_1, \rho_2) P_1(dX_1) = 1$$

ii) $(A^n)_{n\geq 1} \subseteq \mathcal{F}$ adjoin $\implies (\cup_n A^n)_{X_1} = \cup_n A^n_{X_1}$ and

$$\begin{aligned} P(\bigcup_n A^n) &= \int K(X_1, \bigcup_n A^n_{X_1} P_1)(dX_1) \\ &= \sum_n \int K(X_1, A^n_{X_1}) P_1(dX_1) = \sum_n P(A^n) \end{aligned}$$

$\implies$ P is a $\sigma-$mbl.

3) Let $F \in \mathcal{L}^0_+$ Claim:

$$X_1 \mapsto \int f(X_1, X_2) K(X_1, dX_2)$$

is ρ_1-mbl.

proof of claim: By approximation and linearity, it suffices to consider

$$\mathcal{F} = 1_A,\ A \in \mathcal{F}$$

then

$$\int f(X_1, X_2) K(X_1, dX_2) = K(X_1, AX_1)$$

let

$$\mathcal{C} = \{A \in \mathcal{F} : X_1 \mapsto K(X_1, AX_1) \text{ is mbl}\}$$

If $A = A_1 \times A_2 \in \rho_1 \times \rho_2$ then $K(X_1, A_{X_1}) = 1_{A_1}(X_1) K(X_1, A_2)$ is mbl, so $A \in \mathcal{C}$, also $\mathcal{C}$ is a $\lambda-$system. By Dynkin $\mathcal{C} = \mathcal{F} \implies$ claim is proved.

4) The formula (6.2) is true by definition for $f = 1_A$, $A \in \mathcal{F}$. By linearity and approximation, it holds for $f \in \mathcal{L}^0_+$. □

Next we study the "converse" to Fubini:

Theorem 164. *(decomposition of measure) Let P be any probability measure on $(\Omega, \mathcal{F}) = \{S_1 \times S_2, \rho_1 \times \rho_2\}$*

$$P_1(A_1) := P(A_1 \times S_2), \quad A_1 \in \rho_1$$

If S_2 is a complete metric space and $\rho_2 = \mathcal{B}(S_2)$ then there exists a Kernel K s.t.

$$P = P_1 \times K$$

Proof. Exercise. □

We will show how to construct K in two special cases:

i) Discrete Case: If S_1, S_2 are countable, set

$$K(X_1, A_2) = \frac{P\{(X_1, X_2) : X_2 \in A\}}{P\{(X_1, X_2) : X_1 \in S_2)\}}$$

If the denomination is 0, let K be any probability on ρ_2.

ii) Conditional density: suppose we are given ($\sigma-$finite) measure m_1, m_2 on S_1, S_2 and P admits a density wrt $m_1 \times m_2$

$$P(A) = \iint f(X_1 X_2) m_1(dX_1) m_2(dX_2)$$

Note that the marginal P_1 has a density given by

$$f_1(X_1) = \int f(X_1, X_2) m_2(dX_2)$$

Set

$$K(X_1, A_2) = \begin{cases} \frac{f(X_1,X_2)}{f_1(X_1)} m_1(dX_2) & f_1(X_1) \neq 0 \\ \text{any prob on } \rho_2 & f_1(X_1) = 0 \end{cases}$$

then $P = P_1 \times K$.

One can call

Definition 165.

$$f(X_2|X_1) := \frac{f(X_1, X_2)}{f_1(X_1)}$$

conditional density.

6.5 Infinite Product

It has applications to random walk and quantum field theory.

For $i = 0, 1, 2, ...$ let (S_i, ρ_i) be a mbl space and let

$$(S^n, \rho^n) = (S_0 \times ... \times S_n, \rho_0 \times ... \times \rho_n)$$

Given a probability P_0 on (S_0, ρ_0), $\forall\, n \geq 1$ we define a kernel K_n from (S^{n-1}, ρ^{n-1}) to (S_n, ρ_n). This gives a discrete time dynamic system with state space S_i at time i, P_0 initial distribution,

$$K_n(X_0, ..., X_{n-1}, A_n)$$

the probability of being in A_n at time n given that $X_0, ..., X_{n-1}$ were the states at $t = 0, 1, ..., n-1$.

Fubini should have the following form: for each $n \geq 1$, we have a probability P^n on (S^n, ρ^n)

$$\begin{aligned} P^0 : &= P_0 \\ p^n &= P^{n-1} \times K_n = P_0 \times K_1 \times ... \times K_n, \qquad n \geq 1 \end{aligned}$$

If $f \in \mathcal{L}^0_+(S^n, \rho^n)$, then

$$\int f dP^n = \int P_0(dX_0) \int K_1(X_0, dX_1) ... \int K_n\left((X_0, ..., X_{n-1}), dX_n\right) f(X_0, ..., X_n)$$

Let $\Omega = \prod_{i=0}^{\infty} S_i = \{w = (X_0, X_1, ...), X_i \in S_i\}$, this is the space of possible "trajectories".

Let $X_n : \Omega \to S_n$ be the canonical projection i.e.

$$X_n(w) = X_n$$

if $w = (X_0, X_1, ...)$. Let $\mathcal{F}_n = \sigma(X_0, X_1, ..., X_n)$ "event observable up to time n" and

$$\mathcal{F} = \sigma(X_0, X_1, ...) = \sigma(\bigcup_{n \geq 0} \mathcal{F}_n)$$

We want to construct a probability P on $(\Omega, \mathcal{F})$ that behaves like P^n up to time n.

More precisely: each $A \in \mathcal{F}_n$ is of the form

$$A = A^n \times S_{n+1} \times S_{n+2} \times ...$$

for some $A^n \in \rho^n$, and we want

$$P(A^n \times S_{n+1} \times ...) = P^n(A^n)$$

$\forall\, A^n \in \rho^n$, $n \geq 0$.

Lecture 17 (3/31/14)

Theorem 166. *(Ionescu-Tulcea) $\exists$ a unique probability P on $(\Omega, \mathcal{F})$ s.t.*

$$P(A^n \times S_{n+1} \times S_{n+2} \times ...) = P^n(A^n) \tag{6.3}$$

$\forall\, A^n \in \rho^n$, $n \geq 0$ and more generally

$$\int f(X_0, X_1, ..., X_n) dP = \int P_0(dX_0) K_1(X_0, dX_1) ... K_n(X_0 ... X_{n-1}, dX_n) f(X_0, ..., X_n)$$

for all $f \in \mathcal{L}^0_+(S^n, \rho^n)$.

Proof. (Sketch) Uniqueness: P is determined by (6.3) on finite rectangles, they form a $\pi-$system generating $\mathcal{F} \implies P$ is unique.

Existence: (6.3) defines P on the field

$$\bigcup_{n \geq 0} \mathcal{F}_n = \xi$$

One checks that P is $\sigma-$additive on ξ and uses the Caratheodory theorem. $\square$

Definition 167. The stochastic process $(X_{n\geq 0})$ on $(\Omega, \mathcal{F}, P)$ has initial distribution P_0 and transition law (or kernel)

$$K_n(X_0, ..., X_{n-1}, A_n) = P[X_n \in A_n | X_0 = x_0, ..., X_{n-1} = x_{n-1}]$$

if S_i is countable, one can define the RHS directly and show this.

If $K_n(X_0, ..., X_{n-1}, \cdot) = K_n(X_{n-1}, \cdot)$, i.e. K_n depends only on the "current" state X_{n-1}, X is a **Markov** process.

If furthermore $(S_n, \rho_n) = (S, \rho)$ and $K_n = K$, we say that X is time-homogeneous.

Proposition 168. *Let P_n be the $n-$th marginal i.e.*

$$P_n(A) := P(X_n \in A), \ A \in \rho^n$$

then $(X_{n\geq 0})$ are independent (under P) iff

$$P = \prod_{n\geq 0} P_n$$

i.e. $K_n = P_n \ \forall n \geq 1$.

Proof. Let $\tilde{P} = \prod_{n\geq 0} P_n$. (X_n) are independent iff

$$P(X_0 \in A_0, ..., X_n \in A_n) = \prod_{i=0}^{n} P_i(X_i \in A_i) = \tilde{P}(X_0 \in A_0, ..., X_n \in A_n)$$

iff (because of rectangles form $\pi-$system)

$$P = \tilde{P}$$

□

Example 169. Let $S_n = \{\pm 1\} \ \forall n \geq 0, \ S_0 = \{0\}$,$P_n(\{1\}) = p = 1 - P_n(\{-1\})$

Let $\Omega = \pi S_n$, $P = \pi P_n \implies (X_{n\geq 1})$ are iid $P(X_0 = \pm 1) = \begin{cases} p \\ 1-p \end{cases}$,
i.e. $S_n = X_0 + ... + X_n$ is a simple random walk.

6.6 Caratheodory Extension and Existence of Lebesgue-Stieltjes Measure

Theorem 170. *(Caratheodory) Let ξ be a field of subsets of Ω and let $\nu : \xi \to [0, \infty]$ satisfy*

1) $\nu(\phi) = 0$

2) ν is $\sigma-$finite

3) ν is σ additive on ξ if $(A_{n\geq 1}) \subseteq \xi$ are disjoint and $\bigcup_{n\geq 1} A_n \in \xi$, then

i)

$$\nu(\bigcup_n A_n) = \sum_n \nu(A_n)$$

ii) ν extents uniquely to a $\sigma-$finite measure μ on $\sigma(\xi)$. Moreover

$$\mu(B) = \inf\{\sum_n \nu(A_n), A_n \in \xi, B \in \bigcup_n A_n\} \tag{6.4}$$

$B \in \sigma(\xi)$.

Proof. uniqueness: $\pi-$system; existence: one checks that (6.4) defines a measure. ☐

Example 171. Let $F : \mathbb{R} \to \mathbb{R}$ be non-decreasing and right-continuous, define

$$\xi = \{\bigcup_{j=1}^{n} (a_j, b_j] : -\infty \leq a_j \leq b_j \leq \infty, n \in \mathbb{N}\}$$

Lemma 172. *Let $A \in \xi$ then A is a union of disjoint intervals $(a_j, b_j]$, $1 \leq j \leq n$ and*

$$\nu(A) := \sum_{j=1}^{n} F(b_j) - F(a_j)$$

defines a $\sigma-$finite, $\sigma-$additive function on ξ.

Corollary 173. *(Lebesgue-Stielties) Let $F : \mathbb{R} \to \mathbb{R}$ be non-decreasing and right continuous, there exists a unique measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ s.t.*

$$\mu((a, b]) = F(a) - F(b)$$

$\forall\, a < b$. If $F(x) = x$ this is the Lebesgue measure.

Proof. Note that ξ is a field and $\sigma(\xi) = \mathcal{B}(\mathbb{R})$ in view of the lemma 172, we can apply the Caratheodory theorem. ☐

Now we prove lemma 172.

Proof. 1) ν is well-defined: $\nu(A)$ does not depend on the choice of the disjoint intervals.

2) $\sigma-$finite: $\mathbb{R} = \cup_n(-n, n]$ and $\nu(-n, n] = F(n) - F(-n) \in \mathbb{R}$

3) claim: if $-\infty < a < b < \infty$ and $(a, b] = \cup_{n\geq 1}(a_n, b_m]$ is a countable disjoint union, then

$$\nu(a, b] = \sum_{n\geq 1} \nu(a_n, b_n]$$

proof of claim: finite additive is clear;

$$\nu(a, b] \geq \sum_{n\geq 1} \nu(a_n, b_n]$$

follows from definition. Let $0 < \xi < b - a$, choose $\tilde{b}_n > b_n$ s.t.

$$F(\tilde{b}_n) - F(b_n) \leq \xi 2^{-n}$$

which is possible because F is right continuous. Then $\cup_n(a_n, \tilde{b}_n)$ form a covering of $[a+\epsilon, b)$, which is compact, so $\exists$ a finite sub cover $(a_{n_1}, \tilde{b}_{n_1})...(a_{n_e}, \tilde{b}_{n_e})$, so

$$\begin{aligned} \nu(a+\epsilon, b] &= F(b) - F(a+\epsilon) \leq \sum_{j=1}^{e} \nu(a_{n_j}, \tilde{b}_{n_j}] \\ &\leq \epsilon + \sum_{j=1}^{e} \nu(a_{n_j}, b_{n_j}] \end{aligned}$$

letting $\epsilon \to 0$, using right continuous,

$$\nu(a, b] \leq \sum_{j=1}^{e} \nu(a_{n_j}, b_{n_j}]$$

4) ν is $\sigma-$additive on ξ: reduces to 3). □

7 Limit Theorems

7.1 Convergence in Distribution

Lecture 18 (4/6/14)

Definition 174. Let μ_n, $n \geq 0$ be probability on $\mathbb{R}$, then μ_n converges weakly to μ_0 (denoted $\mu_n \Rightarrow \mu_0$) if

$$\int g d\mu_n \underset{n\to\infty}{\longrightarrow} \int g d\mu_0$$

$\forall\, g \in C_b(\mathbb{R})$, bounded continuous function $\mathbb{R} \to \mathbb{R}$. Also called vague convergence.

Definition 175. Let $(\Omega_n, \mathcal{F}, P_n)$, $n \geq 0$ be probability spaces. $X_n : \Omega_n \to \mathbb{R}$ r.v. then X_n converges to X_0 in distribution (or in law$X_n \xrightarrow{\mathcal{L}} X_0,\ X_n \xrightarrow{\mathcal{D}} X_0,\ X_n \implies X_0$) if the corresponding distribution converges weakly i.e.

$$\underbrace{E^{P_n}[g(X_n)]}_{\int g\mu_n} \longrightarrow \underbrace{E^{P_0}[g(X_0)]}_{\int g\mu_0}, \forall\, g \in \mathcal{C}_b(\mathbb{R})$$

Theorem 176. *Let X_n, $n \geq 0$ be rv's on $(\Omega, \mathcal{F}, P)$ then*

$$X_n \xrightarrow{P} X_0$$

converges in measure implies

$$X_n \xrightarrow{\mathcal{L}} X_0$$

Proof. Let $X_n \xrightarrow{P} X_0$, then

$$g(X_n) \xrightarrow{P} g(X_0)$$

for all $g \in C(\mathbb{R})$. If g is bounded, $\{g(X_n)\}$ is UI $\implies$

$$E^P[g(X_n)] \to E^P[g(X_0)]$$

□

Theorem 177. *Let μ_n, μ be probabilities on $\mathbb{R}$ and $F_n(X) = \mu_n((-\infty, X])$, $F(X) = \mu((-\infty, X])$. The following are equivalent:*

1) $\mu_n \implies \mu$

2) $\int g d\mu_n \to \int g d\mu$, $\forall\, g \in C_b^\infty(\mathbb{R})$

3) $F_n(X) \to F(X)$, $\forall\, X \in C = \{X \in \mathbb{R} : F \text{ is continuous at } X\}$

Proof. 1) $\implies$ 2) clear

2) $\implies$ 3) let $X \in C$, $\delta > 0$, let $h \in C_b^\infty$ be s.t. $0 \le h \le 1$, $h = 1$ on $(-\infty, X]$, $h = 0$ on $[X+\delta, +\infty)$.

Then $F_n(X) \le \int h d\mu \le F_n(X+\delta)$, $F(X) \le \int h d\mu \le F(X+\delta)$, then

$$\limsup_{n\to\infty} F_n(X) \le \lim \int h d\mu_n \overset{2)}{\implies} \int h d\mu \le F(X+\delta)$$

$\overset{\text{let } \delta \downarrow 0}{\implies} \limsup F_n(X) \le F(X) \because F$ is RC. Also

$$\begin{aligned} F_n(X-\delta) &\le \int h(\cdot+\delta) d\mu_n \le F_n(X) \\ F(X-\delta) &\le \int h(\cdot+\delta) d\mu_n \le F(X) \end{aligned}$$

$\implies \liminf_{n\to\infty} F_n(X) \ge \lim \int h(\cdot+\delta) d\mu_n = \int h(\cdot+\delta) d\mu \ge F(X-\delta)$

$$\overset{\text{let } \delta \downarrow 0}{\implies} \liminf F_n(X) \ge F(X)$$

$\because X \in C$

3) $\implies$ 2) Note: F is non decreasing $\implies R/C$ is countable $\implies C \subseteq \mathbb{R}$ is dense. Let $\epsilon > 0$, by 3) there exists $a, b \in C$ s.t.

$$\mu([a,b]) > 1-\epsilon$$

and

$$\mu_n([a,b]) > 1-\epsilon \quad \forall n$$

Let $g \in C_b(\mathbb{R})$, since g is uniform continuous on $[a,b]$, $\exists \delta > 0$ s.t.

$$|X - Y| < \delta \implies |g(X) - g(Y)| < \epsilon$$

$\forall X, Y \in [a,b]$. Fix $a = X_0 < X_1 < ... < X_m = b$, $|X_i - X_{i-1}| < \delta$, $X_i \in C$, let

$$\bar{g} = \sum_{i=1}^{m} g(X_{i-1}) 1_{(X_{i-1}, X_i]}$$

then

$$|g(X) - \bar{g}(X)| \le \begin{cases} \epsilon & \text{if } X \in [a,b] \\ \|g\|_\infty & \text{else} \end{cases}$$

As

$$\int \bar{g} d\mu_n = \sum_{i=1}^{n} g(X_{i-1})(F_n(X_i) - F(X_{i-1})),$$

and $X_i \in C$, we have

$$\int \bar{g} d\mu_n \to \int \bar{g} d\mu$$

therefore,

$$\begin{aligned} \left|\int g d\mu_n - \int g d\mu\right| &\leq \left|\int g d\mu_n - \bar{g} d\mu_n\right| + \left|\int \bar{g} d\mu_n - \int \bar{g} d\mu\right| + \left|\int \bar{g} d\mu - \int g d\mu\right| \\ &\leq \epsilon(3 + 2\|g\|_\infty) \text{ for } n \text{ large} \end{aligned}$$

□

Remark 178. $\mu_n \implies \mu$ does not imply

$$\mu_n(A) \to \mu(A)$$

for every $A \in \mathcal{B}(\mathbb{R})$ not even for $A = (-\infty, n]$. For example

$$\mu_n = \mathcal{N}(0, \frac{1}{n})$$

$\mu = \delta_0$, then

$$\mu_n \implies \mu \text{ but } 0 = \mu_n(\{0\}) \not\longrightarrow \mu(\{0\}) = 1$$

but it is true that

$$\mu_n(A) \to \mu(A)$$

for all $A \in \mathcal{B}(\mathbb{R})$ s.t. $\mu(\partial A) = 0$, where $\partial A = \bar{A}/\mathring{A}$.

Also recognize that

$$X_n \xrightarrow{\mathcal{L}} X$$

doesn't imply

$$X_n \xrightarrow{P} X$$

or even

$$X_n \xrightarrow{\text{a.s.}} X$$

However this changes if only the distributions are given and the r.v.'s can be chose.

Theorem 179. *(Skovohod) Let F_n, $n \geq 0$ be cdf's s.t.*

$$F_n(X) \to F_0(X)$$

$\forall X \in C = \{X : F_0$ continuous at $X\}$, there exists a probability space $(\Omega, \mathcal{F}, P)$ and r.v.'s $X_n : \Omega \to \mathbb{R}$ s.t.

$$F_n(\cdot) = P(X_n \geq \cdot)$$

and

$$X_n \to X_0 \text{ a.s.}$$

Proof. Let $(\Omega, \mathcal{F}) = ([0,1], \mathcal{B}[0,1])$ and P =Lebesgue measure. Let

$$\begin{aligned} X_n^+(w) &= \inf\{X : F_n(X) > w\} \\ X_n^-(w) &= \inf\{X : F_n(X) \geq w\} \end{aligned}$$

then $X_n^+ \geq X_n^-$ and $F_n(\cdot) = P(X_n^{\pm} \leq \cdot)$, then

$$X_n^+ = X_n^- \; P - \text{a.s.}$$

Let $w \in \Omega$, $X \in (X_0^+(w), \infty) \cap C$, then

$$X > X_0^+(w) \implies F_0(X) > w \implies F_n(X) > w$$

$\forall\, n$ large, then

$$X \geq X_n^+(w) \;\forall\, n \text{ large}$$

then

$$\limsup X_n^+(w) \leq X_0$$

let

$$X \downarrow X_0^+(w) \implies \limsup X_n^+(w) \leq X_0^+(w)$$

Similarly

$$\liminf X_n^-(w) \geq X_0^-(w)$$

Since $X_n^+ = X_n^-$ a.s., we have

$$X_n^+ \to X_0^+ \text{ a.s.}$$

□

7.2 Law of Large Numbers

Lecture 19 (4/8/14)

Let $(X_n)_{n \geq 1}$ be integrable rv's on $(\Omega, \mathcal{F}, P)$. Assume that

$$E[X_n] = \mu \in \mathbb{R} \;\forall\, n$$

and let $S_n = X_1 + ... + X_n$.

Definition 180. The weak law of large numbers (week LLN) holds for (X_n) if

$$\frac{S_n}{n} \xrightarrow{P} \mu$$

as $n \to \infty$.

Definition 181. The strong law of large numbers (strong LLN) holds for (X_n) if

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} \mu$$

as $n \to \infty$.

The basic idea is sample average $\to$ expectation. This needs integrability and independent assumption.

Example 182. Let (X_n) be iid with density

$$P(X) = \frac{1}{\pi}\frac{1}{1+X^n}$$

Cauchy distribution. Then one can check by convolution that $\frac{S_n}{n}$ has the same distribution for all n. Hence

$$E[|X_n|] = \infty$$

Theorem 183. *(A weak LLN) Let $(X_n)_{n\geq 1}$ be integrable r.v. with $E[X_n] = \mu \; \forall n$*

$$var(X_n) = E[(X_n - E[X_n])^2] = \sigma^2 < \infty$$

$$cov(X_i, X_j) = E[(X_i - E(X_i))(X_j - E(X_j))] \leq R(|i-j|)$$

where $R : \mathbb{N} \to \mathbb{R}$ with $R(n) \to 0$ as $n \to \infty$. Then

$$\frac{S_n}{n} \xrightarrow{P} \mu$$

for $S_n = X_1 + ... + X_n$.

Proof. WLOG $\mu = 0$ By Chebyshev's inequality

$$P(\left|\frac{S_n}{n}\right| > \epsilon) \leq \frac{E(\frac{S_n}{n})^2}{\epsilon^2} = \frac{E(S_n^2)}{n^2\epsilon^2}$$

hence

$$\begin{aligned}
E[S_n^2] &= E[(X_1 + ...X_n)^2] = \sum_{1\le i,j\le n} E[X_iX_j] \\
&= \sum_{i=1}^{n} E[X_i^2] + 2\sum_{1\le i\le j\le n} E[X_iX_j] \\
&\le n\sigma^2 + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} R(j-i) \\
&= n\sigma^2 + \sum_{k=1}^{n-1}(n-k)R(k) \\
&\le n\sigma^2 + 2n\sum_{k=1}^{n}|R(k)|
\end{aligned}$$

Note

$$|R(k)| \to 0 \implies \frac{1}{n}\sum_{k=1}^{n-1}|R(k)| \to 0$$

hence

$$P(\left|\frac{S_n}{n}\right| > \epsilon) \le \frac{n\sigma^2}{n^2\epsilon^2} + \frac{2n}{n\epsilon^2}\frac{1}{n}\sum_{k=1}^{n-1}|R(k)| \to 0 + 0 = 0$$

□

Lemma 184. *$X_k \in \mathbb{R}_+$, $X_k \to 0$, then*

$$\frac{1}{n}\sum_{k=1}^{n-1} X_k \to 0$$

Proof.

$$\frac{X_1 + ... + X_n}{n} = \frac{X_1 + ... + X_N}{n} + \frac{X_{N+1} + ... + X_n}{n}$$

□

Theorem 185. *(Strong LLN) Let $(X_k)_{k\ge1}$be pairwise independent and identically distributed rv's with $E[|X_k|] < \infty$. Let $\mu = E[X_1]$ and $S_1 = X_1 + ... + X_n$, then*

$$\frac{S_n}{n} \xrightarrow{a.s.} \mu$$

We use Etemadi's proof.

Lemma 186. *It is sufficient to prove that*

$$\frac{T_n}{n} \to \mu \ a.s.$$

where $T_n = Y_1 + ... + Y_n$, $Y_k = X_k 1_{|X_k| \le k}$.

Proof.

$$\begin{aligned}
\sum_{k\ge 1} P(\underbrace{|X_k| > k}_{=\{X_k \ne Y_k\}}) &\le \int_0^\infty P(|X_1| > t) = E[|X_1|] < \infty \\
&\implies P(X_k \ne Y_k \text{ i.o.}) = 0 \\
&\implies \sup_n |S_n - T_n| < \infty \text{ a.s.} \\
&\implies \begin{cases} \limsup \frac{S_n}{n} = \limsup \frac{T_n}{n} \\ \liminf \frac{S_n}{n} = \liminf \frac{T_n}{n} \end{cases}
\end{aligned}$$

□

Lemma 187.

$$\sum_{k\ge 1} \frac{E[Y_k^2]}{k^2} \le 4E[|X_1|] < \infty$$

Proof.

$$E[Y_k^2] = \int_0^\infty 2yP(|Y_k| > y)dy \le \int_0^k 2yP(|X_1| > y)dy$$

implies

$$\begin{aligned}
\sum_{k\ge 1} \frac{E[Y_k^2]}{k^2} &\le \sum_{k\ge 1} \frac{1}{k^2} \int_0^\infty 1_{\{y\le k\}} 2yP(|X_1| > y)dy \\
&= \int_0^\infty (\sum_{k\ge 1} k^{-2} 1_{y<k}) 2yP(|X_1| > y)dy \qquad (7.1)
\end{aligned}$$

Claim: $2y \sum_{k>y} k^{-2} \le 4 \ \forall\, y \ge 0$.

Proof of claim: Let $m \ge 2$ then

$$\sum_{k\ge m} k^{-2} \le \int_{m-1}^\infty x^{-2} dx = \frac{1}{m-1}$$

If $Y \geq 1$, sum starts with $k = \lfloor Y \rfloor + 1 \geq 2$,

$$\implies 2Y \sum_{k>Y} k^{-2} \leq \frac{2Y}{\lfloor Y \rfloor} \leq 4$$

If $0 < Y < 1$,

$$2Y \sum_{k>Y} k^{-2} \leq 2(1 + \sum_{k=2}^{\infty} k^{-2}) \leq 4$$

proves the claim. Together with (7.1)

$$\sum \frac{E[Y_k^2]}{k^2} \leq 4 \int_0^{\infty} P(|X_1| > Y) dY = 4E[|X_1|] < \infty$$

□

Now we prove strong LLN

Proof. By treating separately X_k^+, X_k^- we may assume that $X_k \geq 0$, we first prove the result along a sub sequence $k_n = \lfloor \alpha^n \rfloor$, where $\alpha > 1$ is a fixed constraint.

$$\begin{aligned} \sum_{n\geq 1} P(|T_{k_n} - E[T_{k_n}]| > \epsilon k_n) &\leq \epsilon^{-2} \sum_{n\geq 1} \frac{\text{var}(T_{k_n})}{k_n^2} \\ &= \epsilon^2 \sum_{n\geq 1} \frac{1}{k_n^2} \sum_{m=1}^{k_n} \text{var}(Y_m) \because \text{pairwise indep} \\ &= \epsilon^2 \sum_{m\geq 1} \text{var}(Y_m) \sum_{n\in k_n \geq m}^{k_n} \frac{1}{k_n^2} \because \text{Fubini} \end{aligned}$$

also

$$\lfloor \alpha^n \rfloor \geq \frac{\alpha^n}{2} \implies \sum_{n\in k_n\geq m} \lfloor \alpha^n \rfloor^{-2} \leq 4 \sum_{n\in k_n \geq m} \alpha^{-2n} \leq 4(1-\alpha^{-2})^{-1} m^{-2}$$

so

$$\sum_{n\geq 1} P(|T_{k_n} - E[T_{k_n}]| > \epsilon k_n) \leq 4(1-\alpha^{-2})^{-1} \epsilon \sum_{m\geq 1} \frac{E[Y_m^2]}{m^2} < \infty$$

then

$$P(\frac{|T_{k_n} - E[T_{k_n}]|}{k_n} > \epsilon \text{ i.o.}) = 0$$

i.e.

$$\frac{|T_{k_n} - E[T_{k_n}]|}{k_n} \to 0 \text{ a.s.}$$

Since $E[Y_k] \to E[X_1] = \mu$ (DCT), this shows

$$\frac{T_{k_n}}{k_n} \to \mu \text{ a.s.}$$

To handle the indices between the k_n's , sandwich if $k_n \leq m \leq k_{n+1}$, then

$$\frac{T_{k_n}}{k_{n+1}} \leq \frac{T_m}{m} \leq \frac{T_{k_{n+1}}}{k_n}$$

Since $k_n = \lfloor \alpha^n \rfloor$, we have

$$\frac{k_{n+1}}{k_n} \to \alpha$$

and

$$\frac{1}{\alpha}\mu \leq \liminf_{m\to\infty} \frac{T_m}{m} \leq \limsup_{m\to\infty} \frac{T_m}{m} \leq \alpha\mu$$

Now let

$$\alpha \downarrow 1 \implies \lim \frac{T_m}{m} = \mu$$

Now the theorem of strong LLN follows from lemma 186. □

Lecture 20 (4/10/14)

LLN tells us

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} \mu = E[X_1]$$

but how far does S_n deviate from $n\mu$ (fluctuation) ?

Theorem 188. *(Law of the Iterated Log) Let (X_n) iid, $X_n \in \mathcal{L}^2$, let $\mu = E[X_1]$ $\sigma^2 = var(X_1)$, then*

$$\begin{aligned}\limsup \frac{S_n - n\mu}{\sqrt{2\sigma^2 n ln(lnn)}} &= +1 \\ \liminf \frac{S_n - n\mu}{\sqrt{2\sigma^2 n ln(lnn)}} &= -1\end{aligned}$$

Proof. Exercise. □

7.3 Central Limit Theorem

Recall that $X \sim \mathcal{N}(\mu, \sigma^2)$ $(\mu \in \mathbb{R}, \sigma > 0)$ If

$$P(X \leq x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$$

and then $\mu = E[X]$, $\sigma^2 = \text{var}(X)$. Let Φ be the cdf of $\mathcal{N}(0,1)$ i.e.

$$\Phi = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy$$

Note that Φ is continuous.

Theorem 189. *(Central Limit Theorem) Let $(X_i)_{i\geq 1}$ be iid with $E[X_i] = \mu$ and var$(X_i) = \sigma^2 \in (0, \infty)$. For $S_n = X_1 + ... + X_n$,*

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

or equivalent

$$P(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x) \to \Phi(x) \, \forall \, x \in \mathbb{R}$$

Theorem 190. *(Lindeberg) Let $X_{n,i}$, $1 \leq i \leq n$, $n \geq 1$ be r.v. and*

1) $X_{n,1}$,..., $X_{n,n}$ are independent $\forall \, n$

2) $E[X_{n,i}] = 0$, $E[X_{n,i}^2] = \sigma_{n,i}^2 < \infty$, $\sum_{i=1}^{n} \sigma_{n,i}^2 = 1 \; \forall \, n$

3) $\lim_{n\to\infty} \sum_{i=1}^{n} E[X_{n,i}^2 1_{\{|X_{n,i}|>\epsilon\}}] = 0$, $\forall \, \epsilon > 0$

Then $S_n = X_{n,1} + ... + X_{n,n} \implies \mathcal{N}(0,1)$

Note: the iid assumption from the CLT is relaxed here.

Remark 191. Each of the contribution $X_{n,i}$ is small:

$$\begin{aligned} \max_{1\leq i\leq n} \sigma_{n,i}^2 &\leq \max(\epsilon^2 + E[X_{n,i}^2 1_{\{|X_{n,i}|>\epsilon\}}]) \\ &\leq \epsilon^2 + \underbrace{\sum_{i=1}^{n} E[X_{n,i}^2 1_{\{|X_{n,i}|>\epsilon\}}]}_{\to 0 \text{ by } 3)} \end{aligned}$$

We first prove Lindeberg $\implies$ CLT

Proof. Let $X_{n,i} = \frac{X_i - \mu}{\sigma\sqrt{n}}$, then $X_{n,1}...X_{n,n}$ are independent. $E(X_{n,i}) = 0$,

$$\sigma_{n,i}^2 = E[X_{n,i}^2] = \frac{1}{n} \to \sum_{i=1}^{n} \sigma_{n,i}^2 = 1$$

Fix $\epsilon > 0$,

$$\sum_{i=1}^{n} E[X_{n,i}^2 1_{\{|X_{n,i}|>\epsilon\}}] = \frac{1}{\sigma^2} E[(X_1 - \mu)^2 1_{|X_1-\mu|>\epsilon\sigma\sqrt{n}}] \overset{\text{DCT}}{\longrightarrow} 0$$

□

Now prove Lindeberg

Proof. Let f be a bounded smooth function with f', f'', f''' bounded. Choose (possible after enlarging Ω)

$$Y_{n,i} \sim \mathcal{N}(0, \sigma_{n,i}^2)$$

s.t.

$$Y_{n,1}...Y_{n,n} \qquad X_{n,1}...X_{n,n}$$

are independent $\forall\, n$. Hence it suffices to show that

$$E[f(\sum_{i=1}^{n} X_{n,i})] - E[f(\sum_{i=1}^{n} Y_{n,i})] \to 0$$

So the idea is to replace $X_{n,i}$ by $Y_{n,i}$, one after the other. Let

$$Z_{n,i} = X_{n,1} + ... + X_{n,j-1} + Y_{n,i+1} + ... + Y_{n,n}$$

applying Taylor

$$\begin{aligned} f(Z_{n,i} + X_{n,i}) - f(Z_{n,i} + Y_{n,i}) &= f'(Z_{n,i})(X_{n,i} - Y_{n,i}) \\ &\quad + \frac{1}{2} f''(Z_{n,i})(X_{n,i}^2 - Y_{n,i}^2) + R(Z_{n,i}, X_{n,i}) \\ &\quad + R(Z_{n,i}, Y_{n,i}) \end{aligned}$$

implies

$$\begin{aligned} |R(Z,X)| &\le \left| \frac{1}{6} X^3 f'''(Z + \theta X) \right| \le \text{const}\, |X|^2 \\ &\le \left| \frac{1}{2} X^2 (f''(Z + \theta X) - f''(Z)) \right| \le \text{const}\, |X|^2 \\ &\le \text{const}(\epsilon X^2 + X^2 1_{|X|>\epsilon}) \end{aligned}$$

Since $Z_{n,i}$ is independent of $X_{n,i}$ and $Y_{n,i}$, we get

$$\begin{aligned}|E[f(Z_{n,i}+X_{n,i})]-E[f(Z_{n,i}+Y_{n,i})]| &\leq \left|E[f'(Z_{n,i})]\cdot E[_{n,i}-Y_{n,i}]\right| \\ &\quad +\frac{1}{2}\left|E[f''(Z_{n,i})]\cdot\left(E[X_{n,i}^2]-E[Y_{n,i}^2]\right)\right| \\ &\quad +E[|R(Z_{n,i},X_{n,i})|]+E[|R(Z_{n,i},Y_{n,i})|] \\ &\leq \text{const}(\epsilon\sigma_{n,i}^2+E[X_{n,i}^2 1_{|X_{n,i}|>\epsilon}]+E[|Y_{n,i}|^3])\end{aligned}$$

Thus $\forall\,\epsilon>0$

$$\begin{aligned}\left|E[f(\sum_{i=1}^n X_{n,i})-f(\sum_{i=1}^n Y_{n,i})]\right| &= |E[f(Z_{n,n}+X_{n,n})-f(Z_{n,1}+Y_{n,1})]| \\ &\leq \left|\sum_{i=1}^n E[f(Z_{n,i}+X_{n,i})-f(Z_{n,i}+Y_{n,i})]\right| \\ &\leq \text{const}(\epsilon+\sum_i E[X_{n,i}^2 1_{|X_{n,i}|>\epsilon}]+\sum_i E[|Y_{n,i}|^3]) \\ &\leq \text{const}(\epsilon+\epsilon+\epsilon) \text{ for large } n\end{aligned}$$

Using assumption 3) in the theorem

$$\sum_{i=1}^n E[|Y_{n,i}|^3]=\sum_{i=1}^n \sigma_{n,i}^3\sqrt{\frac{8}{\pi}}\leq \underbrace{(\max_i \sigma_{n,i})}_{\to 0 \text{ by rank}}\underbrace{(\sum_i \sigma_{n,i}^2)}_{1}\sqrt{\frac{8}{\pi}}\to 0$$

□

Lecture 21 (4/14/14)

In the CLT, how fast does

$$P(\frac{S_n-n\mu}{\sqrt{n}\sigma}\leq X)$$

converge to $\Phi(X)$? The typical rate is $\sqrt{n}$.

Theorem 192. *(Barry-Esseen) Let X_i be iid, $\mu=E[X_i]$, $\sigma^2=var(X_i)\in(0,\infty)$*

$$S_n=X_1+...+X_n$$

If $E[|X_i|^3]<\infty$, then

$$\sup_{X\in\mathbb{R}}\left|P(\frac{S_n-n\mu}{\sqrt{n}\sigma}\leq X)-\Phi(X)\right|\leq\frac{E[|X_1|^3]}{\sigma^3\sqrt{n}}$$

Proof. Skipped. □

Example 193. (rounding error) Let $X_1, ..., X_n, ...$ iid uniformly distributed in $[-\frac{1}{2}, \frac{1}{2}]$, (that of X_n as rounding error) Then since $\mu = 0$, $\sigma^2 = \frac{1}{12}$

$$\begin{aligned} P(a \le S_n \le b) &= P(a\sqrt{\frac{12}{n}} \le \frac{S_n - n\mu}{\sqrt{n}\delta} \le b\sqrt{\frac{12}{n}}) \\ &\overset{n \text{ lareg}}{\approx} \Phi(b\sqrt{\frac{12}{n}}) - \Phi(a\sqrt{\frac{12}{n}}) \end{aligned}$$

E.g. $P(|S_{100}| \le 5) \approx \Phi(\sqrt{3}) - \Phi(-\sqrt{3}) \approx 0.917$. In probability of $|\text{error}| \le 5$ is about 0.9, comparing to worst error is $n/2 = 50$.

Example 194. (Asymptotic of the median) Let F be a cdf. The median of F is $m = F^{-1}(\frac{1}{2})$ right inverse. Let $X_1, X_2...$ be iid $\sim F$ and median $m = 0$. Assume that $F'(0)$ exists and $F'(0) > 0$.

Let Z_n be the empirical median

$$Z_n = X_{(k)}$$

for $k = \lfloor \frac{n}{2} + 1 \rfloor$, where $X_{(k)}(w)$ is the $k-$th smallest value of $X_1(w), ..., X_n(w)$, then

$$\sqrt{n} Z_n \implies \mathcal{N}(0, \frac{1}{4F'(0)^2})$$

Proof. We shall use Lindeberg's theorem. Fix $X \in \mathbb{R}$. Let $Y_{n,i} = 1_{\{X_i > \frac{X}{\sqrt{n}}\}}$. Then

$$\sqrt{n} Z_n \le X_n \iff X_{(k)} \le \frac{X}{\sqrt{n}} \iff \sum_{i=1}^{n} Y_{n,i} \le n - k$$

with $P_n = 1 - F(\frac{X}{\sqrt{n}})$, we have

$$E[Y_{n,i}] = P_n, \ \text{var}(Y_{n,i}) = P_n(1 - P_n)$$

Standardize to use Lindeberg

$$X_{n,i} = \frac{Y_{n,i} - P_n}{\sqrt{nP_n(1 - P_n)}}$$

Since

$$\sqrt{n} Z_n \le X \iff S_n = \sum_{i=1}^{n} X_{n,i} \le a_n = \frac{n - k - nP_n}{\sqrt{nP_n(1 - P_n)}}$$

As $n \to \infty$, and substituting $Y = 1/\sqrt{n}$

$$\frac{n-k-nP_n}{\sqrt{nP_n(1-P_n)}} \sim \frac{n-\frac{n}{2}-nP_n}{\sqrt{nP_n(1-P_n)}} = \frac{-\frac{n}{2}-nF(\frac{X}{\sqrt{n}})}{\sqrt{n(1-F(\frac{X}{\sqrt{n}}))F(\frac{X}{\sqrt{n}})}}$$
$$= \frac{F(XY)-\frac{1}{2}}{Y\sqrt{1-F(XY)}\sqrt{F(XY)}} \overset{L'Hopital}{\to} \frac{XF'(0)}{\sqrt{1-F(0)}\sqrt{F(0)}} = 2XF'(0)$$

Hence for $\delta > 0$, n large

$$P(S_n \leq 2F'(0)X - \delta) \leq P(S_n \leq a_n) = P(\sqrt{n}Z_n \leq X) \leq P(S_n \leq 2F'(0)X + \delta)$$

apply Lindeberg to both sides,

$$\Phi(2F'(0)X - \delta) \leq \Phi(2F'(0) + \delta)$$

Let $\delta \to 0$

$$P(\sqrt{n}Z_n \leq X) \to \Phi(2F'(0)X)$$

□

8 Markov Chains

8.1 Notations

Lecture 22 (4/16/14)

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let S be a countable set, the state space, let $(X_n)_{n\geq 0}$ be $S-$valued rv's.

Definition 195. (X_n) has the Markov property if

$$P(X_{n+1} = x_{n+1}|X_0 = x_0, ..., X_n = x_n) = P(X_{n+1} = x_{n+1}|X_n = x_n)$$

for all $X_0, ..., X_{n+1} \in S$, $n \in \mathbb{N}$ s.t.

$$P(X_{n+1} = x_{n+1}|X_n = x_n) > 0$$

Definition 196. (X_n) is a time-homogeneous Markov chain if in addition

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

$\forall\, n \geq 0$, $X, Y \in S$, or set

$$K(X, A) = P(X_{n+1} \in A|X_n = x)$$

The matrix

$$\mathrm{P} = (P_{XY})_{X,Y\in S}$$

$$\mathrm{P}_{XY} = P(X_1 = Y|X_0 = x)$$

is the transition probability matrix of (X_n).

Recall P is a stochastic matrix, i.e. $P_{XY} \in [0,1]$ and

$$\sum_{Y\in S} P_{XY} = 1 \qquad \forall\, X \in S$$

Example 197. The simple random walk on $\mathbb{Z}$ is a Markov chain on $S = \mathbb{Z}$ here

$$p_{n,n+1} = p, \qquad p_{n,n-1} = 1 - p$$

$$p_{n.m} = 0 \text{ if } |n - m| \neq 1$$

Definition 198. The distribution μ of X_0 is called initial distribution of the chain.

Note that μ and P characterize the distribution of the chain. Indeed

$$P(X_0 = x_0, ..., X_n = x_n) = \mu(\{X_0\})P_{X_0,X_1}...P_{X_{n-1},X_n}$$

In the sequel, (X_n) is a time-homogeneous Markov chain on S, with given μ and P.

Theorem 199. *(Markov property) Let $X \in S$ and $N \in \mathbb{N}$ conditional on $\{X_N = x\}$, the sequence $X_N,\ X_{N+1},...$ is a Markov chain with transition probability P and initial distribution δ_k. Moreover, conditionally on $\{X_N = x\}$, $X_N,\ X_{N+1},...$ is independent of $X_0, ..., X_{N-1}$.*

Proof. Exercise. (This is a special case of a result proved later) □

Theorem 200. *(Chapman-Kolomogorov) The $n-$step transition probability $P^n_{XY} = P(X_n = y|X_0 = x)$ satisfy the Chapman-Kolmogoro equation*

$$P^{(m+n)}_{XZ} = \sum_{Y\in S} p^{(m)}_{XY} p^{(n)}_{YZ}\ \forall\, X, Y, Z \in S,\ m, n \geq 1$$

In matrix notation

$$P^n = (p^{(n)}_{XY})_{X,Y\in S}$$

is given by

$$P^{(n)} = P^n = P = ... = P$$

Moreover if μ *is identified with the row vector*

$$\mu = (\mu(\{X\}))_{X \in S}$$

then

$$P(X_n = Y) = (\mu P^n)_Y$$

$Y-$*th entry of row vector* μP^n.

Proof. We have

$$\begin{aligned} P_{XZ}^{(m+n)} &= \sum_{Y \in S} \underbrace{P(X_{m+n} = z | X_0 = x, X_m = y)}_{P(X_{m+n}=z|X_m=y)} P(X_m = y | X_0 = x) \\ &= \sum_{Y \in S} P_{YZ}^{(n)} P_{XY}^{(m)} \end{aligned}$$

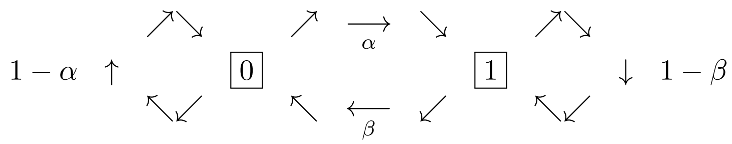
and

$$P(X_n = y) = \sum_{X \in S} P(X_n = y | X_0 = x) P(X_0 = x) = (\mu P^{(n)})_Y$$

□

Example 201. Let $S = \{0, 1\}$ and $0 < \alpha < \beta < 1$

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$



Hence

$$\begin{aligned} P^n &= \frac{1}{\alpha+\beta} \left[\begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + (1-(\alpha+\beta))^n \begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix} \right] \\ &\overset{n\to\infty}{\longrightarrow} \frac{1}{\alpha+\beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} \end{aligned}$$

hence

$$\lim_{n\to\infty} P(X_n = 0) = \frac{\beta}{\alpha+\beta}, \ \lim_{n\to\infty} P(X_n = 1) = \frac{\alpha}{\alpha+\beta}$$

Notation 202.

$$\begin{aligned} P_X(B) &= P(B|X_0 = x) \forall B \in \mathcal{F} \\ E_X(Z) &= E^{P_X}[Z] \text{ expectation under } P_X \\ &= E[Z|X = x] = \text{value of } E[Z|X_0] \text{ on } [X_0 = x] \end{aligned}$$

Definition 203. Let $A \subseteq S$ then

$$H^A = \inf\{n \in \mathbb{N}_0 : X_n \in A\}$$

($\inf \varnothing = +\infty$) is the hitting time of A.

$$\begin{aligned} h^A(X) &= P(H^A < \infty | X_0 = x) = P_X(H^A < \infty) \\ K^A(X) &= E_X(H^A) \end{aligned}$$

Lecture 23 (4/22/14)

Theorem 204. *a)* $h^A = (h^A(X))_{X \in S}$ *is the minimal non-negative solution of*

$$\begin{cases} h^A(X) = 1 & X \in A \\ h^A(X) = \sum_{Y \in S} P_{XY} h^{(A)}(Y) & X \notin A \end{cases}$$

the second case is called "mean value property"

b) $k^A = (k^A(X))_{X \in S}$ *is the minimal non-negative solution of*

$$\begin{cases} k^A(X) = 0 & X \in A \\ k^A(X) = 1 + \sum_{Y \notin A} P_{XY} k^{(A)}(Y) & X \notin A \end{cases}$$

Proof. a) h^A is a solution for $X \in A$ is clear. For $X \notin A$

$$h^A(X) = P_X(H^A < \infty) = \sum_{Y \in S} \underbrace{P_X(H^A < \infty | X_1 = Y)}_{h^{(A)}(Y)} \underbrace{P_X(X_1 = Y)}_{P_{XY}}$$

if g is any non negative solution, we can show that $g \geq h^A$. Clearly $X \in A$, $g(X) = 1 = h^A(X)$. For $X \notin A$

$$\begin{aligned} g(X) &= \sum_{Y \in S} P_{XY} g(Y) = \sum_{Y \in S} P_{XY} 1 + \sum_{Y \notin A} P_{XY} g(Y) \\ &= \sum_{Y \in A} P_{XY} + \sum_{Y \notin A} P_{XY} (\sum_{Z \in A} P_{YZ} + \sum_{Z \notin A} P_{YZ} g(Z)) \\ &= P_X(X_1 \in A) + P_X(X_1 \notin A, X_2 \in A) + \sum_{Y \notin A} \sum_{Z \notin A} P_{XY} P_{YZ} g(Z) \\ &\dots \\ &= P_X(X_1 \in A) + \dots + P_X(X_1 \notin A, \dots, X_{n-1} \notin A, X_n \in A) + \sum_{Y_1 \notin A} \dots \sum_{Y_n \notin A} P_{XY_1} \dots P_{Y_{n-1} Y_n} g(Y \\ &\geq P_X(H^A \leq n) \; \forall n \geq 1 \end{aligned}$$

b) Exercise. □

8.2 Strong Markov Property

We know that let (X_n) be a Markov chain with μ, P. Define $F_n = \sigma(X_0, ..., X_n)$ $n \geq 0$, then $(F_n)_{n\geq 0}$ is a filtration, i.e. $F_0 \subseteq F_1 \subseteq ....$ Note since S is discrete, $A \in F_n \iff A$ is a countable union of sets $\{X_0 = x_0, ..., X_n = x_n\}$.

Definition 205. A r.v. $\tau : \Omega \to \{0, 1, 2, ..., +\infty\}$ is a stopping time if $\{\tau = n\} \in F_n$, $n \geq 0$, which is of course equivalent to say $\{\tau \geq n\} \in F_n$, $n \geq 0$.

Example 206. 1) $\tau \equiv t \in \mathbb{N}_0 \cup (+\infty)$. 2) $H^A : \{H^A = n\} = \{X_0 \notin A, ..., X_{n-1} \notin A, X_n \in A\} \in F_n$.

3) $\tau =$last visit to $A : \{t = n\} = \{X_n \in A, X_{n+1} \notin A, X_{n+2} \notin A, ...\}$ in general $\notin F_n$

Notation 207. X_τ is the r.v.

$$X_\tau(w) := (X_{\tau(w)})(w) \quad \nearrow \longrightarrow \searrow \quad X_\tau = X_0(\tau, \mathrm{id})$$

Theorem 208. *(Strong Markov Property) Let τ be a stopping time conditionally on $\{\tau < \infty\} \cap \{X_\tau = x\}$, $(\hat{X}_n)_{n\geq 0}$, $\hat{X}_n := X'_{\tau+n}$ is a Markov chain with matrix P and initial distribution $\mu = \delta_X$ for any X s.t.*

$$P(\{\tau < \infty\} \cap X_\tau = x) \neq 0$$

Moreover, $(\hat{X}_n)_{n\geq 0}$ and $(X_{\tau\cap n})_{n\geq 0}$ are conditionally independent given $\{\tau < \infty\} \cap \{X_\tau = x\}$.

Proof. 1) We first consider $\tau \equiv N \in \mathbb{N}_0$ (deterministic time), we show

$$\begin{aligned} P(\{X_N &= x_N, X_{N+1} = x_{N+1}, ..., X_{N+n} = x_{N+n}\} \cap A | X_N = x) \\ &= 1_{X=X_N} P_{X_N X_{N+1}} ... P_{X_{N+n-1} X_{N+n}} P(A | X_N = x) \quad \forall A \in F_N \end{aligned} \tag{8.1}$$

It suffices to consider $A = \{X_0 = x_0, ..., X_N = x_N\}$.

(8.1) is clear if $X \neq X_N$. Whereas if $X = X_N$, it becomes

$$P(X_0 = x_0, ..., X_N = x_N, ..., X_{N+n} = x_{N+n}) = \underbrace{P(X_0 = x_0, ..., X_N = x_N)}_{\mu(X_0) P_{X_0 X_1} ... P_{X_{N-1} X_N}} P_{X_N X_{N+n}} ... P_{X_{N+n-1} X_{N+n}}$$

2) Now let τ be a stopping time, let $A \in \sigma(X_{n\cap\tau}, n \geq 0)$ and fix $m \in \mathbb{N}_0$. Then $A \cap \{\tau = m\} \in \mathcal{F}_m$. We have

$$\begin{aligned} P[\{\hat{X}_1 &= x_1, ..., \hat{X}_n = x_n\} \cap A \cap \{\tau = m\} \cap \{X_\tau = x\}] \\ &= P[\{\hat{X}_1 = x_1, ..., \hat{X}_n = x_n\} \cap A \cap \{\tau = m\} \cap \{X_m = x\}] \\ &= P_X(X_1 = x_1, ..., X_n = x_n) P[A \cap \{\tau = m\} \cap \{X_m = x\}] \text{ by 1)} \\ &= P_X(X_1 = x_1, ..., X_n = x_n) P[A \cap \{\tau = m\} \cap \{X_\tau = x\}] \end{aligned}$$

Adding over $m \geq 0$, since

$$\bigcup_{m \geq 0} \{\tau = m\} = \{\tau < \infty\}$$

we get

$$\begin{aligned} P[\{\hat{X}_1 &= x_1, ..., \hat{X}_n = x_n\} \cap A \cap \{\tau = m\} \cap \{X_\tau = x\}] \\ &= P_X(X_1 = x_1, ..., X_n = x_n) P[A \cap \underbrace{\{\tau = m\} \cap \{X_\tau = x\}}_{B}] \end{aligned}$$

Divide by $P(B) \implies$

$$P[\{\hat{X}_1 = x_1, ..., \hat{X}_n = x_n\} \cap A | B] = P_X(X_1 = x_1, ..., X_n = x_n) P[A|B] \quad (8.2)$$

For $A = \Omega$, we get the first claim:

$$P[\{\hat{X}_1 = x_1, ..., \hat{X}_n = x_n\} | B] = P_X(X_1 = x_1, ..., X_n = x_n) \quad (8.3)$$

The second claim follows by combining (8.2) and (8.3)

$$P[\{\hat{X}_1 = x_1, ..., \hat{X}_n = x_n\} \cap A | B] = P[\{\hat{X}_1 = x_1, ..., \hat{X}_n = x_n\} | B] P[A|B]$$

□

8.3 Recurrence and Transience

Lecture 24 (4/24/14)

Definition 209. Let $i, j \in S$

1) i leads to j, denoted $(i \to j)$ if

$$P_i(H^{\{j\}} < \infty) = P_i(X_n = j \text{ for some } n \geq 0) > 0$$

2) i commutes with j if $i \to j$ and $j \to i$, denoted $(i \leftrightarrow j)$.

3) (X_n) is irreducible if $i \leftrightarrow j$ for all $i, j \in S$.

4) The equivalent classes of $\leftrightarrow$ are called communicating classes.

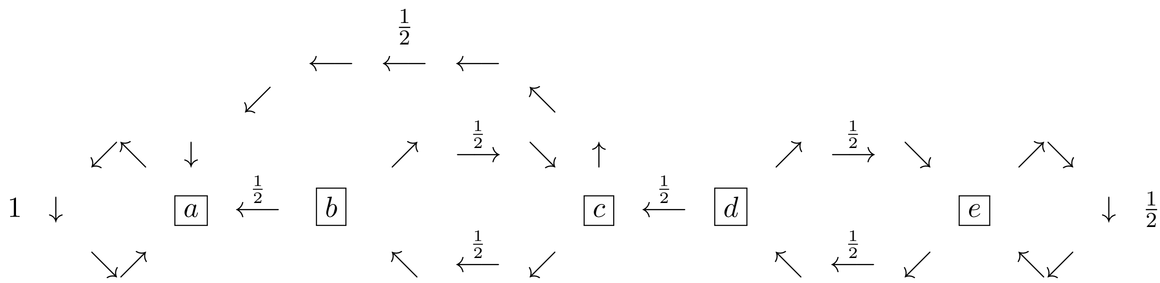
5) A communicating class C is closed if

$$i \in C, i \to j \implies j \in C$$

i.e. chain cannot leave C.

6) $i \in C$ is absorbing if $\{i\}$ is a closed class.

Example 210.



a is absorbing. Classes are $\{a\}, \{b, c\}$,$\{d, e\}$. $\{a\}$ is closed, $\{b, c\}, \{d, c\}$ are not.

Definition 211. Let $i \in S$

$$\tau_i = \min\{n \geq 1, X_n = i\} = \text{"first passage time"}$$

More passage times:

$$\begin{aligned} \tau_i^{(0)} &= 0 \\ \tau_i^{(1)} &= \tau_i \\ \tau_i^{(r+1)} &= \min\{n \geq \tau_i^{(r)} + 1, X_n = i\} \text{ "first return after } \tau_i^{(r)}\text{"} \end{aligned}$$

Length of r-th excursion from i:

$$\sigma_i^{(r)} = (\tau_i^{(r)} - \tau_i^{(r-1)}) 1_{\{\tau_i^{(r-1)} < \infty\}}$$

Lemma 212. *Let $r \geq 2$. Conditional on $\{\tau_i^{(r-1)} < \infty\}$, $\sigma_i^{(r)}$ is independent of $\{X_{k \cap \tau_i^{(r-1)}}, k \geq 0\}$ and*

$$P\{\sigma_i^{(r)} = n | \tau_i^{(r-1)} < \infty\} = P_i(\underset{\substack{\uparrow \\ \tau_i}}{\sigma_i^{(1)}} = n), n \geq 0$$

$$P\{\sigma_i^{(r)} < \infty | \tau_i^{(r-1)} < \infty\} = P_i(\tau_i < \infty)$$

Proof. Exercise by applying strong Markov property. □

Definition 213.

$$V_i = \sum_{n\geq 0} 1_{\{X_n=i\}} = \#\text{ visit to state } i$$

Remark 214.

$$E_i(V_I) = \sum_{n\geq 0} P_i(X_n = i) = \sum_{n\geq 0} P_{ii}^{(n)}$$

Definition 215. i is recurrent if $P_i(V_i = \infty) = 1$, i.e. chain will return infinitely after, a.s.

i is transient if $P_i(V_i < \infty) = 1$, i.e. chain will return only finitely many times, a.s.

Theorem 216. *(Recurrence/transience Dichotomy) Let $i \in S$ either*

1) $P_i(\tau_i < \infty) = 1$, in which case i is recurrent and $\sum_{n\geq 0} P_{ii}^{(n)} = \infty$, or

2) $P_i(\tau_i < \infty) < 1$, in which case i is transient and $\sum_{n\geq 0} P_{ii}^{(n)} < \infty$.

Lemma 217. *Let $f_i = P_i(\tau_i < \infty)$, $r < \mathbb{N}$, then $P_i(V_i \geq r) = f_i^r$.*

Proof. We have $\{V_i \geq r\} = \{\tau_i^{(r)} < \infty\}$ P_i- a.s. $\implies r = 1$ is clear.

For $r \geq 2$

$$\begin{aligned}
P_i(V_i \geq r+1) &= P_i(\tau_i^{(r)} < \infty, \sigma_i^{(r+1)} < \infty) \\
&= P_i(\sigma_i^{(r+1)} < \infty | \tau_i^{(r)} < \infty) P_i(\tau_i^{(r)} < \infty) \\
&= P_i(\tau_i < \infty) P_i(\tau_i^{(r)} < \infty) \text{ by lemma 212} \\
&= f_i \cdot f_i^r \text{ by induction} \\
&= f_i^{r+1}
\end{aligned}$$

□

We now prove the theorem.

Proof. If $f_i = P_i(\tau_i < \infty) = 1$, lemma 217 yields

$$P_i(V_i = \infty) = \lim_{r\to\infty} P_i(V_i \geq r) = \lim_{r\to\infty} 1^r = 1$$

hence

$$\infty = E_i[V_i] = \sum_{n\geq 0} P_{ii}^{(n)}$$

If $f_i < 1$,

$$P_i(V_i < \infty) = 1 - P_i(V_i = \infty) = 1 - \lim_{r\to\infty} 1^r = 1$$

and

$$\sum_{n\geq 0} P_{ii}^{(n)} = E_i[V_i] = \sum_{r\geq 0} P_i(V_i \geq r) = \sum_r f_i^r = \frac{f_i}{1-f_i} < \infty$$

□

Definition 218. 1) $C \subseteq S$ is recurrent / transient if all $i \in C$ are recurrent / transient.

2) (X_n) is recurrent / transient if all $i \in S$ are recurrent / transient.

Example 219. Let (X_n) be simple symmetric random walk on $\mathbb{Z}^d$, then (X_n) is recurrent if $d \leq 2$ and transient if $d \geq 3$.

Proof. Sketch of proof.

dimension = 1:

$$P_{ii}^{(2n)} = \frac{2n!}{n!n!}\frac{1}{2^{2n}} \sim \frac{1}{\sqrt{\pi n}} \text{ by stirling}$$

and $P_{ii}^{(2n)} = 0$ imply

$$\sum_{n\geq 0} P_{ii}^{(n)} < \infty \implies (X_n) \text{ is recurrent}$$

dimension = 2: Trick: let X, Y be independent 1-dim random walks

$$Z = \frac{X+Y}{\sqrt{2}}$$

is 2-dim random walk, so

$$P_{ii}^{(2n)}(\text{dim} = 2) = (P_{ii}^{(2n)}(\text{1-dim}))^2 \sim (\frac{1}{\sqrt{\pi n}})^2 = \frac{1}{\pi n}$$

thus

$$\sum P_{ii}^{(2n)} = \infty \implies \text{recurent.}$$

dimension $= 3$:

$$P_{ii}^{(2n)} = \sum_{j,k,l\geq 0 j+k+l=n} \frac{(2n)!}{(j!k!j!)^2}\frac{1}{6^{2n}} = \frac{(2n)!}{n!n!}\frac{1}{2^{2n}}\sum(\frac{n!}{j!k!l!})^2\frac{1}{3^{2n}}$$

Use

$$\sum\frac{n!}{j!k!l!} = 3^n$$

and

$$\frac{n!}{j!k!l!} \leq \frac{n!}{(m!)^3} \text{ for } m = \left\lceil\frac{n}{3}\right\rceil$$

to get

$$P_{ii}^{(2n)} \leq \underbrace{\frac{(2n)!}{n!n!}\frac{1}{2^{2n}}}_{\leq \text{const}\frac{1}{\sqrt{n}}}\underbrace{\frac{n!}{(m!)^3}\frac{1}{3^n}}_{\leq\text{const}\frac{1}{n}} \leq \text{const}\frac{1}{n^{3/2}} \implies \text{summable} \implies \text{transient}$$

dimension ≥ 4: if dim $A \geq 4$ were recurrent, then $d = 3$ would also be recurrent. $\tau_i = \inf\{n \geq 1, X_n = i\}$. □

Lecture 25 (4/27/14) -Last Lec-

Theorem 220. *(Recurrent is contagious) Let $i, j \in S$. Suppose that i is recurrent and $i \to j$, then*

1)

$$P_j(X_n = i \text{ i.o.}) = 1$$

in particular

$$P_j(\tau_i < \infty) = 1 \text{ and } j \to i$$

2) j is also recurrent.

Proof. 1) Since $i \to j$, $P_{ij}^{(m)} > 0$ for some $m \geq 1$. As $P_i(X_n = i \text{ i.o.}) = 1$

$$\begin{aligned}
P_{ij}^{(m)} &= P_i(X_m = j) = P_i(X_m = j, X_n = i \text{ i.o.}) \\
&= P_i(X_m = j, X_{m+k} = i \text{ i.o.}) \\
&= P_i(X_m = j)P_i(X_{m+k} = i \text{ i.o.}|X_m = j) \\
&= P_{ij}^{(m)} \cdot P_j(X_k = i \text{ i.o.}) \text{ by Markov prop}
\end{aligned}$$

thus

$$P_j(X_k = i \text{ i.o.}) = 1$$

2) Suppose for contract that j is transient, then

$$\sum_{c\geq 0} P_{jj}^{(c)} < \infty$$

we have $i \leftrightarrow j$, hence

$$P_{ij}^{(m)} > 0,\ P_{ji}^{(n)} > 0$$

Note $P_{jj}^{(m+n+l)} \geq P_{ij}^{(m)} P_{ii}^{(l)} P_{ji}^{(n)}$ $\forall\, l \geq 0$, thus

$$\sum_{l\geq 0} P_{ii}^{(l)} \leq \frac{1}{P_{ij}^{(m)} P_{ji}^{(n)}} \underbrace{\sum_{l\geq 0} P_{jj}^{(m+n+l)}}_{<\infty \text{ by assump.}} < \infty$$

thus contradict to i is recurrent. □

Corollary 221. *Let $i \in S$ be recurrent, then*

$$C_j = \{j \in S, i \to j\} = \{j \in S, i \leftrightarrow j\} = \textit{equiv class of } j$$

and C_j is recurrent.

Corollary 222. *If (X_n) is irreducible thus:*

$$\textit{One state is recurrent} \iff (X_n) \textit{ is recurrent.}$$

9 Facts about Characteristic Functions

Let $X = (X_1, ..., X_n)$ be $n-$dim random vector, i.e.

$$X_k : (\Omega, \mathcal{F}, P) \to \mathbb{R}$$

is a r.v. for $k = 1, ..., n$.

Let μ be the distribution of X on $\mathbb{R}^n$:

$$\mu(B) := P(X \in B), B \in \mathcal{B}(\mathbb{R}^n)$$

Definition 223. The characteristic function of X (or of μ) (or Fourier transform) is

$$\begin{aligned} \phi : \mathbb{R}^n &\to \mathbb{C} \\ \phi(u) &= \int_{\mathbb{R}^n} e^{i\langle u,x\rangle} \mu(dx) = E[e^{i\langle u,x\rangle}] \\ &= \int \cos(\langle u, x\rangle)\mu(dx) + i \int \sin(\langle u, x\rangle)\mu(dx) \end{aligned}$$

Notation 224. ϕ_X, ϕ_μ, $\hat{\mu}$

Note if μ has a density f,

$$\phi(u) = \int \underbrace{e^{i\langle u,x\rangle} f(x)}_{\langle e^{i\langle u,\cdot\rangle}, f\rangle_{L^2}} dx$$

basis of a space $V = \sum_{i=1}^{d} V_i e_i = (V_1, ..., V_d)$, $V_i = \langle V, e_i\rangle$.

Because of short of time, we will only state the theorems.

Theorem 225. *1)*

$$\phi : \mathbb{R}^n \to \mathbb{C}$$

is continuous.

2)

$$|\phi(u)| = 1 \;\forall\, u$$

3)

$$\phi(0) = 1$$

Theorem 226. *Suppose that* $E[|X|^m] < \infty$ *for some* $m \in \mathbb{N}$*, then*

$$\frac{\partial^m}{\partial X_{j_1} ... \partial X_{j_m}} \phi(u) = i^m E[X_{j_1} ... X_{j_m} e^{i\langle u,X\rangle}]$$

In particular if X *is one-dim* $(n = 1)$

$$\frac{\partial^m}{\partial X^m} \phi(0) = i^m E[X^m]$$

which is the formula for moments.

Remark 227. $a \in \mathbb{R}^m$,

$$A \in \mathbb{R}^{m\times n} \implies \phi_{a+AX}(u) = e^{i\langle u,a\rangle} \phi_X(A^T u)$$

Theorem 228. ϕ_μ *characterizes* μ *i.e.*

$$\phi_\mu = \phi_{\tilde{\mu}} \implies \mu = \tilde{\mu}$$

Theorem 229. $X_1, ..., X_n$ *are independent iff*

$$\phi_X(u) = \phi_{X_1}(u_1) ... \phi_{X_n}(u_n)$$

$\forall\, u = (u_1, ..., u_n) \in \mathbb{R}^n$

Theorem 230. *Let μ_n, μ are probabilities on $\mathbb{R}$ and let ϕ_n, ϕ be the corresponding characteristic functions*

1) if $\mu_n \implies \mu$ then

$$\phi_n(t) \to \phi(t) \forall\, t \in \mathbb{R}$$

2) if $\tilde{\phi}(t) := \lim_{n\to\infty} \phi_n(t)$ exists $\forall\, t$, and $\tilde{\phi}$ is continuous at $t = 0$, then $\tilde{\phi}$ is the characteristic function of a distribution $\tilde{\mu}$ and

$$\mu_n \implies \tilde{\mu}$$

Example 231. 1)

$$\mu = \mathcal{N}(b, \sigma^2) \implies \phi(u) = \exp(ibu - u^2 \frac{\sigma^2}{2})$$

$b = 0$, $\sigma^2 = 1$, so

$$\phi(u) = e^{-u^2/2}$$

2) $\mu = \text{uniform}(-a, a) \implies$

$$\phi(u) = \frac{\sin(au)}{au}$$