106349 - Advanced probability

Nick Crawford

January 24, 2019

Abstract

1 Introduction. Summary of course through an example. Branching process

We have an individual that gives a birth to a random number of offsprings – random variable X. X define a distribution, i.e., $P : \mathbb{Z}^+ \to [0, 1]$, i.e., $P(X = k) \in [0, 1]$, and $\sum_{k=0}^{\infty} P(X = k) = 1$.

Definition 1.1. $f_X(\theta) = \sum_{k=0}^{\infty} \theta^k P(X = k)$ – moment-generating function.

The series is absolutely convergent for $\theta \in [-1, 1]$ since k sums to 1. For $\theta \in (-1, 1)$, f_x is analytic, thus we can differentiate it term-by-term:

$$f'_X(\theta) = \sum_{k\geq 1} \theta^{k-1} P(X = k)$$

Since, f_X is analytic, knowing it means knowing $P(X = k)$ and vice versa.
Note that $f_X(0) = P(X = 0)$ and $f_X(1) = 1$. Also

$$f'_X(1) = \sum_{k\geq 0}^{\infty} kP(X = k) = \mathbb{E}X = \mu$$

$$\lim_{\theta\to 1} \frac{f_X(1) - f_X(\theta)}{1 - \theta} = \lim_{\theta\to 1} \frac{1 - f_X(\theta)}{1 - \theta}$$

Note also that f_X is convex, since second derivative is positive.

Size of n^{th} generation Let $\left(X_r^{(n)}\right)_{n,r=1^\infty}$, where n is generation and r is offspring number (index) in n^{th} generation.
Assume $X_r^{(n)}$ are i.i.d. (independent, identically distributed) random variables.
Identically distributed means

$$P(X_n^r = k) = P(X = k)$$

.
Independence means

$$P\big(\forall i < J\ X_{r_i}^{n_i} = k\big) = \prod_{i=1}^{J} P\big(X_{r_i}^{n_i} = k\big)$$

.
Define $z_1 = X_1^1$. $z_2 = \sum_{r=1}^{z_1} X_r^2$ an so on:

$$z_{n+1} = \sum_{r=1}^{z_n} X_r^n$$

We want to study asymptotics of z_n.
Given U and V taking values in $\mathbb{Z}^+$,

$$\mathbb{E}][U|V = k] = \sum_{j=0}^{\infty} jP(U = j|V = k)$$

, where

$$P(U = j|V = k) = \frac{P(U = j, V = k)}{P(V = k)}$$

If U, V are independent, $P(U=j|V=k)=P(U=j)$ and thus $\mathbb{E}[U|V=k]=\mathbb{E}U$.

Definition 1.2. Define random variable $\mathbb{E}[U|V]$ such that

$$\mathbb{E}[U|V]=\mathbb{E}[U|V=k]$$

if $V=k$.

Definition 1.3 (Tower property)**.**

$$\mathbb{E}\big[\mathbb{E}[U|V]\big]=\mathbb{E}U$$

Define

$$f_n=\sum_{k=0}^{\infty}\sum_{k=0}^{\infty}\theta^k P(z_n=k)=\mathbb{E}\theta^{z_n}$$

.

Theorem 1.1.

$$f_{n+1}(\theta)=f_n(f_X(\theta))$$

or

$$f_n(\theta)=\underbrace{f\circ f\circ\cdots\circ f}_{n \text{ times}}(\theta))$$

Proof. Use tower property with $U^{z_{n+1}}$ and $V=\theta^{z_n}$. By tower property

$$\mathbb{E}[\theta^{z_{n+1}}]=\mathbb{E}\big[\mathbb{E}[\theta^{z_{n+1}}|\theta^{z_n}]\big]$$

$$\mathbb{E}\big[\mathbb{E}[\theta^{z_{n+1}}|\theta^{z_n}]\big]=\sum_{k=0}^{\infty}P(z_n=k)\mathbb{E}[\theta^{z_{n+1}}|\theta^{z_n}=k]$$

What is $\mathbb{E}[\theta^{z_{n+1}}|\theta^{z_n}=k]$?

$$\mathbb{E}[\theta^{z_{n+1}}|\theta^{z_n}=k]=\mathbb{E}\Big[\theta^{\sum_{j=1}^{k}X_j^{n+1}}|\theta^{z_n}=k\Big]\overset{\text{independence}}{=}\mathbb{E}\Big[\theta^{\sum_{j=1}^{z_n}X_j^{n+1}}\Big]\overset{\text{independence}}{=}\prod_{j=1}^{k}\mathbb{E}\Big[\theta^{X_j^{n+1}}\Big]\overset{\text{i.d.}}{=}(f_X(\theta))^k$$

Thus

$$\mathbb{E}\big[\mathbb{E}[\theta^{z_{n+1}}|\theta^{z_n}]\big]=\sum_{k=0}^{\infty}P(z_n=k)(f_X(\theta))^k=f_n(f(\theta))$$

Also we can say

$$\mathbb{E}[\theta^{z_{n+1}}|z_n]=(f_X(\theta))^{z_n}$$

□

Study of z_n What is $\pi_n=P(z_n=0)=f_n(0)=f(\pi_{n-1})$, probability that population is extinguished. Since $z_{n-1}=0\Rightarrow z_n=0$, i.e. π_n is non-decreasing.
Let $P(z_n=0 \text{ for some n})=\pi$.
We hope that $\{z_n=0\}$ such that

$$\bigcup_n\{z_n=0\}=\{z_n=0 \text{ for some n}\}$$

i.e., $\pi=\lim_{n\to\infty}\pi_n$. We call π the extinction probability.

Theorem 1.2. If $\mu=\mathbb{E}>1$ then π is a unique root of $\pi=f(\pi)$ and $\pi\in[0,1)$. If $\mu\le 1$, $\pi=1$.
If we look at $f(\pi)$ and π, they intersect in 1, and they can intersect in two points since $f(x)$ is convex. There is second intersection iff $f'(1)=\mu>1$.

Construction of X_n^r Construct set Ω, $f_{n,r}:\Omega\to\mathbb{Z}^+$ and $\mathcal{F}$ a collection of subsets of Ω with $P:\mathcal{F}\to[0,1]$.
Let $\Omega=\mathbb{Z}^+\times\mathbb{Z}^+$, $\mathcal{F}=\{0,1\}^{\Omega}$.
The problem is when we have infinitely number of variables.

Example Example of not well-behaved triple $(\Omega, \mathcal{F}, P)$. $\Omega = \mathbb{N}$. Now $\mathcal{F} = \{C \subset \mathbb{N} : C \text{ has density}\}$.
C has density means

$$\frac{|C \cap \mathbb{N}|}{n} \overset{n\to\infty}{\to} \rho(C)$$

However, for $C(m) = \{1, 2, \dots, m\}$, $\forall m \quad \rho(c_m)$, and

$$\rho\Big(\bigcup C_m\Big) = 1$$

Thus $(\mathbb{N}, \mathcal{F}, \rho)$ is not a good probability space, since it doesn't fulfills this $\pi_n \to \pi$ property. Note we can define other probabilities on naturals, for example

$$P(\{i\}) = 2^{-i}$$

Asymptotics of z_i Assuming $\pi \in (0, 1)$, what is behavior of z_n?

Definition 1.4. z_n is a Markov chain if

$$P(z_{n+1} = j | z_i = k_i \quad \forall i \leq n) = P(z_{n+1} = j | z_n = k_n)$$

We can use to compute expectation:

$$\mathbb{E}[z_{n+1} | z_i = k_i \quad \forall i < n] = E[z_{n+1} | z_n = k_n]$$

Then, since $E\Big[\sum_{i=1}^{J} X_i^n\Big] = J\mu$

$$E[z_{n+1} | z_n] = \mu z_n$$

Let $M_n = \frac{z_n}{\mu^n}$ then $\mathbb{E}[M_n] = 1$. Also

$$\mathbb{E}[M_{n+1} | z_0, \dots, z_n] = M_n$$

This is a definition of martingale with respect to $z_0, \dots, z_n$.
Let $(\Omega, \mathcal{F}, P)$ we say S happens almost surely (a.s.) if

$$P(\{w \in \Omega : S \text{ is true for w}\}) = 1$$

Theorem 1.3 (Martingale convergence theorem)**.** If M_n is a positive martingale then $\lim_{n\to\infty} M_N = M_\infty$ exists a.s. and

- $\mu \leq 1$. $M_\infty = 0$ a.s. That means $\mathbb{E}M_\infty = 0$ but $\mathbb{E}M_{=}1$, i.e.,

$$\mathbb{E}\Big[\liminf_{n\to\infty} M_n\Big] < [\liminf_{n\to\infty} \mathbb{E}[M_n]$$

- $\mu > 1$. If $M_\infty > 0$ with positive probability then $z_n \sim \mu^n M_\infty$.

Lemma 1.1 (Fatou's lemma)**.**

$$\mathbb{E}\Big[\liminf_{n\to\infty} M_n\Big] \leq \liminf_{n\to\infty} \mathbb{E}[M_n]$$

Theorem 1.4.

$$\mathbb{E}[M_\infty] = 1 \iff \mu > 1 \quad \text{ and } \mathbb{E}[X\log(X)] < \infty$$

2 Overview of measure theory

Notation

- S is a set.
- $\mathcal{A}$ is algebra of subsets of S
    1. $S \in \mathcal{A}$
    2. $$E \in \mathcal{A} \Rightarrow E^C \in \mathcal{A}$$
       , where $E^C = S \setminus E$
    3. $$E_1, E_2 \in \mathcal{A} \Rightarrow E_1 \cup E_2 \in \mathcal{A}$$
       meaning
       $$E_1, E_2 \in \mathcal{A} \Rightarrow E_1 \cap E_2 \in \mathcal{A}$$

- $\mathcal{F}$ is a σ-algebra if the last item works for countable union.
- $E \Delta F = E \setminus F \cup F \setminus E$

Definition 2.1. A measurable space is a pair $\{S, \mathcal{F}\}$.

Proposition 2.1. If we have $(\mathcal{F}_i)_{i\in I}$, then $\bigcap_{i\in I} \mathcal{F}$ is also a σ-algebra.

Definition 2.2. Let C be a collection of subsets of S. $\sigma(C)$ is a smallest σ-algebra containing C (σ-algebra generated by C). It is easy to construct one

$$I = \{\mathcal{F} : \mathcal{F} \supset C\}$$

and then

$$\sigma(C) = \bigcap_{\mathcal{F}\in I} \mathcal{F}$$

Definition 2.3. Let $\{S, \mathcal{F}\}$ be a topological space. $\mathcal{B}(X)$ (Borel σ-algebra) is defined as σ-algebra generated by open sets. We denote $\mathcal{B} = \mathcal{B}(\mathbb{R})$.

Exercise

$$\pi(\mathbb{R}) = \{(-\infty, x],\ x \in \mathbb{R}\}$$

Show that $\sigma(\pi(\mathbb{R})) = B$

Definition 2.4. Additive set function on a collection of sets $\mathcal{F}$ is

$$\mu : \mathcal{F} \to [0, \infty)$$

$$\forall E, F \in \mathcal{F}\ E \cap F = \emptyset \quad \mu(E \cup F) = \mu(E) + \mu(F)$$

We say μ is σ-additive if same holds of countable infinite sets

$$\forall \{E_i\}_{i=1}^{\infty}\ E_i \cap E_j = \emptyset \quad \mu(E \cup F) = \sum_{i=1}^{\infty} \mu(E_i)$$

Definition 2.5. A triple $(S, \mathcal{F}, \mu)$ is a measure space if $\mathcal{F}$ is a σ-algebra on S and μ is σ-additive on $\mathcal{F}$.

Definition 2.6. $(S, \mathcal{F}, \mu)$ is finite if $\mu(S) < \infty$
$(S, \mathcal{F}, \mu)$ is σ-finite if

$$\exists \{E_i,\ \mu(E_i) < \infty\}_{i=1}^{\infty} \quad S = \bigcup_{i=1}^{\infty} E_i$$

Definition 2.7. If $\mu(S) = 1$, $(S, \mathcal{F}, \mu)$ is probability space.

Definition 2.8. E is null if $\mu(E) = 0$.

Definition 2.9. ϕ is said to be true almost everywhere with respect of μ if

$$\mu(\{X : \phi(X) = \text{False}\}) = 0$$

2.1 Results from measure theory

Definition 2.10. A collection of sets $\mathcal{D}$ is called a π-system if $E, F \in D \Rightarrow E \cap F \in \mathcal{D}$

Theorem 2.2 (Uniqness)**.** Let $\mathcal{D}$ be a π-system generating a σ-algebra $\mathcal{F}$. Let μ_1 and μ_2 be two finite measures on $\mathcal{F}$ which agree on $\mathcal{D}$. Then $\mu_1 = \mu_2$.

Collary 2.2.1. $(S, \mathcal{F}, P_1)$, $(S, \mathcal{F}, P_2)$ probability spaces, $P1 = P2$ on π-system $\mathcal{D}$, then $P_1 = P_2$.

Theorem 2.3 (Carathéodory's extension theorem)**.** Let $\mathcal{A}$ be an algebra of sets. $\mu_0 : \mathcal{A} \to \mathbb{R}^+$ σ-additive set function on $\mathcal{A}$. Then exists unique extension $\bar{\mu} : \sigma(\mathcal{A}) \to \mathbb{R}^+$ such that $\bar{\mu} = \mu_0$.

Homework Lebesgue on $\mathbb{R}$. $\mathcal{A} = \{\text{open set}\}$. If we have

$$O = \bigcup_{i=1}^{\infty} (a_i, b_i)$$

then

$$\mu_0(O) = \sum_{i=1}^{\infty} b_i - a_i$$

Check that μ_0 is well defined and σ-additive.

Lemma 2.1. $(S, \mathcal{F}, \mu)$ measure space. $A, B \in \mathcal{F}$, then

$$\mu(A \cup B) \leq \mu(A) + \mu(B)$$

$$\mu\left(\bigcup_{i=1}^{\infty} F_i\right) \leq \sum_{i=1}^{\infty} \mu(F_i)$$

If $\mu(S) < \infty$

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$$

From that we get inclusion-exclusion:

$$\mu\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mu(A)_i - \sum_{i \neq j} \mu(A_i \cap A_j) + \cdots + (-1)^{n-1} \mu\left(\bigcap_{i=1}^{n} A_i\right)$$

Exercise Proof the lemma

Lemma 2.2. If $F_n \subseteq F_{n+1}$ then

$$\mu\left(\bigcup_{i=1}^{\infty} F_i\right) = \lim_{n \to \infty} \mu(F_n)$$

If $\mu(S) < \infty$ and $F_n \supseteq F_{n+1}$ then

$$\mu\left(\bigcap_{i=1}^{\infty} F_i\right) = \lim_{n \to \infty} \mu(F_n)$$

Proof. Assume $\mu(S) < \infty$. Define $F_\infty = \bigcup_{i=1}^{\infty} F_i$. Let $G_n = F_n \setminus F_{n+1}$. Then

$$F_\infty = \bigcup_{i=1}^{\infty} G_i$$

Meaning

$$\mu(F_\infty) = \sum_{i=1}^{\infty} G_i$$

$$\mu(F_n) = \sum_{k=1}^{n} G_k$$

Since measure is finite, the tail of series tends to 0, thus

$$\mu(F_\infty) - \mu(F_n) = \sum_{k=n}^{\infty} G_k \to 0$$

Then we can take complements and get the second statement. □

Exercise Proof unconditionally

3 Recasting measure theory as probability

Definition 3.1. A probability space is a $(\Omega, \mathcal{F}, P)$ is a measure space such that $P(\Omega) = 1$. We call $\omega \in \Omega$ an outcome. $E \in \mathcal{F}$ is an event. $P(E)$ is probability of the event.

Example Tossing finite or infinite sequence of coins.

Tossing 4 coins

$$\Omega = \{HHHH, HHHT, HHTH, \dots, TTTH, TTTT\}$$
$$\mathcal{F} = 2^{\Omega}$$
$$P(\omega \in \Omega) = \frac{1}{|\Omega|}$$

Tossing infinite number coins

$$\Omega = \{0,1\}^{\mathbb{N}}$$

Ω has a natural topology which is called a product topology. It is coarsest topology such that $\pi_i : \Omega \to \{0,1\}$ $\pi_i(\omega) = \omega_i$ is continuous.
Let $\mathcal{F} = \mathcal{B}(\Omega)$.
Smallest σ-algebra such that

$$\pi_i^{-1}(0) \subset \Omega \in \mathcal{F}$$
$$\pi_i^{-1}(1) \subset \Omega \in \mathcal{F}$$

$$\pi_i(\Omega, \mathcal{F}) \to \left(\{0,1\}, \{0,1\}^{\{0,1\}}\right)$$

Natural π-system $\mathcal{F}_n$ smallest σ-algebra making $\pi_1, \dots, \pi_n$ measurable.
Note that

Proposition 3.1.

$$\bigcup_n \mathcal{F}_n \neq \mathcal{F}$$

Proof. Define $S_n(\omega) = \sum_{i=1}^n \omega_n$.

$$X_n = \frac{S_n(\omega)}{n}$$

Define

$$Y(\omega) = \limsup X_n(\omega)$$
$$E = \left\{\omega : Y(\omega) \geq \frac{1}{3}\right\}$$
$$E \in \mathcal{F} \setminus \bigcup_n \mathcal{F}_n$$

□

What $\mathcal{F}_n$ looks like? For example, $\mathcal{F}_2$ has 4 outcomes, deciding only first two tosses.

Note If we take $(\Omega_4, \mathcal{F}^{(4)}, P_4)$, restricting to $(\Omega_3, \mathcal{F}^{(3)}, P_3)$

$$P_4(\{(0,0,0,\omega_4)\}) = P_3(\{(0,0,0)\})$$

Thus we want P_{fair} defined on Ω to fulfill same property:

$$P_{fair}(E) = P_n(\tilde{E})$$

where $E \in \mathcal{F}_n$ and $\tilde{E} \in F^{(n)}$.

Definition 3.2. $E \subset \mathcal{F}$ occures almost surely (a.s.) if $P(E) = 1$.

Definition 3.3 (lim sup and lim inf)**.** Let $\{E_n\}$ be a sequence of events.

$$\limsup E_n = \bigcap_m \bigcup_{n \geq m} E_n = \{E_n \text{ occurs infenetely often (i.o.)}\} = \{\omega \in \Omega : \forall m \, \exists n(\omega) > m \quad \omega \in E_n(\omega)\}$$

Alternatively, $(\Omega, \mathcal{F})$ and $\{E_n\}$ there is a natural map

$$I : \Omega \to \{0,1\}^N$$

$$\omega \mapsto \{1_{E_n}(\omega)\}$$

where

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

Now

$$\liminf E_n = \bigcup_m \bigcap_{n\geq m} E_n = \{E_n \text{ occurs eventually}\} = \{\omega \in \Omega : \exists m(\omega)\, \forall n \geq m(\omega) \quad \omega \in E_n(\omega)\}$$

Remark Since everything is countable, if $E_n \in \mathcal{F}$, then $\limsup E_n, \liminf E_n \in \mathcal{F}$
We can write

$$\left\{\frac{S_n}{n} \to \frac{1}{2}\right\} = \left\{\limsup \frac{S_n}{n} \leq \frac{1}{2}\right\} \cap \left\{\liminf \frac{S_n}{n} \geq \frac{1}{2}\right\}$$

.
Choose $q \in \mathbb{Q}^+$ and take a look at

$$\left\{\liminf \frac{S_n}{n} > q\right\} = \liminf E_n(q)$$

where $E_n = \left\{\omega : \frac{S_n}{n} > q\right\}$.
In addition

$$\left\{\limsup \frac{S_n}{n} < q\right\} = \liminf F_n(q)$$

where $F_n = \left\{\omega : \frac{S_n}{n} < q\right\}$.
Therefore $\left\{\liminf \frac{S_n}{n} > q\right\} \in \mathcal{F}$.
Finally,

$$\left\{\liminf \frac{S_n}{n} \geq \alpha\right\} = \bigcap_{q<\alpha} \left\{\liminf \frac{S_n}{n} > q\right\}$$

Lemma 3.1 (Fatou's lemma)**.**

$$P\left[\liminf_{n\to\infty} E_n\right] \leq \liminf_{n\to\infty} p(E_n)$$

Proof.

$$\liminf_{n\to\infty} E_n = \bigcup_m \bigcap_{n\geq m} E_n$$

Sets $F_m = \bigcap_{n\geq m} E_n$ are increasing and $F_n \subseteq E_n$, thus

$$P\left[\liminf_{n\to\infty} E_n\right] = \lim_{n\to\infty} P(F_n) \leq \liminf_{n\to\infty} P(E_n)$$

□

Lemma 3.2 (Fatou's lemma)**.**

$$P\left[\limsup_{n\to\infty} E_n\right] \geq \limsup_{n\to\infty} p(E_n)$$

Proof. Note that $(\limsup E_n)^C = \liminf E_n^C$, thus this is straightforward form previous lemma. □

Lemma 3.3 (First Borel-Cantelli lemma)**.** Let $\{E_n\} \subseteq \mathcal{F}$ be a sequence of events s.t. $\sum_n P(E_n) < \infty$, then

$$P(E_n \text{ happens i.o.}) = 0$$

Proof.

$$P(E_n \text{ i.o.}) = P\left(\bigcap_m \bigcup_{n\geq m} E_n\right) \leq P\left(\bigcup_{n\geq m} E_n\right) \leq \sum_{n=m}^{\infty} P(E_n) \overset{m\to\infty}{\to} 0$$

Since $P(E_n \text{ i.o.})$ is independent on m, it got to be 0. □

Example Fix $\epsilon > 0$. Look at $P\left(\left|\frac{S_n(\omega)}{n} - \frac{1}{2}\right| > \epsilon\right)$.

Claim

$$P\left(\left|\frac{S_n(\omega)}{n}-\frac{1}{2}\right|>\epsilon\right)\leq\frac{12}{\epsilon^4}\frac{1}{n^2}$$

By 3.3 $P\left(\left|\frac{S_n(\omega)}{n}-\frac{1}{2}\right|>\epsilon \text{ i.o.}\right)=0$ thus

$$\left\{\frac{S_n}{n}\to\frac{1}{2}\right\}=\bigcap_{\epsilon>0}\left\{\left|\frac{S_n(\omega)}{n}-\frac{1}{2}\right|<\epsilon \text{ eventually}\right\}=1$$

Definition 3.4. Let $(S,\mathcal{F})$, $(\Omega,\mathcal{B})$ be measurable spaces.

$$\phi:S\to\Omega$$

ϕ is $\big((\mathcal{F},\mathcal{B})\big)$-measurable if $\forall B\in\mathcal{B}\quad\phi^{-1}(B)\in\mathcal{F}$.

Remark $\mathcal{C}$ is collection of sets in Ω. $\phi^{-1}(\mathcal{C})=\left\{\phi^{-1}(C):C\in\mathcal{C}\right\}$.

- $$\phi^{-1}\left(\bigcap_{i\in I}B_i\right)=\bigcap\phi^{-1}(B_i)$$
- $$\phi^{-1}\left(\bigcup_{i\in I}B_i\right)=\bigcup\phi^{-1}(B_i)$$
- $$\phi^{-1}\left(B^C\right)=\left[\phi^{-1}(B)\right]^c$$

Lemma 3.4. Let $\sigma(\mathcal{C})=\mathcal{B}$. ϕ is measurable iff $\phi^{-1}(\mathcal{C})\subseteq\mathcal{F}$.

Collary 3.1.1. $\Omega=\mathbb{R}$, $\mathcal{B}(\mathbb{R})$ then ϕ is measurable iff

$$\forall x\ \phi^{-1}((-\infty,x])\subseteq\mathcal{F}$$

Lemma 3.5. Let $(S,\mathcal{F})$, $(T,\mathcal{T})$, $(\Omega,\mathcal{B})$ be measurable spaces. Let $\phi_1:S\to T$ and $\phi_2 T\to\Omega$ measurable. Then $\phi_2\circ\phi_1$ is measurable.

Proof. Let $B\in\mathcal{B}$. Then $\phi_2^{-1}(B)\in\mathcal{T}$, and thus $\phi_1^{-1}\big(\phi_2^{-1}(B)\big)\in\mathcal{F}$, meaning $(\phi_2\circ\phi_1)^{-1}(B)\in\mathcal{F}$. □

Lemma 3.6. $\Omega=\mathbb{R}$. Then $\{\phi|\phi \text{ is } \mathcal{F},\mathcal{B}\text{-measurable}\}$ is an algebra over $\mathbb{R}$.

Proof. Using previous lemma and the fact $+$ is continuous, and thus measurable, we define $\Psi(s)=(\phi_1(s),\phi_2(s))$.
Ψ is measurable. Take a look at

$$\Psi^{-1}\big((-\infty,x_1]\times(-\infty,x_2]\big)=\{s:\phi_1(s)\in(-\infty,x_1],\ \phi_2(s)\in(-\infty,x_2]\}$$

□

Notation

$$\phi:(S,\mathcal{F})\to(\Omega,\mathcal{B})$$

We write $\phi\in\mathcal{F}$ for ϕ is $\mathcal{F},\mathcal{B}$ measurable.

Constructions preserved by measurability

Proposition 3.2. If $\{\phi_n\}_{n=1}^{\infty}$ measurable maps $(S, \mathcal{F}) \to (\Omega, \mathcal{B})$, then $\liminf \phi_n$, $\limsup \phi_n$, $\inf \phi_n$, $\sup_n$ are also measurable.

Proof. For example, fpr infimum, we need to show that

$$\left\{s | \inf_n \phi_n(s) \leq c\right\} \in \mathcal{F}$$

or alternatively,

$$\left\{s | \inf_n \phi_n(s) > c\right\} \in \mathcal{F}$$

which is just countable intersection:

$$\bigcap_n \{s : \phi_n(s) > c\}$$

Same for lim sup, which is just infimum of supremum:

$$\limsup \phi_n = \inf_m \left(\sup_{n \geq m} \phi_n\right)$$

□

Recall

$$S_n = \text{number of 1's until n}$$

We can view s_n as a composition of projection and sum:

$$\omega \mapsto \big(\pi_1(\omega), \ldots \pi_n(\omega)\big) \mapsto \sum_{i=1}^{n} \pi_i(\omega)$$

Both are continuous (projection from the definition of product topology) and thus measurable, and so is $\frac{S_n}{n}$.

4 Random variables

Definition 4.1. Let $(\Omega, \mathcal{F}, P)$ be a probability space. $X : \Omega \to (S, \mathcal{S})$ measurable is called a random variable.

Notation

$$\{\omega : X(\omega) \in A\} = X^{-1}(A)$$

We use notation like $X \in A$.

Basic constructions with random variables

Definition 4.2. Given a probability space $(\Omega, \mathcal{F}, P)$ and measurable $(S, \mathcal{S})$, X induces measure $\mathcal{L}_X$ on $(S, \mathcal{S})$ via

$$\mathcal{L}_X(E) = P(X \in E)$$

$\mathcal{L}_X$ is called marginal distribution of X or law of X.

Proposition 4.1. $\mathcal{L}_X$ is countably additive set function.

If $(S, \mathcal{S})$ is $\mathbb{R}, \mathcal{B}$. By uniqueness theorem, $\mathcal{L}_X$ if defined by

$$F_X(x) = \mathcal{L}_X\big((-\infty, x]\big) = P\big(X \in (-\infty, x]\big)$$

Proposition 4.2. $\mathcal{L}_X \mapsto F_X$ is 1-1 and onto.

Proof. Uniqueness:
If μ, ν exists such that

$$\mu\big((-\infty, x]\big) = F_X(x) = \nu\big((-\infty, x]\big)$$

then, since they agree on π-system, and thus are equal by uniqueness theorem.
Existence $\mu\big((-\infty, x]\big)$ fulfills Carathéodory's extension theorem requirements, thus there exists unique extension. □

We assume there exists Lebesgue measure on Borel sets $([0, 1], \mathcal{B}, \lambda)$.

Definition 4.3. A coupling of X,Y is $(\Omega, \mathcal{F}, P)$ and $\tilde{X}, \tilde{Y} : \Omega \to \mathbb{R}$ such that $\mu_{\tilde{X}} = \mu_X$ and $\mu_{\tilde{Y}} = \mu_Y$.

Theorem 4.3 (Skorokhod's representation (of a random variable X))**.** Given μ_X, μ_Y can we construct $(\Omega, \mathcal{F}, P)$ and $\tilde{X}, \tilde{Y} : \Omega \to \mathbb{R}$ such that $\mu_{\tilde{X}} = \mu_X$ and $\mu_{\tilde{Y}} = \mu_Y$, i.e., a coupling of X,Y.

Proof. Given increasing, right continuous F such that $F(-\infty) = 0$ and $F(\infty) = 1$. We want to define $X : [0, 1] \to \mathbb{R}$ s.t. $F_X = F$.

$$X^-(\omega) = \inf \{x : F(x) \geq \omega\}$$

(We could also choose $X^+(\omega) = \inf \{x : F(x) \geq \omega\}$, which is a bit different)
We want to show that

$$\{\omega : X^-(\omega) \leq x\} = \{\omega \leq F(x)\}$$

$X^-(\omega) \leq x$ means that $F(x) \geq \omega$ (by definition), verifying $\{\omega : X^-(\omega) \leq x\} \subseteq \{\omega \leq F(x)\}$
$F(x) \geq \omega$ means $X^-(\omega) \leq x$ (since F is increasing), finishing the proof.
Note, that X^+ and X^- disagree only on countable number of points. □

F_X is cumulative distribution function (CDF) or distribution function of X.
We ask the question: what are set properties distinguish F_X?

Proposition 4.4 (Properties of CDF)**.**

1. F_X is non-decreasing
2. F_X is right continuous
3. $F_X(-\infty) = 0$

Proof.

1. If $x < y$, $(-\infty, x] \subseteq (-\infty, y]$, thus, from monotonicity of measure $F_X(x) \leq F_X(y)$
2. we want to show
$$\lim_{x \downarrow x_0} F(x) = F(x_0)$$
Since if $E_n \downarrow E$, then $\mu(E_n) \to \mu(E)$
We can look on sequence $\{x_n\}$:
$$\omega \in \bigcap_n \{X \in (-\infty, x_n)\} \Rightarrow \forall n \quad X(\omega) \leq x_n \Rightarrow X(\omega) \leq x \Rightarrow \bigcap_n \{X \in (-\infty, x_n)\} \subseteq \{X \in (-\infty, x)\}$$
The other direction is obvious.
3. Since
$$\bigcap_x X^{-1}\big((-\infty, x]\big) = \emptyset$$

□

4.1 Independence

$(\Omega, \mathcal{F}, P)$ probability space. Let $\{\mathcal{J}_i\}_{i \in I}$ be a collection of sub-σ-algebras of $\mathcal{F}$.

Definition 4.4 (Independence of σ-algebras)**.** Say $\{\mathcal{J}_i\}_{i \in I}$ are independent if

$$\forall i_1, \ldots i_k \, \forall j \quad G_{ij} \in \mathcal{J}_i \quad P\left(\bigcap_{j=1}^{k} G_{ij}\right) = \prod_{j=1}^{k} P(G_{ij})$$

Definition 4.5 (Independence of random variables)**.** Say $\{X_i\}_{i \in I}$ are independent if $\sigma(X_i)$ are independent.

Definition 4.6 (Independence of sets)**.** Say $\{E_i\}_{i \in I}$ are independent if random variables $\mathbb{1}_{E_i}$ are independent.

Connection with elementary probability theory

Lemma 4.1 (Checking independence)**.** Let $\mathcal{F}_1$, $\mathcal{F}_2$ be σ-algebras, $\mathcal{A}_1$, $\mathcal{A}_2$ π-systems such that $\sigma(\mathcal{A}_\infty) = \mathcal{F}_i$. Then $\mathcal{F}_1$, $\mathcal{F}_2$ are independent iff

$$\forall A_1 \in \mathcal{A}_1,\ A_2 \in \mathcal{A}_2 \quad P(A_1 \cap A_2)P(A_1)P(A_2)$$

Proof. Given $E \in \mathcal{F}$ let $P_E(A) = P(A \cap E)$, a measure on $(\Omega, \mathcal{F})$.
Given $A_1 \in \mathcal{A}_1$ consider $P_{A_1}\big|_{\mathcal{F}_2}$.

$$\forall\ A_2 \in \mathcal{A}_2\ P_{A_1}(A_2) = P(A_1)P(A_2)$$

Thus P_{A_1} and $P(A_2) \times P$ are measures on $\mathcal{F}_2$ agreeing on $\mathcal{A}_2$.

$$\forall A_1 \in \mathcal{A}_1, E_2 \in \mathcal{F}_2 \quad P(A_1 \cap E_2) = P(A_1)P(E_2)$$

Next iterate argument argument for $\mathcal{F}_1$

$$\forall\ E_2 \in \mathcal{F}_2\ P_{E_2} = P(E_2)P$$

By uniqueness

$$\forall E_i \in \mathcal{F}_i \quad P(E_1 \cap E_2) = P(E_1)P(E_2)$$

□

Collary 4.4.1. To check $X_1, \dots, X_k$ are independent random variables it suffices

$$\forall x \in \mathbb{R}^k \quad P(X_i \leq x_i) = \prod_{i=1}^{k} P(X_1 \leq x_i)$$

Lemma 4.2 (Second Borel-Cantelli lemma)**.** If $\sum_i P(E_i) = \infty$ and $\{E_i\}_{i=1}^\infty$ are independent, then

$$P(E_n \text{ i.o.}) = 0$$

Proof.

$$\{E_i \text{ i.o.}\}^C = \left\{E_i^C \text{ eventually}\right\}$$

It's enough to show

$$P\left(\bigcap_{i\geq n} E_i^C\right) = 0$$

or, by truncating

$$P\left(\bigcap_{i\geq n}^{k} E_i^C\right) = 0$$

By independence

$$P\left(\bigcap_{i\geq n} E_i^C\right) \prod_{i=n}^{k} P(E_i^C) = \prod_{i=n}^{k} [1 - P(E_i)] \leq e^{-\sum_{i=n}^{k} P(E_i)}$$

Since the sum tends to infinity, the exponent tends to 0. □

Example Let $\{X_i\}_{i=1}^\infty$ be i.i.d. $Exp(1)$ random variables, i.e.

$$P(X_i > x) = e^{-x}$$

We are interested in growth rate of $X_n \leq f(n)$.
If $f(n) = \alpha \log(n)$

$$P(X_n > f(n)) = e^{-f(n)} = n^{-\alpha}$$

Thus, from Lemmas 3.3, 4.2

$$P(X_n \geq \alpha \log(n) \text{ i.o.}) = \begin{cases} 0 & \alpha > 1 \\ 1 & \alpha \leq 1 \end{cases}$$

Define $L = \limsup_{n\to\infty} \frac{X_n}{\log(n)}$, then

$$P(L \geq 1) = P(X_n \geq \log(n) \text{ i.o.}) = 1$$

Finally, if we look at $E = \bigcup_{k=1}^{\infty} \left\{L \geq 1 + \frac{1}{k}\right\}$,

$$P(E) \leq \sum_{k=1}^{\infty} P\left(L \geq 1 + \frac{1}{k}\right) \leq \sum_{k=1}^{\infty} P\left(X_n \geq \left(1 + \frac{1}{2k}\right)\log(n)\right) = 0$$

thus $P(L \leq 1) = 1$.

Method of generation of i.i.d. uniform $[a,b]$ variables We write $\omega = \sum_{i=1}^{\infty} \frac{\omega_i}{2^i}$

$$\begin{cases} w^{(1)} = \omega_1\omega_3\omega_6\omega_{10}\omega_{15}\dots \\ w^{(2)} = \omega_2\omega_5\omega_9\omega_{14}\dots \\ w^{(3)} = \omega_4\omega_8\omega_{13}\omega_{19}\dots \\ \vdots \end{cases}$$

Theorem 4.5 (Kolmogorov's zeroone law)**.**

$$\mathcal{T}_n = \sigma(X_n, X_{n+1}, \dots)$$
$$\mathcal{T} = \bigcap_n \mathcal{T}_n$$

Suppose $\{X_n\}_{n=1}^{\infty}$ are independent, then $\forall A \in \mathcal{T}$, $P(A) \in \{0,1\}$.

Proof. We show that $P(A) = P(A)^2$.
We show that $\mathcal{T}$ is independent of itself, $\forall A, B \in \mathcal{T} \quad P(A \cup B) = P(A) \cdot P(B)$.
$\mathcal{T} \subset \mathcal{T}_1$. Consider $\mathcal{I}_{l,n} = \sigma(X_l, X_{l+1}, \dots, X_n)$
For $n < k$, take $A \in \mathcal{I}_{l,n}$, $B \in \mathcal{I}_{k,m}$, then A, B are independent.
Let $\Pi_n = \bigcup_{m>n+1} \mathcal{I}_{n+1,m}$. Π_n is π-system, and $\sigma(\Pi_n) = \mathcal{T}_n$.
By lemma 4.1, $\mathcal{I}_{l,n}$ is independent on $\mathcal{T}_n$. Thus $\forall A \in \mathcal{I}_{l,n}$, $B \in \mathcal{T}_n$,

$$P(A \cap B) = P(A) \cdot P(B)$$

$\mathcal{T}_1 = \sigma(\bigcup_n \mathcal{I}_{1,n})$, and thus $\mathcal{T}_1$ is independent from $\mathcal{T}$ and since $\mathcal{T} \subseteq \mathcal{T}_1$, thus $\mathcal{T}$ is independent of itself and

$$\forall A \in \mathcal{T} \quad P(A) = P(A \cap A) = P(A)^2$$

□

Collary 4.5.1. If $\{X_n\}_{n=1}^{\infty}$ i.i.d. $s_n = \sum_{i=1}^{n} X_i$ then $\forall c \in \mathbb{R} \quad P(\limsup \frac{S_n}{n} \geq c) \in \{0,1\}$.

Proposition 4.6. Let $\{X_n\}_{n=1}^{\infty}$ i.i.d. $s_n = \sum_{i=1}^{n} X_i$ then $P(\limsup \frac{S_n}{n}$ exists$) \in \{0,1\}$.
If $P(\limsup \frac{S_n}{n}$ exists$) = 1$, then $\exists c \in [-\infty, \infty]$ such that $P(\limsup \frac{S_n}{n} = c) = 1$.

5 Integration theory

Definition 5.1 (Notation)**.** $(S, \mathcal{F}, \mu)$. Given $f : S \to \mathbb{R}$ measurable. We define

$$\mu(f) = \int_S f(S)\,\mathrm{d}\mu(S)$$

If $A \in \mathcal{F}$:

$$\mu(f_j A) = \mu(f \cdot \mathbb{1}_A) = \int_A f(S)\,\mathrm{d}\mu(S) = \int_A f\,\mathrm{d}\mu$$

Desirable properties of integral

1. Linearity
$$\int \alpha f + g \,\mathrm{d}\mu = \alpha \int f \,\mathrm{d}\mu + \int g \,\mathrm{d}\mu$$
2. Positivity:
$$f > 0 \Rightarrow \int f \,\mathrm{d}\mu > 0$$
3.
$$\int \mathbb{1}_A \,\mathrm{d}\mu = \mu(A)$$

Classes of functions we are going to consider:

1.
$$\mathcal{S} = \left\{ f(x) = \sum_{k=1}^{m} a_k \mathbb{1}_{A_k} \, a_k > 0, A_k \in \mathcal{F} \right\}$$
2.
$$\mathcal{P} = \{\text{positive measurable functions}\}$$
3.
$$\mathcal{I} = \left\{ f(x) = g(x) - h(x) |\, g, h \in \mathcal{P}, \int g \,\mathrm{d}\mu \text{ or } \int h \,\mathrm{d}\mu \text{ finite.} \right\}$$

Definition 5.2. For $\phi \in \mathcal{S}$ let $\mu_0(\phi) = \int_S \phi(S) \,\mathrm{d}\mu_0 = \sum_{k=1}^{m} a_k \mu(A_k)$.
Definition 5.3. For $f \in \mathcal{P}$ let $\mu(\phi) = \sup\limits_{\substack{\phi\in\mathcal{S}\\ \phi\le f}} \mu_0(\phi)$.
Proposition 5.1. If $f = g$ μ a.e., then $\mu(f) = \mu(g)$.
Lemma 5.1. If $\mu(f) = 0$ then $f = 0$ a.e.
Lemma 5.2.
$$\mu\Big((\min\{f, k\}) \cdot \mathbb{1}_{\frac{1}{k} \le f}\Big) \to \mu(f)$$

Proof. Given $\phi \in S^+$, $\phi \le f$, for k large enough, $\phi \le (\min\{f, k\}) \cdot \mathbb{1}_{\frac{1}{k} \le f}$. Since $\phi = \sum_{l=1}^{m} a_l \mathbb{1}_{A_l}$ $\phi \le f \cdot \mathbb{1}_{\frac{1}{k} \le f}$ and for k large enough $\phi \le \min\{f, k\}$. Thus
$$\phi \le (\min\{f, k\}) \cdot \mathbb{1}_{\frac{1}{k} \le f}$$
and
$$\mu_0(\phi) \le \mu\Big((\min\{f, k\}) \cdot \mathbb{1}_{\frac{1}{k} \le f}\Big)$$
Taking the limit
$$\mu_0(\phi) \le \lim_{k\to\infty} \mu\Big((\min\{f, k\}) \cdot \mathbb{1}_{\frac{1}{k} \le f}\Big)$$
By taking supremum over ϕ
$$\mu(f) \le \lim_{k\to\infty} \mu\Big((\min\{f, k\}) \cdot \mathbb{1}_{\frac{1}{k} \le f}\Big)$$
Other direction is trivial and thus
$$\mu\Big((\min\{f, k\}) \cdot \mathbb{1}_{\frac{1}{k} \le f}\Big) \to \mu(f)$$
□

Theorem 5.2 (Monotone convergence theorem)**.** Let $0 < f_n$ and $f_n \uparrow f$, then $\mu(f_n) \uparrow \mu(f)$

Proof. From lemma we can assume f is bonded and
$$\exists \epsilon \quad \{f > 0\} = \{f > \epsilon\}$$
If $\mu(f > \epsilon) = \infty$, then $\mu(f) = \infty$. Also
$$\left\{f_n > \frac{\epsilon}{2}\right\} \uparrow \left\{f \ge \frac{\epsilon}{2}\right\} \Rightarrow \mu\Big(f_n > \frac{\epsilon}{2}\Big) \to \mu(f \ge) \frac{\epsilon}{2})$$

Thus $\mu(f_n > \epsilon) \to \infty$
If $\mu(f > \epsilon) < \infty$, given $\delta > 0$, let

$$C_n = \{|f - f_n| > \delta, f > \epsilon\}$$

Then $C_n \downarrow \emptyset$, thus $\forall \delta > 0$ $\mu(C_n) \to 0$.
Given $S \ni \phi \le f$,

$$\phi = \phi \mathbb{1}_{f>\epsilon} = (\phi - \delta)\mathbb{1}_{f>\epsilon} + \delta \mathbb{1}_{f>\epsilon}$$

Let $\phi_n = (\phi - \delta)\mathbb{1}_{f>\epsilon}\mathbb{1}_{|f-f_n|<\delta}$. Obviously $\phi_n \le f_n$.
We claim

$$\exists C > 0 \quad |\mu_0(\phi_n) - \mu_0(\phi)| \le C \cdot (\delta + \mu(C_n))$$

since

$$\mu_0(\phi_n) = \mu_0(\phi) - \delta\mu(f > \epsilon, |f - f_n| > \delta) - \delta\mu(f > \epsilon)$$
$$|\mu_0(\phi_n) - \mu_0(\phi)| \le \delta\mu(f > \epsilon) + M\mu(C_n)$$

for $M \ge f$.
Since δ is arbitrary, $\lim_{n\to\infty} C \cdot (\delta + \mu(C_n)) = 0$, thus

$$\mu_0(\phi) \le \lim_{n\to\infty} \mu_0(f_n)$$

Optimizing over ϕ we get

$$\mu(f) \le \lim_{n\to\infty} \mu(f_n)$$

□

Collary 5.2.1. $f, g \in \mathcal{P}$ and $a \ge 0$ then

$$\mu(af + g) = a\mu(f) + \mu(g)$$

Proof. Taking staircase functions $\alpha_f^{(r)}$ and $\alpha_g^{(r)}$. Then $a\alpha_f^{(r)} + \alpha_g^{(r)} \uparrow af + g$ and

$$\mu\left(a\alpha_f^{(r)} + \alpha_g^{(r)}\right) = a\mu\left(\alpha_f^{(r)}\right) + \mu\left(\alpha_g^{(r)}\right)$$

By 5.2

$$\lim_{r\to\infty} \mu\left(a\alpha_f^{(r)} + \alpha_g^{(r)}\right) = \lim_{r\to\infty} a\mu\left(\alpha_f^{(r)}\right) + \mu\left(\alpha_g^{(r)}\right)$$
$$\mu(af + g) = a\mu(f) + \mu(g)$$

□

Collary 5.2.2.

$$\mu(\liminf f_n) \le \liminf \mu(f_n)$$

Proof.

$$\liminf f_n = \sup_k \left[\underbrace{\inf_{n\ge k} f_n}_{g_k}\right]$$
$$g_k \uparrow \liminf f_n$$

From 5.2

$$\mu(\liminf f_n) = \lim_k \mu(g_k) \le \liminf_k \mu(f_k)$$

□

Definition 5.4. Let $f = g - h$ a.s. such that $g, h \ge 0$, and at most one of $\int g \,\mathrm{d}\mu$, $\int h \,\mathrm{d}\mu$ is infinite. Then define

$$\int f \,\mathrm{d}\mu = \int g \,\mathrm{d}\mu - \int h \,\mathrm{d}\mu$$

Proposition 5.3. $\int f \,\mathrm{d}\mu$ is well defined.

Proof. If $g_1 - h_1 = f = g_2 - f_2$ a.s. it is true that

$$g_1 - g_2 + h_2 - h_1 = 0$$

$$g_1 + h_2 = g_2 + h_1$$

$$\int g_1 \, d\mu + \int h_2 \, d\mu = \int g_2 \, d\mu + \int h_1 \, d\mu$$

Since maximum one term on each side is infinite we can move the other one to the second side, getting the

$$\int g_1 \, d\mu - \int g_2 \, d\mu = \int h_1 \, d\mu - \int h_2 \, d\mu$$

as required □

Definition 5.5.

$$f^{\pm}(\omega) = \max\{\pm f(\omega), 0\}$$

Definition 5.6. We say $f \in L^1(\mu)$ if $\exists g, h$ such that $f = g - h \int g \, d\mu + \int h \, d\mu < \infty$.
For $f \in L^1(\mu)$, $\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu$.
$|f| = f^+ + f^-$ and $f \in L^1(\mu) \iff \int |f| \, d\mu < \infty$.

Lemma 5.3. $L^1(\mu)$ is a vector space.

Proof. $f, g \in L^1(\mu)$ thus, since $|f + g| \leq |f| + |g|$, $f + g \in L^1(\mu)$.

$$\int f + g \, d\mu = \int f^+ + g^+ \, d\mu - \int f^- + g^- \, d\mu = \int f \, d\mu + \int g \, d\mu$$

□

If $\|f\|_1 = \int f \, d\mu$ then $\|\|_1$ is a norm on $L^1(\mu)$
Further $L^1(\mu)$ is complete, i.e., each Cauchy sequence converges.

Lemma 5.4 (Reverse Fatou's Lemma)**.** Let $\{f_n\}$ be a sequence of functions such that $0 \leq f_n \leq g$ such that $\int g \, d\mu < \infty$. Then

$$\int \limsup f_n \, d\mu \geq \limsup \int f_n \, d\mu$$

Proof. Let $h_n = g - f_n$. By 3.2

$$\int \liminf h_n \, d\mu \leq \liminf \int h_n \, d\mu$$

Using the fact $\liminf h_n = g - \limsup f_n$ we get the result. □

Theorem 5.4 (Lebesgue's dominated convergence theorem)**.** Let f_n be a sequence such that $|f_n| \leq g$ and $g \in L^1(\mu)$ and $f_n \to f$ then

$$\int f_n \to \int f$$

and

$$\int |f_n - f| \to 0$$

Proof. We first proof that $\int f_n \to \int f$.
Sine $|f_n| < g$, $g \pm f_n \geq 0$. Applying 3.2 to $g \pm f_n$:

$$\int \liminf f_n \leq \liminf \int f_n$$

$$\int f \leq \liminf \int f_n$$

and

$$\int \liminf(-f_n) \leq \liminf \int (-f_n)$$

$$\int \limsup f_n \geq \limsup \int f_n$$

$$\int f \geq \limsup \int f_n \geq \liminf \int f_n \geq \int f$$

We have $h_n = |f_n - f|$, and $h_n \overset{a.s.}{\to} 0$ and $h_n \leq 2g$ so by first statement

$$0 = \lim_{n\to\infty} \int h_n$$

□

5.1 Integration on probability spaces and integration

Definition 5.7 (Expectation)**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X : \Omega \to \mathbb{R}$ be a random variable.

$$\mathbb{E}[X] = \int X(\omega)\,\mathrm{d}P(\omega)$$

Theorem 5.5 (Bounded convergence theorem)**.** Let $X_n \to X$ a.s. and $|X|_n \leq C$. Then $\mathbb{E}[|X|_n - X]to0$.

independent of DCT. Define $E_\epsilon = \{\omega : |X_n(\omega) - X(\omega)| < \epsilon\}$.

$$\mathbb{E}[|X_n(\omega) - X|] = \mathbb{E}[|X_n - X|\mathbb{1}_{E_\epsilon}] + \mathbb{E}\big[|X_n - X|\mathbb{1}_{E_\epsilon^C}\big]$$

Since $|X_n - X|\mathbb{1}_{E_\epsilon} \leq \epsilon\mathbb{1}_{E_\epsilon}$ and $|X_n - X|\mathbb{1}_{E_\epsilon^C} \leq 2C\mathbb{1}_{E_\epsilon^C}$:

$$\mathbb{E}[|X_n(\omega) - X|] \leq \epsilon\mathbb{1}_{E_\epsilon} + 2C\mathbb{1}_{E_\epsilon^C} \leq \epsilon + 2cP\big((E_\epsilon^n)^C\big)$$

For some $m > n$

$$(E_\epsilon^n)^C = \{\omega : |X_n - X| > \epsilon\} \subseteq \{\omega : |X_m - X| > \epsilon\}$$

$$\bigcap_n F_{n,\epsilon} = \{\omega : \limsup |X_n - X| => \epsilon\}$$

By continuity of measure

$$\lim_{n\to\infty} P(F_{n,\epsilon}) = 0$$

$$\lim_{n\to\infty} P\big((E_\epsilon^n)^C\big) = 0$$

□

Definition 5.8. Let $A \in \mathcal{F}$.

$$\int_A f\,\mathrm{d}\mu = \mu(f; A) = \mu(f\mathbb{1}_A)$$

We can look on it as constructing a new measure space: $\big(S \cap A = A, \mathcal{F}_A, \mu\big|_A\big)$
We claim that

$$\mu\Big|_A(f) = \mu(f\mathbb{1}_A)$$

Proposition 5.6 (The standard machine)**.**

1. Check for $\mathbb{1}_E$.
2. Check for simple functions (use linearity)
3. Use MCT to check positive functions.
4. Use linearity to extend to L^1.

Proposition 5.7. If $h, g : \mathbb{R} \to \mathbb{R}$ are Borel-measurable and X,Y are independent, then $h(X)$, $g(Y)$ are independent.

Proof. $h(X)$, $g(Y)$ are measurable, and σ-algebras generated by then are sub-σ-algebras of original ones, and thus they're independent. □

Proposition 5.8. Suppose $X, Y \in L^1$ are independent and in L_1, then $X \cdot Y \in L_1$ and

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

Proof. $X = X^+ - X_-$ and $Y = Y^+ - Y^-$, by linearity its enough to check for $X^\pm$, $Y^\pm$.
So we can assume $X, Y > 0$. Denote $X_N = \max\{X, N\}$, $Y_N = \max\{X, N\}$, from 5.2 if the claim holds for X_N, Y_N, then it holds for X, Y.
Since now X, Y are bounded, we can find simple functions $\alpha^{(r)}(X) \to X$, $\alpha^{(r)}(Y) \to Y$.
From 5.5 the identity would hold for bounded functions if it holds for simple functions. From linearity it's enough to show for indicators:

$$\mathbb{E}\mathbb{1}_E(X)\mathbb{1}_F(Y) = P(X \in E, Y \in F) = P(X \in E)P(Y \in F) = \mathbb{E}\mathbb{1}_E(X) \cdot \mathbb{E}\mathbb{1}_F(Y)$$

□

Proposition 5.9. Let $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$. Let μ_X be a law of X on $\mathbb{R}$:

$$\mu_X(B) = P(X \in B)$$

let $h : \mathbb{R} \to \mathbb{R}$, then $\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x)\mu_X(\mathrm{d}x)$

Proof. Use the Standard machine. □

Proof. □

Definition 5.9. Say $\nu \lhd \mu$ if $\mu(A) = 0 \Rightarrow \nu(A) = 0$. We say μ is absolutely continuous with respect to μ.

Theorem 5.10. $S, \mathcal{F}$ is nice. $\nu \lhd \mu \iff \exists f \in L^1(\mu)$ and $f \geq 0$, $\int f\,\mathrm{d}\mu = 1$ such that $\nu(A) = \int_A f\,\mathrm{d}\mu$. We call $f = \frac{\partial\nu}{\partial\mu}$ a Radon-Nikodym derivative.

$P(X = \cdot | Z =_j$ is probability measure on $\{x_1, \ldots, x_m\}$ for j fixed. $\mathbb{E}[X|z = z_j] = \sum_i x_i P(X = x_i | z = z_j$ is a conditional expectation. note that it is random variable: $\mathbb{E}[X|Z] = \sum_j \mathbb{E}[X|Z = z_j]$.
Properties

1. $\mathbb{E}[X|Z] \in \sigma(Z)$
2. $\forall A \in \sigma(Z)$

$$\mathbb{E}\Big[\mathbb{1}_A \cdot \mathbb{E}[X|Z]\Big] = \mathbb{E}[\mathbb{1}_A X]$$

It's enough to check for $A = \{z_j\}$:

$$\mathbb{E}\Big[\mathbb{1}_A \mathbb{E}[X|Z]\Big] = P(Z = z_j)\mathbb{E}\Big[\mathbb{E}[X|z_j]\Big] = \sum_i x_i P(X = x_i, Z = z_j) = \mathbb{E}[\mathbb{1}_A X]$$

Definition 5.10. We say that Y is a conditional expectation of Z given $\mathcal{J}$ if $Y \in \mathcal{J}$, $Y \in L^1(P)$ and $\forall A \in \mathcal{J}$ $\mathbb{E}[Y\mathbb{1}_A] = \mathbb{E}[X\mathbb{1}_A]$

Lemma 5.5. If conditional expectation Y exists, it is unique up to sets of measure 0. Henceforth call Y by $\mathbb{E}[X|\mathcal{J}]$

Proof. Let Y, Y' two conditional expectations. We'll show that $Y > Y'$ P a.s. then by symmetry $Y = Y'$ P a.s.
Let $A_n = \left\{Y > Y' + \frac{1}{n}\right\}$, $A_n \in \mathcal{J}$. □

$L^2(P)$ remark For $X, Y \in L^2(P)$ define $\langle X, Y\rangle = \mathbb{E}[XY]$ and $\|X\| = \sqrt{\langle X, X\rangle} = \sqrt{\mathbb{E}[X^2]}$. Parallelogram law:

$$\|X + Y\|^2 + \|X - Y\|^2 = 2\|X\|^2 + 2\|Y\|^2 2$$

Theorem 5.11. Let $K \subseteq L^2(P)$ be a closed convex subset, $X \in L^2(P)$ then $\exists! Y \in K$ such that $\inf\{\|X - Z\| : z \in K\} = \|X - Y\|$, call $Y = P_K(X)$.

Proof. Let $Y_n \in K$ such that $\|X - Y_n\| \to \Delta = \inf\{\|X - Z\| : z \in K\}$
We claim Y_n is Cauchy.

$$\|X - Y_n\|^2 + \|X - Y_m\|^2 = 2\left\|X - \frac{Y_n + Y_m}{2}\right\|^2 + \frac{1}{2}\|Y_n - Y_m\|^2$$

Since $\frac{Y_n + Y_m}{2} \in K$ by convexity, thus

$$2\left\|X - \frac{Y_n + Y_m}{2}\right\|^2 + \frac{1}{2}\|Y_n - Y_m\|^2 \geq 2\Delta^2 + \frac{1}{2}\|Y_n - Y_m\|^2$$

and

$$\|X - Y_n\|^2 + \|X - Y_m\|^2 \to 2\Delta^2$$

thus

$$\frac{1}{2}\|Y_n - Y_m\|^2 \to 0$$

i.e., $\{Y_n\}$ is Cauchy and thus the limit exists and is in K.
If there are two different sequences, then from same identity, the distance between limits goes to 0. □

Theorem 5.12. If $X \in L^1$ and $\mathcal{J} \subseteq \mathcal{F}$ then $\mathbb{E}[X|\mathcal{J}]$ exists

Definition 5.11. $P_K(X) = Y$, $P_K : L^2(P) \to V$. Then $P_K(X)$ is linear, contractive and self-adjoint.

We want to use P_K to define $\mathbb{E}[X|\mathcal{J}]$.

- For all $\mathcal{J} \subset \mathcal{F}$ containing all sets of measure 0 $\left\{Y, \mathbb{E}\left[Y^2\right]\right\}$ forms a closed subspace of $L^2(P)$ and $L^2(\Omega, \mathcal{J}, P)$ is called complete measure space.
- $L^2(\Omega, \mathcal{J}, P)$ is complete by Lemma.
- $L^2(\Omega, \mathcal{J}, P) \subseteq L^2(\Omega, \mathcal{F}, P)$

Definition 5.12. For $X \in L^2(\Omega, \mathcal{F}, P)$ and $\mathcal{J} \subseteq \mathcal{F}$ is complete then

$$\mathbb{E}[X|\mathcal{F}] = P_{L^2(\Omega,\mathcal{J},P)}(X)$$

Proposition 5.13.

$$\mathbb{E}[\mathbb{1}_A X] = \mathbb{E}[\mathbb{1}_A P \cdot (X)]$$

Proof. The statement is equivalent to

$$\langle \mathbb{1}_A, X\rangle = \left\langle \mathbb{1}_A, P_{L^2(\mathcal{J})} X\right\rangle$$

However for $Y \in L^2(\Omega, \mathcal{J}, P)$, $P \cdot Y = Y$:

$$\langle \mathbb{1}_A, X\rangle = \langle P \cdot \mathbb{1}_A, X\rangle = \langle \mathbb{1}_A, P \cdot X\rangle$$

□

Lemma 5.6. If $X \geq 0$ a.s. and $X \in L^2$ then $\mathbb{E}[X|\mathcal{J}] \geq 0$ a.s.

Proof. Let $A_n = \left\{\omega : \mathbb{E}[X|\mathcal{J}] < \frac{1}{n}\right\}$. We clame $P(A_n = 0)$.

$$0 \leq \mathbb{E}[\mathbb{1}_{A_n} X] = \mathbb{E}[\mathbb{1}_{A_n}\mathbb{E}[X|\mathcal{J}]] \leq -\frac{1}{n}\mathbb{1}_{A_n} \Rightarrow P(A_n) = 0$$

□

Note that to define conditional expectation $\mathbb{E}[X|\mathcal{J}]$ for $X \in L^1$, we can assume $X > 0$, since we can define

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

So, let $X_n = \max\{X, n\}$, then $X_n \in L^2(\mathcal{F})$ so $\mathbb{E}[X_n|\mathcal{J}]$ exists, and $X_n \uparrow X$. We want that (a.s.) $\mathbb{E}[X_n|\mathcal{J}] \uparrow_n$
Take a look at $X_n - X_m \geq 0$ for $n > m$:

$$\mathbb{E}[X_n - X_m|\mathcal{J}] \geq 0$$

thus

$$\mathbb{E}[X_n|\mathcal{J}] \geq \mathbb{E}[X_m|\mathcal{J}]$$

so the sequence is increasing.
Let us define

$$\mathbb{E}[X|\mathcal{J}] = \lim_{n\to\infty} \mathbb{E}[X_n|\mathcal{J}]$$

Now $\forall A \in \mathcal{J}$, from monotone convergence,

$$\mathbb{E}[\mathbb{1}_A X_n] \to \mathbb{E}[\mathbb{1}_A X]$$
$$\mathbb{E}[\mathbb{1}_A X_n] = \mathbb{E}[\mathbb{1}_A \mathbb{E}[X_n|\mathcal{J}]]$$

and also

$$\mathbb{E}[\mathbb{1}_A \mathbb{E}[X_n|\mathcal{J}]] \to \mathbb{E}[\mathbb{1}_A \mathbb{E}[X|\mathcal{J}]]$$

and thus

$$\mathbb{E}[\mathbb{1}_A X] = \mathbb{E}[\mathbb{1}_A \mathbb{E}[X|\mathcal{J}]]$$

i.e., it is the conditional expectation.

Proposition 5.14 (Properties of conditional expectations).

1. If $Y = \mathbb{E}[X|\mathcal{J}]$ a.s., then $\mathbb{E}[X] = \mathbb{E}[Y]$
2. If $X \in \mathcal{J}$, $\mathbb{E}[\mathbb{X}|\mathcal{J}] = X$ a.s.
3. $$\mathbb{E}[aX + Y|\mathcal{J}] = a\mathbb{E}[X|\mathcal{J}] + \mathbb{E}[Y|\mathcal{J}]$$
4. If $X \geq 0$ then $\mathbb{E}[X|\mathcal{J}] \geq 0$
5. If $X_n \uparrow X$ then $\mathbb{E}[X_n|\mathcal{J}] \uparrow \mathbb{E}[X|\mathcal{J}]$
6. If $X_n \geq 0$
$$\mathbb{E}[\liminf X_n|\mathcal{J}] \leq \liminf \mathbb{E}[X_n|\mathcal{J}]$$
7. $|X_n| \leq V(w)$ and $\mathbb{E}[V] < \infty$, $X_n \to X$ a.s., then
$$\mathbb{E}[X_n|\mathcal{J}] \to \mathbb{E}[X|\mathcal{J}]$$
8. $$\mathbb{E}[c(X)|\mathcal{J}] \geq c(\mathbb{E}[X|\mathcal{J}])$$
9. For $\mathcal{H} \subseteq \mathcal{J} \subset \mathcal{F}$
$$\mathbb{E}[\mathbb{E}[X|\mathcal{J}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$$
10. For $Z \in \mathcal{J}$:
$$\mathbb{E}[ZX|\mathcal{J}] = Z\mathbb{E}[X|\mathcal{J}]$$
11. For $\mathcal{H}$ independent on $\sigma(X, \mathcal{J})$:
$$\mathbb{E}[X|\sigma(\mathcal{J}, \mathcal{H})] = \mathbb{E}[X|\mathcal{J}]$$

Proposition 5.15 (Jensen's inequality). *Proof.* So let

$$S = \{a, b | ax + b \leq c(x)\}$$

Let $S' \subset S$ be a countable dense subset.

$$\forall a, b \in S' \quad a\mathbb{E}[X|\mathcal{J}] + b \leq \mathbb{E}[c(X)|\mathcal{J}]$$

Since for convex function

$$c(x) = \sup_{ax+b \leq c(x)} \{ax + b\}$$

Optimizing over S' we get the Jensen inequality. $\square$

Proposition 5.16 (Hölder's inequality).

$$|\mathbb{E}[X|\mathcal{J}]|^p \leq \mathbb{E}[X^p|\mathcal{J}]$$

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

$$|\mathbb{E}[XY|\mathcal{J}]| \leq (\mathbb{E}[|X|^p|\mathcal{J}])^{\frac{1}{p}} + (\mathbb{E}[|Y|^q|\mathcal{J}])^{\frac{1}{q}}$$

Proposition 5.17 (Minkowski's inequality).

$$|\mathbb{E}[(X + Y)^p|\mathcal{J}]|^{\frac{1}{p}} \leq |\mathbb{E}[X^p|\mathcal{J}]|^{\frac{1}{p}} + |\mathbb{E}[Y^p|\mathcal{J}]|^{\frac{1}{p}}$$

Example Let X, Z be random variable.

$$P((X, Z) \in A) = \int_A f_{(X,Z)}(x, z) \, \mathrm{d}x \, \mathrm{d}z$$

where $f_{(X,Z)}$ is called the joint distribution

Proposition 5.18.

$$\mathbb{E}[h(x)|\sigma(Z)] = \frac{\int f(x, z)h(x) \, \mathrm{d}x}{\int f(x, z) \, \mathrm{d}x} \mathbb{1}_{\int f(x,z)\mathrm{d}x \neq 0} = \phi(z)$$

Proof. We want to check that

$$\mathbb{E}[h(X); z \in \mathcal{B}] = \mathbb{E}[\phi(z); z \in \mathcal{B}$$

$$\mathbb{E}[\phi(z); z \in \mathcal{B}] = \mathbb{E}[\phi(z)\mathbb{1}_{z\in\mathcal{B}}] = \int \phi(z)\mathbb{1}_{z\in\mathcal{B}} f(x,z)\,\mathrm{d}x\,\mathrm{d}z$$

$$\mathbb{E}[h(X); z \in \mathcal{B}] = \int \mathrm{d}z \left[\int f(x,z)h(x)\,\mathrm{d}x\right]\mathbb{1}_{z\in\mathcal{B}} = \int \left[\frac{\int f(x,z)h(x)\,\mathrm{d}x}{\int f(x,z)\,\mathrm{d}x}\right]\mathbb{1}_{z\in\mathcal{B}} f(u,z)\,\mathrm{d}u\,\mathrm{d}z = \int \phi(z)\mathbb{1}_{z\in\mathcal{B}} f(x,z)\,\mathrm{d}x\,\mathrm{d}z$$

□

Suppose $\mathcal{H}, \mathcal{J} \subseteq \mathcal{F}$ σ-algebras. We want to regard

$$X \in \mathcal{H} \mapsto \mathbb{E}[X|\mathcal{J}]$$

as the expectation corresponding to some probability distribution.

$$P(A|\mathcal{J}) = \mathbb{E}[\mathbb{1}_A|\mathcal{J}]$$

So $A, \omega \mapsto P(A||\mathcal{J})(\omega)$ is random set function on $\mathcal{H}$. For ω fixed a.s. is this a probability measure? Generally, no.

Definition 5.13. Let $P(\cdot,\cdot) : \mathcal{F} \times \Omega \mapsto [0,1]$ We say that P is a regular conditional probability distribution for $\mathcal{J}$ if

1. $\forall F \in \mathcal{F}$, $P(F,\cdot)$ is a version of $\mathbb{E}[\mathbb{1}_F|\mathcal{J}]$.
2. a.e. $\omega \in \Omega$, $P(\cdot,\omega)$ is a probability measure on $(\Omega,\mathcal{F})$.

Proposition 5.19. Let $X_1, \ldots, X_k$ be independent random variables. Given $h \in \mathcal{B}(\mathbb{R}^k)$

$$\gamma_n(x) = \mathbb{E}[h(x, X_2, \ldots X_k)]$$

then

$$\gamma(X_1) = \mathbb{E}[h(X_1,\ldots,X_k)|\sigma(X_1)]$$

Proof. If $C = \big\{h \in \mathcal{B}(\mathbb{R}^k) \text{ s.t. identity holds } \big\}$ then we check that C is monotone class.
Note, for $A = A_1 \times \cdots \times A_k$ then

$$\mathbb{E}[\mathbb{1}_{A_1} \ldots \mathbb{1}_{A_k}|\sigma(X_1)] = \mathbb{1}_{A_1}\mathbb{E}[\mathbb{1}_{A_2} \ldots \mathbb{1}_{A_k}|\sigma(X_1)] = \mathbb{1}_{A_2} P(A_2 \cap \cdots \cap A_k) = \gamma_n(x)$$

□

Example Let $\{X_i\}_{i=1}^{\infty}$ i.i.d. random variables and define

$$S_n = \sum_{i=1}^{n} X_i$$

Let

$$\mathcal{G}_n = \sigma(S_n, S_{n+1}, \ldots) == \sigma(S_n, X_{n+1}, \ldots)$$

What is

$$\mathbb{E}[X_1|\mathcal{G}_n] = \mathbb{E}[X_1|\sigma(\sigma(S_n), \sigma(X_{n+1}, \ldots))]$$

Since both X_1 andS_n is independent on $\sigma(X_{n+1}, \ldots)$ we can rewrite as

$$\mathbb{E}[X_1|\mathcal{G}_n] = \mathbb{E}[X_1|\sigma(\sigma(S_n), \sigma(X_{n+1}, \ldots))] = \mathbb{E}[X_1|\sigma(S_n)]$$

We claim that

$$\mathbb{E}[X_1|S_n] = \frac{1}{n} S_n$$

since for $i \le n$

$$\mathbb{E}[X_1|S_n] = \mathbb{E}[X_i|S_n]$$

Since

$$\mathbb{E}[\mathbb{E}[X_i|S_n]; \mathbb{1}_{S_n\in B}] = \mathbb{E}[X_i; \mathbb{1}_{S_n\in B}] = \int_{\mathbb{R}^n} x_i \mathbb{1}_B(x_1+x_2+\ldots x_n)\prod_{i=1}^{n}\mathrm{d}\mu(x_i) = \int_{\mathbb{R}^n} x_1 \mathbb{1}_B(x_1+x_2+\ldots x_n)\prod_{i=1}^{n}\mathrm{d}\mu(x_i) = \mathbb{E}[X_1; \mathbb{1}_{S_n\in B}] = \mathbb{E}[\mathbb{E}[X$$

thus

$$S_n = \mathbb{E}[S_n|S_n] = \sum_{k=1}^{n} \mathbb{E}[X_k|S_n] = n\mathbb{E}[X_1|S_n]$$

6 Martingales

Definition 6.1 (Filtration)**.** $\{\mathcal{F}_i\}_{i=0}^{\infty}$ sequence of σ-algebras is a filtration if

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_\infty \subseteq \mathcal{F}$$

where $\mathcal{F}_\infty = \sigma(\bigcup_{i=1}^{\infty} \mathcal{F}_i)$

Example Let $\{W_i\}_{i=0}^{\infty}$ sequence of random variables (stochastic process). Let $\mathcal{F}_i = \sigma(W_0, \ldots W_i)$, then $\{\mathcal{F}_i\}_{i=0}^{\infty}$ is a filtration.

Definition 6.2 (Martingale)**.** A sequence $\{X_n\}_{n=0}^{\infty}$ (sub-/super-) martingale with respect to $\{\mathcal{F}_n\}_{n=0}^{\infty}$ if

1. $X_n \in L^1(\mathcal{F}_n, P)$
2. $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$ ($\geq$ for sub- and $\leq$ for super-).

We may assume $X_n = 0$ since if X_n is martingale, so is $Y_n = X_n - X_0$.

Example Let X_i be i.i.d. and $X_i \geq 0$ with $\mathbb{E}[X_i] = 1$.

$$M_n = \prod X_i$$

Then

$$\mathbb{E}[M_{n+1}|\mathcal{F}_n] = \mathbb{E}\left[\prod_{1}^{n+1} X_i|\mathcal{F}_n\right] = \mathbb{E}[X_{n+1}|\mathcal{F}_n]\prod X_i = \prod 1 \cdot \prod X_i = M_n$$

Example Consider $\vec{\mathbf{X}}_i$ i.i.d. vectors in $\mathbb{R}^d$ with natural filtration $\mathcal{F}_n$ and

$$\vec{\mathbf{S}}_n = \sum \vec{\mathbf{X}}_i$$

$$\mathbb{E}\left[\vec{\mathbf{S}}_{n+1}|\mathcal{F}_n\right] = \vec{\mathbf{S}}_n + \mathbb{E}\left[\vec{\mathbf{X}}_{n+1}|\mathcal{F}_n\right]$$

Thus if $\mathbb{E}\left[\vec{\mathbf{X}}_{n+1}|\mathcal{F}_n\right] = 0$, $\vec{\mathbf{S}}_n$ is martingale.

Example Let $d = 2$ and $\vec{\mathbf{X}}_i$ be equiprobable out of $\pm\epsilon\hat{\mathbf{x}}$, $\pm\epsilon\hat{\mathbf{y}}$.
Given $f \in \mathcal{C}^3(\mathbb{R}^2)$ consider $Z_n = f(S_n^\epsilon)$

Definition 6.3. f is (sub-/super-) harmonic if Z_n is a (sub-/super-) martingale

What happens for small ϵ:

$$\mathbb{E}[f(X_{n+1}|\mathcal{F}_n)] = \frac{1}{4}(f(\epsilon\hat{\mathbf{x}} + S_n) + f(-\epsilon\hat{\mathbf{x}} + S_n) + f(\epsilon\hat{\mathbf{y}} + S_n) + f(-\epsilon\hat{\mathbf{y}} + S_n)) =$$

$$= f(S_n) + \frac{1}{4}\Big(\big[f(S_n + \epsilon\hat{\mathbf{x}}) - f(S_n)\big] + \big[f(S_n - \epsilon\hat{\mathbf{x}}) - f(S_n)\big] + \big[f(\epsilon\hat{\mathbf{y}} + S_n) - f(S_n)\big] + \big[f(-\epsilon\hat{\mathbf{y}} + S_n)) - f(S_n)\Big]$$

For Taylor expansion we get

$$f(\epsilon\hat{\mathbf{x}} + \vec{\mathbf{S}}_n) = \epsilon\frac{\partial f}{\partial x} + \frac{\epsilon^2}{2}\frac{\partial^2 f}{\partial x^2} + \mathcal{O}(\epsilon^3)$$

thus

$$\mathbb{E}[f(X_{n+1}|\mathcal{F}_n)] = f(S_n) + \frac{\epsilon^2}{4}\nabla^2 f + \mathcal{O}(\epsilon^3)$$

i.e., we get a condition on the Laplacian of $f$which decides whether Z_n is martingale.

Information exposure martingale For some $\{\mathcal{F}_n\}_{n=1}^{\infty}$ and $X \in L^1(\Omega, \mathcal{F}, P)$

- $M_n = \mathbb{E}[X|\mathcal{F}_n]$ is a martingale by tower property.
- From Jensen $Z_n = \phi(M_n)$ is submartingale if ϕ is convex.

We'll show that

$$M_n \overset{L_1 \text{ a.s.}}{\to} \mathbb{E}[X|\mathcal{F}_\infty]$$

We are interested when $X = \mathbb{E}[X|\mathcal{F}_\infty]$.

Example Let X_n, $\mathcal{F}_n$ be martingale. $X_n - X_{n-1}$ is an amount we won in n^{th} game.

Definition 6.4. C_n is predictable if $C_n \in \mathcal{F}_{n-1}$.

Definition 6.5 (Discrete stochastic integral)**.** Discrete stochastic integral with respect to X is

$$(C \circ X)_n = \sum_{j=1}^{n} C_j (X_j - X_{j-1})$$

i.e., total amount won at time n won using strategy using gambling strategy C.

Proposition 6.1. If $X_n - X_{n-1}$ is (super)martingale with respect to $\mathcal{F}_n$ and C_n are bounded and positive (not necessarily uniformly) then so is $C \circ X$.

Proof.

$$\mathbb{E}[(C \circ X)_n | \mathcal{F}_{n-1}] = \sum_{j=1}^{n} \mathbb{E}[C_j (X_j - X_{j-1}) | \mathcal{F}_{n-1}] = (C \circ X)_{n-1} + C_n \mathbb{E}[(X_n - X_{n-1}) | \mathcal{F}_{n-1}]$$

□

So what is expected winning $\mathbb{E}[(C \circ X)_n]$?

$$\mathbb{E}[(C \circ X)_n] = \mathbb{E}[\mathbb{E}[(C \circ X)_n | \mathcal{F}_{n-1}]] = \mathbb{E}[(C \circ X)_{n-1}]$$

by induction $\mathbb{E}[(C \circ X)_n] = \mathbb{E}[(C \circ X)_0] = 0$.
Similarly, if X is a supermartingale and $C \geq 0$,

$$\mathbb{E}[(C \circ X)_n] \leq 0$$

Definition 6.6. $T : \Omega \to \mathbb{N}$ is called stopping time with respect to $\mathcal{F}_n$ if $\forall n \, \{\omega : T(\omega) \leq n\} \in \mathcal{F}_n$. Equivalently is $\{T = n\} \in \mathcal{F}_n$. Note that $\{T \geq n\} \in \mathcal{F}_{n-1}$, since its complementary of $\{T \leq n-1\}$.

Definition 6.7. Given a process $\{X_n\}_n$ we say X_n is adapted to $\{\mathcal{F}_n\}_n$ if $\forall n \quad X_n \in \mathcal{F}_n$.

Definition 6.8. Given an adapted process $\{X_n\}_n$ and stopping time $T$the stopped process

$$X_n^{(T)} = X_{\min\{T,n\}}$$

Lemma 6.1. If X_n is (super-/sub-)martingale, so is $X_n^{(T)}$.

Proof. Let $C_n = \mathbb{1}_{\{T \geq n\}}$ predictable, then

$$(C \circ X)_n = \sum_{k \leq n} C_k (X_k - X_{k-1}) = \sum_{1 \leq k \leq n} \mathbb{1}_{T \geq k} (X_k - X_{k-1}) = X_{\min\{T,n\}} - X_0$$

By we already know that $(C \circ X)_n$ preserve martingale property. □

Theorem 6.2. Let X_n be a supermartingale, then $\forall n \; \mathbb{E}\left[X_{\min\{T,n\}}\right] \leq \mathbb{E}[X_0]$.
Would the property survive under $n \to \infty$? No!

Example Let X_n be a SRW on $\mathbb{Z}$ starting from 0.

$$T_1 = \inf \{n : X_n = 1\}$$

By theorem, $\mathbb{E}\left[X_{\min\{T,n\}}\right] = 0$. On the other hand, since $P(T < \infty) = 1$, $\mathbb{E}[X_T] = 1$.
The problem is T doesn't have expectation.

Theorem 6.3 (Doob's optional sampling theorem)**.** Let T be a stopping time and let X_n be a supermartingale. If one of the following holds

1. T is bounded
2. X is bounded
3. $\mathbb{E} < \infty$ and $|X_n - X_{n-1}| \leq K \; \forall n \; P$ a.s.

then $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$.

Proof. 1.

$$\mathbb{E}[X_0] \geq \mathbb{E}\left[X_{\min\{T,n\}}\right]$$

Since

$$X_{\min\{T,n\}} \to X_T$$

a.s. since T is bounded, $\left|X_{\min\{T,n\}}\right| \leq \max|X_k|$ by dominated convergence theorem,

$$\mathbb{E}\left[X_{\min\{T,n\}}\right] \to \mathbb{E}[X_T]$$

2. Bounded convergence

3.

$$\left|X_{\min\{T,n\}}\right| \leq K \cdot \min\{T,n\} \leq KT$$

By DCT,

$$\mathbb{E}\left[X_{\min\{T,n\}}\right] \to \mathbb{E}[X_T]$$

□

Collary 6.3.1. If M is a martingale and $|M_n - M_{n-1}| \leq K$ and C is predictable and $|C| \leq K$ and $\mathbb{E}[T] < \infty$ then $\mathbb{E}[C \circ X] = 0$

Collary 6.3.2. If X is a positive supermartingale and $T < \infty$ a.s. then

$$\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$$

Proof. Fatou lemma □

Lemma 6.2. Let T be a stop time such that $\exists k > 0$ so that $\exists \epsilon > 0$ $\forall n > 0$

$$P(T \leq n + k|\mathcal{F}_n) > \epsilon$$

Then $\mathbb{E}[T] < \infty$

Proof.

$$\mathbb{E}[T] = \sum_j P(T \geq j)$$

Consider J,

$$P(T \geq kJ) = \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{T\geq kJ}\mathbb{1}_{T\geq k(J-1)}|\mathcal{F}_{k(J-1)}\right]\right] = \mathbb{E}\left[\mathbb{1}_{T\geq k(J-1)} \underbrace{\mathbb{E}\left[\mathbb{1}_{T\geq kJ}|\mathcal{F}_{k(J-1)}\right]}_{\leq 1-P(T\leq kJ|\mathcal{F}_{k(J-1)})=1-\epsilon}\right] \leq (1-\epsilon)P(T \geq k(J-1)) \leq (1-\epsilon)^J$$

Thus $P(T \geq kJ)$ decays exponentially. □

6.1 Markov chains

Let $\{X_n\}$ be a stochastic process taking values in $(E, \mathcal{E})$ and $\{\mathcal{F}_n\}$ be a filtration such that $X_n \in \mathcal{F}_n$.

Definition 6.9. $p : E \times \mathcal{E} \to [0,1]$ is a transition kernel on E if

1. $\forall e \in E$ $P(e, \cdot)$ is a probability measure.
2. $\forall A \in \mathcal{E}$ $p(\cdot, A) \in \mathcal{E}$

Definition 6.10. $\{X_n\}$ is a Markov chain with respect to $\mathcal{F}_n$ with transition kernel p if $\forall A \in \mathcal{E}$

$$P(X_{n+1} \in A|\mathcal{F}_n) = p(X_n, A)$$

We can acquire

$$\mathbb{E}[h(X_{n+1}|\mathcal{F}_n)] = \int_E p(X_n, \mathrm{d}e)h(e)$$

Given $(E, \mathcal{E}, p)$ does $\exists X_n$, $\mathcal{F}_n$ Markov with kernel p? The answer is yes.
Let $\Omega = E^{\mathbb{N}}$, $\mathcal{F} = E^{\bigotimes \mathbb{E}}$ and $\mathcal{F}_n = \sigma(X_i(\omega) : i \leq n)$. We want $X(\omega) = \omega_n$.

We want to define law of $\{X_n\}$ by first specifying

$$P(A_0 \times A_1 \times \dots A_n \times E \times E \times \dots) = \mathbb{E}\left[\prod_{i=0}^{n} \mathbb{1}_{X_i \in A_i}\right] = \mathbb{E}\left[\prod_{i=0}^{n-1} \mathbb{1}_{X_i \in A_i} \mathbb{E}[\mathbb{1}_{X_n \in A_n} | \mathcal{F}_{n-1}]\right] = \mathbb{E}\left[\prod_{i=0}^{n-1} \mathbb{1}_{X_i \in A_i} p(X_{n-1}, A_n)\right] =$$

$$= \mathbb{E}\left[\prod_{i=0}^{n-2} \mathbb{1}_{X_i \in A_i} \int\limits_{A_{n-1}} p(X_{n-2}, \mathrm{d}e) p(e, A_n)\right] = \dots = \mathbb{E}\left[\mathbb{1}_{X_0 \in A_0} \int\limits_{A_1} \dots \int\limits_{A_n} p(X_0, \mathrm{d}e_1) p(e_1, \mathrm{d}e_2) \dots p(e_{n-1}, \mathrm{d}e_n)\right]$$

Let law of X_0 be μ on E. Then

$$P(A_0 \times A_1 \times \dots A_n \times E \times E \times \dots) = \int\limits_{A_0} \mu(\mathrm{d}e_0) \int\limits_{A_1} p(X_0, \mathrm{d}e_1) \int\limits_{A_2} p(e_1, \mathrm{d}e_2) \dots \int\limits_{A_n} \dots p(e_{n-1}, \mathrm{d}e_n)$$

From now on we'll assume E is either finite or countable. In this case, Markov condition is $\exists p(i,j)$ such that

$$P(X_{n+1} = j | \mathcal{F}_n) = p(X_n, j)$$

And then

$$\mathbb{E}[h(X_{n+1}) | \mathcal{F}_n] = \sum_{j \in E} P(X_n, j) h(j) = p \cdot j$$

where p is matrix and h is a vector.

Definition 6.11. h is called p-superharmonic if $p \cdot h \leq h$ or alternatively, if $Y_n = h(X_n)$ is a p-supermartingale.

Definition 6.12. Let $T_i = \inf \{n \geq 1 : X_n = i\}$.

Definition 6.13. We say X_n is irreducible if $P_i(T_j < \infty) > 0$ where P_i is a law of X_n started with $X_0 = i$ with probability i.

Definition 6.14. We say X_n is irreducible recurrent if $\forall i, j \in E \quad P_i(T_j < \infty) = 1$.

Theorem 6.4. $\{X_n\}$ is irreducible recurrent on E iff all positive superharmonic are constant.

Proof. $\Rightarrow$: Let $\{X_n\}$ is irreducible recurrent on E and h a positive p-superharmonic function. Consider

$$\mathbb{E}_i\left[h\left(X_n^{T_j}\right)\right] \leq h(i)$$

Thus by Fatou

$$\mathbb{E}_i\left[\liminf h\left(X_n^{T_j}\right)\right] \leq h(i)$$

and now $h(i) \leq h(j)$. By symmetry, $h(j) = h(i)$. ∎

How to produce p-harmonic functions? Let $A \subseteq E$ be some set and $g : A \to \mathbb{R}$ be a bounded function. Assume $\forall i P_i(T_A < \infty) = 1$ and let $h(i) = \mathbb{E}_i[g(X_{T_A})]$.

Lemma 6.3. h is p-harmonic on A^c.

Proof. For $i \notin A$, $T_A \geq 1$.

$$P(X_{n+1} \in A | \mathcal{F}_n) = p(X_n, A)$$

$$\forall i \quad P_i(X_{n+1} \in A_i, X_{n+2} \in A_2 | \mathcal{F}_n) = \int\limits_{A_1} p(X_n, \mathrm{d}e_1) \int\limits_{A_2} p(X_n, \mathrm{d}e_2) = P(\hat{X}_1 \in A_1, \hat{X}_2 \in A_2)$$

where $\hat{X}_i = X_{n+i}$.
Moreover

$$P(X_{n+i} \in F | \mathcal{F}_i) = P_{X_n}(\hat{X}_i \in F)$$

Thus

$$\mathbb{E}[g(X_{T_A}) | \mathcal{F}_1] = \mathbb{E}_{X_1}[g(X_{T_A})] = h(X_1)$$

and

$$h(i) = \mathbb{E}_i[\mathbb{E}[g(X_{T_A}) | \mathcal{F}_1]] = \mathbb{E}_i[h(X_1)] = \sum_j p_{ij} h_j = p \cdot h$$

∎

Theorem 6.5 (Martingale convegrence)**.** X is a supermartingale, $\sup_n \mathbb{E}[X_n] < \infty$ then $\lim_{n\to\infty} X_n = X_\infty$ exists a.s. (but not neccesary in L^1)

Definition 6.15 (Uniform integrability)**.** A collection C is uniformly integrable if $\forall \epsilon > 0\ \exists K$ such that $\forall X \in C\ \mathbb{E}[|X|; |X| > K] < \epsilon$.

Example

$$X \in L^1 \Rightarrow C = \{\mathbb{E}[X|\mathcal{J}] : \mathcal{J} \subseteq \mathcal{F} \text{ subalgebra}\}$$

Lemma 6.4. $X \in L^1$ then $\forall \epsilon > 0\ \exists \delta$ such that $P(F) < \delta$ then $\mathbb{E}[|X|; F] < \epsilon$.

Theorem 6.6. $X_n {}_{L^1}\!\to X \iff$

1. $\forall \epsilon > 0\ P(|X|_n - X > \epsilon) \to 0$
2. $\{X_n\}, X$ are uniformly integrable.

Proof. $\Rightarrow$:
From Markov we get

$$\epsilon P(|X_n - X| < \epsilon) \leq \mathbb{E}[|X_n - X|]$$

Now we want to show uniform integrability:

$$\mathbb{E}[|X_n - X + X|; |X|_n > K] \leq \mathbb{E}[|X_n - X|; |X_n| > K] + \mathbb{E}[|X|; |X_n| > K] \leq \epsilon + \mathbb{E}[|X|; |X_n| > K]$$

Choose $K(\delta)$ such that $\sup_n \mathbb{E}|X_n| < \delta(\epsilon)$ use lemma for X to say $\mathbb{E}[|X|; |X_n| > K] < \epsilon$
Says $\forall \epsilon > 0\ \exists n(\epsilon), K(\epsilon)$ such that $\mathbb{E}[|X_n|; |X_n| > K] < \epsilon$. $\Leftarrow$: $\square$

Theorem 6.7. Let M_n be u.i. martingale. Then $\lim_{n\to\infty} M_n = M_\infty$ a.s. and in L_1.

Proof. A.s. convergence follows from 6.5. Since convergence in probability follows from a.s. convergence, we get L_1 convergence. $\square$

Theorem 6.8 (Levy's upwards theorem)**.** Let $\mathcal{F}_n$ filtration $\eta \in L^1$ and $\mathcal{F}_\infty = \sigma(\bigcup_n \mathcal{F}_n)$.
$M_n = \mathbb{E}[\eta|\mathcal{F}_n]$ is a u.i. martingale and $M_n \to M_\infty$ a.s. and L^1 and moreover, $M_\infty = \mathbb{E}[\eta|\mathcal{F}_\infty]$

Proof. If $F \in \mathcal{F}_n\ \forall r > n$:

$$\mathbb{E}[M_r; F] = \mathbb{E}[\mathbb{E}[M_r|\mathcal{F}_n]; F] = \mathbb{E}[M_n; F]$$

Thus

$$\mathbb{E}[M_\infty; F] = \mathbb{E}[M_n; F] = \mathbb{E}[\mathbb{E}[\eta|\mathcal{F}_\infty]; F]$$

So

$$F \mapsto \mathbb{E}[M_\infty; F]$$
$$F \mapsto \mathbb{E}[\mathbb{E}[\eta|\mathcal{F}_\infty]; F]$$

agree on $\bigcup \mathcal{F})n$ by π-λ they agree on $\mathcal{F}_\infty$, $\square$

Definition 6.16 (Backward filtration)**.** $\mathcal{F}_n$ is a backward filtration if $\mathcal{F}_n \supset \mathcal{F}_{n+1}$

Definition 6.17 (Backward martingale)**.**

$$\mathbb{E}[M_n|\mathcal{F}_{n+1}] = M_{n+1}$$

Theorem 6.9 (Levy's downwards theorem)**.** Let $\mathcal{F}_n$ backward filtration $\eta \in L^1$ and $\mathcal{F}_\infty = \bigcap_n \mathcal{F}_n$ $M_n = \mathbb{E}[\eta|\mathcal{F}_n]$ is a u.i. martingale and $M_n \to M_\infty$ a.s. and L^1 and moreover, $M_\infty = \mathbb{E}[\eta|\mathcal{F}_\infty]$

Proof. Recheck upcrossing inequality works for backward martingale. $\square$

Theorem 6.10 (Kolmogorov's zero-one law)**.** Let $\{X_i\}$ independent and $\mathcal{T} = \bigcap_n \sigma(X_n, \dots)$

$$\forall F \in \mathcal{T} \quad P(F) \in \{0, 1\}$$

Proof. Let $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ and $\eta = 1_F$ for $F \in \mathcal{T}$.

$$P(F) = \mathbb{E}[\eta|\mathcal{F}_n]$$

From Levy upward

$$\lim_{n\to\infty} \mathbb{E}[\eta|\mathcal{F}_n] = \mathbb{E}[\eta|\mathcal{F}_\infty] = \eta = \mathbb{1}_F$$

i.e. it's either 0 or 1. $\square$

Theorem 6.11 (Strong law of large numbers)**.** X_i i.i.d. with $\mathbb{E}[|X_n|] < \infty$.

$$\frac{S_n}{n} \to \mathbb{E}[X_1]$$

Proof. Let $\mathcal{F}_n = \sigma(S_n, S_{n+1}, \dots)$.

$$\frac{S_n}{n} = \mathbb{E}\left[\frac{S_n}{n}|\mathcal{F}_n\right] = \frac{\sum \mathbb{E}[X_i|\mathcal{F}_n]}{n} = \mathbb{E}[X_1|\mathcal{F}_n]$$

We've shown that $\mathbb{E}[X_1|\mathcal{F}_n] = \frac{S_n}{n}$. Therefore by Levy downward

$$\frac{S_n}{n} \overset{\text{a.s.},L^1}{\to} A_\infty = \mathbb{E}[X_1|\mathcal{F}_\infty]$$

Since both $\limsup \frac{S_n}{n} \in \mathcal{T}$, $\liminf \frac{S_n}{n} \in \mathcal{T}$ we get $\mathbb{E}[X|\mathcal{F}_\infty] \in \mathcal{T}$. By Kolmogorov 0-1 $\mathbb{E}[X|\mathcal{F}_\infty]$ is constant a.s. and thus equals $\mathbb{E}[X_1]$. □

Let M_n be a martingale with respect to $\mathcal{F}_n$, $X_n = M_n -_{n+1}$ and $q_n = \mathbb{E}\left[X^2|\mathcal{F}_{n-1}\right]$.

Lemma 6.5.

$$\forall r < s \leq t < u \quad \mathbb{E}[(M_u - M_t)(M_s - M_r)] = 0$$

$$\mathbb{E}\left[(M_u - M_t)^2\right] = \sum_{k=t+1}^{u} \mathbb{E}[X_k^2]$$

Proof.

$$\mathbb{E}[M_u - M_t|\mathcal{F}_t] = 0$$

$$\mathbb{E}[(M_u - M_t)(M_s - M_r)|\mathcal{F}_t] = (M_s - M_r)\mathbb{E}[(M_u - M_t)|\mathcal{F}_t]$$

and

$$(M_u - M_t)^2 = \left(\sum_{k=t+1}^{u} X_k\right)^2 = \sum_{k,l=t+1}^{u} X_k X_l$$

□

Proposition 6.12. Let $N_t = M_t^2 - \sum_{i \leq t} q_i$ then N_t is a martingale.

Proof.

$$N_{t+1} - N_t = M_{t+1}^2 - M_t^2 - q_{t+1}$$

$$\langle N_{t+1} - N_t|\mathcal{F}_t\rangle = \left\langle M_{t+1}^2 - M_t^2|\mathcal{F}_t\right\rangle = -q_{t+1}$$

But

$$M_{t+1}^2 = M_t^2 + (M_{t+1} - M_t)^2 + 2M_t(M_{t+1} - M_t)$$

□

Collary 6.12.1.

$$\sup_n \mathbb{E}[M_n^2] < \infty \iff \sum_k \mathbb{E}\left[(M_{t+1} - M_t)^2\right] < \infty$$

Proof.

$$\mathbb{E}(M_{n+k} - M_n)^2 = \sum_{j=n}^{n+k} \mathbb{E}[(M_{j+1} - M_j)^2] \leq \sum_{j=n}^{\infty} \mathbb{E}[(M_{j+1} - M_j)^2]$$

By Fatou

$$\mathbb{E}(M_\infty - M_n)^2 \leq \sum_{j=n}^{\infty} \mathbb{E}[(M_{j+1} - M_j)^2]$$

Moreover by the equality at start of the proof we get

$$\mathbb{E}(M_\infty - M_n)^2 = \sum_{j=n}^{\infty} \mathbb{E}[(M_{j+1} - M_j)^2]$$

Also if $M - 0 = 0$

$$\mathbb{E}M_\infty^2 = \sum_{j=n}^{\infty} \mathbb{E}[X)^2]$$

$\square$

Theorem 6.13. Let X_k be independent random variable such that $\mathbb{E}[X_k] = 0$ and $\sigma_k^2 < \infty$. Let $M_n = \sum X_k$

1. If $\sum \sigma_k^2 < \infty$, $M_n \to M_\infty$ a.s. and L^2.
2. If $|X_i| \leq K < \infty$, then $\sum X_i$ converges a.s.

Proof. Since $\mathbb{E}[X_i] = 0$

$$\mathbb{E}[M_n|\mathcal{F}_{n-1}] = M_{n-1} + \mathbb{E}[X_n|\mathcal{F}_{n-1}]$$

By independence, M_n is $\mathcal{F}_n$ martingale.
Recall N_t and note $q_{k+1} = \mathbb{E}[X_{k+1}] < \infty$.
Define $T_c = \inf\{i : |M_i| \geq c\}$

$$0 = \mathbb{E}\big[N_{\min\{T_c,n\}}\big] = \mathbb{E}[M^2_{\min\{T_c,n\}}] - \mathbb{E}\left[\sum_{k=1}^{\min\{T_c,n\}} \sigma_k^2\right]$$

Thus

$$\mathbb{E}\left[\sum_{k=1}^{\min\{T_c,n\}} \sigma_k^2\right] = \mathbb{E}[M^2_{\min\{T_c,n\}}] \leq (c+k)^2$$

By Fatou

$$\mathbb{E}\left[\sum_{k=1}^{T_c} \sigma_k^2\right] \leq (c+k)^2$$

If $\exists c < \infty$ such that $P(T_c = \infty) = 0$ we are done.
By assumption of a.s. convergence, $M_i \to M_\infty$ thus $\exists c$ such that $P(|M_i| < c) > 0$. $\square$

Lemma 6.6. If $|X_i| \leq K$, $\sum X_i$ converges, then $\sum \mathbb{E}[X_i]$ converges and $\sum \sigma_i^2 < \infty$

Proof. Let $X_i^* = X_i$, then both series converges, and we define $Y_i = X_i - X_i^*$ for which $\mathbb{E}[Y_i] = 0$.
Thus, $\sum \sigma^2(Y_i) < \infty$ and thus $2\sum \sigma^2(X_i) < \infty$. Now from part one, $Z_i = X_i - \mathbb{E}[X_i]$, we get $\sum X_i - \mathbb{E}[X_i]$ converges a.s. and so does $\sum \mathbb{E}[X_i]$ $\square$

Lemma 6.7 (Cesaro's lemma)**.** Let b_n be strictly increasing and positive and $b_0 = 0$. Let $\{V_n\}$ be a convergent sequence $V_n \to V_\infty$. Then

$$\frac{1}{b_n}\sum_{k=1}^{n}(b_k - b_{k-1})V_k \to V_\infty$$

Proof.

$$1 = \sum_{k=1}^{n} \frac{b_k - b_{k-1}}{b_n}$$

$$V_\infty = \sum_{k=1}^{n} \frac{b_k - b_{k-1}}{b_n} V_\infty$$

$$\left|V_\infty - \frac{1}{b_n}\sum_{k=1}^{n}(b_k - b_{k-1})V_k\right| \leq \frac{1}{b_n}\sum_{k=1}^{n}(b_k - b_{k-1})|V_k - V_\infty|$$

$\square$

For $\epsilon > 0$ choose N such that $|V_n - V_\infty| < \epsilon$, let $V_k, V_\infty < M$:

$$\frac{1}{b_n}\sum_{k=1}^{n}(b_k - b_{k-1})|V_k - V_\infty| \leq 2M \cdot \frac{b_N}{b_n} + \epsilon \leq 2\epsilon$$

Lemma 6.8 (Kroniker's lemma)**.** Let $\{b_n\}$ be a sequence increasing to ∞ and $S - n = \sum_{i=1}^n x_i$. Then if

$$\sum \frac{x_n}{b_n}$$

converges, $\frac{S_n}{b_n} \to 0$.

Proof. Let $u_n = \sum_{k=1}^n \frac{x_k}{b_k}$, then $u_n - u_{n-1} = \frac{x_n}{b_n}$ and $u_n \to u_\infty = \sum_{k=1}^\infty \frac{x_k}{b_k}$. Then $x_n = b_n(u_n - u_{n-1})$ and

$$S_n = \sum_{k=1}^n x_k = \sum_{k=1}^\infty b_k(u_k - u_{k-1}) = b_n u_n - \sum_{k=1}^n (b_k - b_{k-1}) u_{k-1}$$

$$\frac{S_n}{b_n} = u_n - \sum_{k=1}^n \left(\frac{b_k - b_{k-1}}{b_n}\right) u_{k-1}$$

By Cesaro we get the required. ☐

Lemma 6.9. Let $\mathbb{E}[W_i] = 0 \sum_k \frac{\mathbb{E}[W_k^2]}{k^2} < \infty$. Then $\frac{\sum_{k=1}^n W_i}{n} \to 0$

Proof. By Kroniker's lemma it's enough to show that $\frac{W_k}{k}$ converges. By previous discussion of random series this follows from $\sum_i \frac{\mathbb{E}W_k^2}{k^2} < \infty$. ☐

Theorem 6.14 (Kolmogorov 3 series theorem)**.** Let $\{X_i\}_{i=1}^\infty$ be independent random variables. Then $\sum_{i=1}^\infty X_i$ converges iff exists K such that

1. $\sum P(|X_i| \geq K) < \infty$
2. $\sum_i \mathbb{E}\left[X_i^K\right]$ convergent
3. $\sum_i \sigma^2(\left[X_i^K\right])$ convergent

Proof. If $\sum X_i$ converges, $\lim |X_i| = 0$ a.s., thus $\forall i > I(k)$ sufficiently large $\sum_i X_i^k$ converges a.s. Since $\sum \sigma^2(X_i^k) < \infty$ by a previous lemma $\sum_i (X_i^k - \mathbb{E}X_i^K)$ converges a.s. But also by second part $\sum_i \mathbb{E}\left[X_i^K\right]$ convergent thus $\sum_{i=1}^\infty X_i$ converges

$$Y_i = X \cdot \mathbb{1}_{|X| \leq i}$$
$$\mathbb{E}[Y_i] = \mathbb{E}[X \mathbb{1}_{|X| \leq i}]$$
$$\lim \mathbb{E}[Y_i] = \mathbb{E}[X]$$

$$P(X_i \neq Y_i) = P(|X_i| \geq i) = P(|X| \geq i)$$

So

$$\sum P(X_i \neq Y_i) = \sum_i P(|X| \geq i) \leq \mathbb{E}[X] < \infty$$

$$\frac{\sigma^2(Y_i)}{i^2} = \frac{\sigma^2(Y_i)}{i^2} \leq \frac{\mathbb{E}[X^2 \leq i]}{i^2}$$

By MCT

$$\sum_i \frac{\sigma^2(Y_i)}{i} = \mathbb{E}\left[X^2 \sum \frac{\mathbb{1}_{|X| \leq i}}{i^2}\right]$$

$$\sum \frac{\mathbb{1}_{|X| \leq i}}{i^2} \leq C \frac{1}{1 + |X|}$$

So

$$\sum_i \frac{\sigma^2(Y_i)}{i} = C\mathbb{E}\left[\frac{X^2}{1 + |X|}\right] < \infty$$

☐

Theorem 6.15 (SLLN)**.** Let $\{X_i\}_{i=1}^\infty$ be iid $X \in L^1$. Let $S_n = \sum_{i=1}^n X_i$ then $\frac{S_n}{n} \xrightarrow{a.s.} \mu = \mathbb{E}[X]$

Proof. Let $Y_i = X \cdot \mathbb{1}_{|X| \leq i}$. Then we have $\frac{1}{n} \sum Y_i \xrightarrow{a.s.} \mu$ ☐

7 Weak convergence and CLT

Modes of convergence

1. $X_n \overset{a.s.}{\to} X$
2. $X_n \overset{prob.}{\to} X$
3. $X_n \overset{dist.}{\to} X$

Definition 7.1. $\mathcal{S}$ is called Polish space if it is complete, separable metric space.

Definition 7.2. Let μ_n and μ be measures on $(\mathcal{S}, \mathcal{B})$. We say $\mu_n \overset{d}{\to} \mu$ if $\forall f \in C_b(\mathcal{S})$:

$$\int_{\mathcal{S}} f(s)\,\mathrm{d}\mu_n(s) \to \int_{\mathcal{S}} f(s)\,\mathrm{d}\mu(s)$$

Reminder $C_b(\mathcal{S})$ is a Banach space with $\|f\|_\infty = \sup_{s\in\mathcal{S}}$. $C_b^*(S)$ is dual space of signed measures and $P(S) \subset C_b^*(S)$ is a closed subspace.

Proposition 7.1. If $X_n \sim \mu_n$ and $X_n \to X$ a.s., then $\mu_n \to \mu$ weakly.

Proof. Let $f \in C_b(\mathbb{R})$

$$f(X_n) \to f(X)$$

a.s. and by BCT

$$\mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$$

□

Also the proposition works for convergence in probability.

We might guess that $F_n(x) \to F(x)$ pointwise is equivalent to weak convergence.

Theorem 7.2. If $\mu_n \overset{w}{\to} \mu$ then $\forall x\, F_n(x) \to F(x)$ in continuity points of F.

Proof. We want

$$\mathbb{E}_{\mu_n} \mathbb{1}_{(-\infty,x)} \to \mathbb{E}_{\mu} \mathbb{1}_{(-\infty,x)}$$

$$\forall \delta \in \mathbb{R}\, \mu_n(f_\delta) \to \mu(f_\delta)$$

$$\mu(f_\delta(x)) \geq \limsup \mu_n(\mathbb{1}_{(-\infty,x)})$$

also $\mu(f_\delta) \leq \mu_n(\mathbb{1}_{(-\infty,x+\delta)}) = F(x+\delta)$

$$\limsup F_n(x) \leq \liminf F(x+\delta) = F(x)$$

we can do the same with negative δ if F is continuous, and thus

$$\mu_n \overset{w}{\to} \mu$$

□

Definition 7.3. Given $(\mathcal{S}_1, \mathcal{B}_1, P_1)$ and $(\mathcal{S}_2, \mathcal{B}_2, P_2)$ a measure Q on $\mathbb{Q}$ is called a coupling of P_1, P_2 if letting $X_i(S_1, S_2) = S_i$ the distribution of X_i under Q is P_i.

Theorem 7.3 (Skorohod representation theorem)**.** Let μ_n, $n \in \mathbb{N}$ be a sequence of probability measures on a metric space S such that μ_n converges weakly to some probability measure μ_∞ on S as $n \to \infty$. Suppose also that the support of μ_∞ is separable. Then there exist random variables X_n defined on a common probability space $(\Omega, \mathcal{F}, \mathbf{P})$ such that the law of X_n is μ_n for all n (including $n = \infty$) and such that X_n converges to X_∞, $\mathbf{P}$-almost surely.

Theorem 7.4 (Helly's selection theorem)**.**

7.1 Characteristic function

Definition 7.4. Given $\mu \sim X$ then $\phi_\mu(\theta) = \int e^{i\theta x}\,\mathrm{d}\mu(x) = \mathbb{E}[e^{i\theta X}]$.

- $\phi(0) = 1$
- $|\phi(\theta)| \leq 1$
- $\phi(\theta)$ is continuous in θ
- $\phi_X(-\theta) = \overline{\phi_X(\theta)}$
- $\phi_{aX+b}(\theta) = e^{i\theta b}\phi_X(a\theta)$

Theorem 7.5.

$$\phi_{X+Y}(\theta) = \phi_X(\theta)\phi_Y(\theta)$$

Theorem 7.6. $X \sim \mu \mapsto \phi_X(\theta)$ is one-to-one and invertible.

Theorem 7.7. $X_n \xrightarrow{w} X$ then $\phi_{X_n} \to \phi_X(\theta)$.

Theorem 7.8 (CLT). Let X_i be iid with mean 0 and variance 1. Let $S_n = \sum_n X_i$ then

$$\frac{S_n}{\sqrt{n}} \xrightarrow{w} N(0,1)$$

Theorem 7.9 (Lévy Inversion). Let $a < b \in \mathbb{R}$. Then

$$\lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-i\theta a} - e^{-i\theta b}}{i\theta}\phi_X(\theta)\,\mathrm{d}\theta = \frac{\mu(\{a\})}{2} + \mu((a,b)) + \frac{\mu(\{b\})}{2} = \frac{1}{2}[F(b) + F(b^-)] - \frac{1}{2}[F(a) + F(a^-)]$$

Proof.

$$\frac{e^{-i\theta a} - e^{-i\theta b}}{i\theta} = \int_a^b e^{-i\theta\lambda}\,\mathrm{d}\lambda$$

$$\frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-i\theta a} - e^{-i\theta b}}{i\theta}\phi_X(\theta)\,\mathrm{d}\theta = \frac{1}{2\pi}\int_{-T}^{T}\left[\int_a^b e^{-i\theta\lambda}\,\mathrm{d}\lambda\right]\phi_X(\theta)\,\mathrm{d}\theta = \int_a^b \mathbb{E}\left[\frac{1}{2\pi}\int_{-T}^{T} e^{i\theta(X-\lambda)}\,\mathrm{d}\theta\right]\mathrm{d}\lambda$$

□