MATH 5214
Intro to Stochastic Analysis

Intro:

Eyvindur Ari Palsson [~~N~~ Kavy]
Harmonic Analysis

Applications:

Uncertainty is all around us.

Course Plan:

[Basics of Probability Theory, Limit Theorems, Markov Processes, Brownian Motion/Wiener Process, Stochastic Diff. Eq??]

Course Policies:

[At most 6 HW sets]
[1 midterm, take home]
[Project - Decide by Sept 18]
[Office Hours]

An Elementary Example

Flipping a fair coin.

Perform n independent tosses

$$X_j = \begin{cases} 1 & \text{if } j\text{-th flip results in a head} \\ 0 & \quad - \quad '' \quad - \quad - \quad \text{tail} \end{cases}$$

$$S_n = X_1 + X_2 + \dots + X_n$$

$$\boxed{\binom{n}{k} = \frac{n!}{k!\,(n-k)!} \qquad n! = 1 \cdot 2 \cdot \ldots \cdot n}$$

Then $\quad \text{Prob}(S_n = k) = \dfrac{\binom{n}{k}}{2^n}$

Do we expect $S_{2n} = n$ as $n \to \infty$?

$$\text{Prob}(S_{2n} = n) = \frac{\binom{2n}{n}}{2^{2n}}$$

$$= \frac{1}{2^{2n}} \frac{(2n)!}{n!\,\underbrace{(2n-n)!}_{n!}}$$

$$\sim \frac{1}{2^{2n}} \frac{\sqrt{2\pi \cdot 2n}\left(\frac{2n}{e}\right)^{2n}}{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \cdot \sqrt{2\pi n}\left(\frac{n}{e}\right)^n}$$

$$= \frac{1}{\sqrt{\pi n}} \longrightarrow 0 \text{ as } n \to \infty$$

Stirling's formula

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n, \; n \to \infty$$

Say $c_n \sim d_n$ as $n \to \infty$ if $\displaystyle\lim_{n\to\infty} \frac{c_n}{d_n} = 1$

We will see later $\text{Prob}\left(\left|\frac{S_{2n}}{2n} - \frac{1}{2}\right| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0$

for any $\varepsilon > 0$. [Weak Law of Large Numbers]

Probability Spaces

A <u>probability space</u> is a triple $(\Omega, \mathcal{F}, \mathbb{P})$.

Ω is called the <u>sample space</u>.

It is the set of all possible outcomes.

$\omega \in \Omega$ is called a sample point.

$\mathcal{F}$ is called a <u>σ-algebra</u> (or σ-field)

It is a collection of subsets of Ω that satisfy

(i) $\Omega \in \mathcal{F}$

(ii) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ $\leftarrow$ $\boxed{A^c := \Omega \setminus A}$

(iii) $\underbrace{A_1, A_2, \ldots}_{\text{countable}} \in \mathcal{F} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$

Call $A \in \mathcal{F}$ an <u>event</u>.

$\mathbb{P}$ is called a <u>probability measure</u>.

It is a (set) function $\mathbb{P}: \mathcal{F} \to [0,1]$ s.t.

(a) $\mathbb{P}(\varnothing) = 0$, $\mathbb{P}(\Omega) = 1$

(b) If $A_1, A_2, \ldots \in \mathcal{F}$ are $\overbrace{\text{pairwise disjoint}}^{A_i \cap A_j = \varnothing \ i \neq j}$ then

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n) \quad \text{[countable additivity]}$$

<u>Note</u>: Call $(\Omega, \mathcal{F})$ a measurable space.

<u>Ex</u>: Fair coin.

$\Omega = \{H, T\}$

$\mathcal{F} = \{\emptyset, \{H\}, \{T\}, \Omega\}$

$\mathbb{P}(\emptyset) = 0, \quad \mathbb{P}(\{H\}) = \mathbb{P}(\{T\}) = \frac{1}{2}, \quad \mathbb{P}(\Omega) = 1$

For n independent tosses

$\Omega = \{H, T\}^n$

$\mathcal{F}$ = all subsets of Ω [There are 2^n of them]

$\mathbb{P}(A) = \frac{|A|}{2^n}$ for $A \subseteq \Omega$ where $|A|$ denotes cardinality of A.

<u>Notation</u>: If $\mathcal{B}$ is a collection of subsets of Ω we denote by $\sigma(\mathcal{B})$ the smallest σ-algebra that contains $\mathcal{B}$.

Ex: Uniform measure on $[0,1]$

$$\Omega = [0,1]$$

[Borel σ-algebra could also use the Lebesgue one]

$$\mathcal{F} = \sigma(\text{collection of all open subintervals of } [0,1])$$

$$\mathbb{P}(A) = \underbrace{\int_0^1 1_A(x)\,dx}_{} \quad \text{for } A \in \mathcal{F}$$

Technically the Lebesgue integral

Gives same answer as Riemann integral

for all $A \in \mathcal{F}$ s.t. 1_A is Riemann integrable.

Properties: $(\Omega, \mathcal{F}, \mathbb{P})$ probability space.

(i) $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$

(ii) $\underset{\in \mathcal{F}}{A} \subseteq \underset{\in \mathcal{F}}{B} \Rightarrow \mathbb{P}(A) \le \mathbb{P}(B)$ [Monotonicity]

(iii) $\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \le \sum_{n=1}^{\infty} \mathbb{P}(A_n)$ [subadditivity]

(iv) If $A_n \uparrow A$, i.e. $A_1 \subseteq A_2 \subseteq \ldots$ and $\bigcup_{n=1}^{\infty} A_n = A$

then $\mathbb{P}(A_n) \uparrow \mathbb{P}(A)$, i.e. $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$ from below

[Continuity from below]

(v) If $A_n \downarrow A$ then $\mathbb{P}(A_n) \downarrow \mathbb{P}(A)$

[Continuity from above]

Proof: (i) $\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c\right)^c$

$\boxed{\begin{array}{l}\text{De Morgan's Law} \\ (A \cap B)^c = A^c \cup B^c\end{array}}$

(ii) $\mathbb{P}(B) = \mathbb{P}(\underbrace{A \cup (B \setminus A)}_{\text{disjoint}}) \overset{\text{def}}{=} \mathbb{P}(A) + \underbrace{\mathbb{P}(B \setminus A)}_{\geq 0} \geq \mathbb{P}(A)$

(iii) Set $B_1 = A_1$, $B_2 = A_2 \setminus A_1$, $B_3 = A_3 \setminus (A_1 \cup A_2), \ldots,$
$B_n = A_n \setminus \bigcup_{k=1}^{n-1} A_k, \ldots$

Then $B_1, B_2, \ldots$ disjoint, $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$ and $B_n \subseteq A_n$

so

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) \overset{\text{def}}{=} \sum_{n=1}^{\infty} \mathbb{P}(B_n) \overset{\text{(ii)}}{\leq} \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

(iv) Set $B_1 = A_1$, $B_2 = A_2 \setminus A_1$, $B_3 = A_3 \setminus A_2, \ldots, B_n = A_n \setminus A_{n-1}, \ldots$
Then $B_1, B_2, \ldots$ disjoint, $\bigcup_{k=1}^{\infty} B_k = \bigcup_{n=1}^{\infty} A_n = A$ and
$\bigcup_{k=1}^{n} B_k = A_n$ so

$$\mathbb{P}(A) = \sum_{k=1}^{\infty} \mathbb{P}(B_k) = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{P}(B_k) = \lim_{n\to\infty} \mathbb{P}(A_n)$$

(v) Note $A_1 \setminus A_n \uparrow A_1 \setminus A$ so by (iv) $\mathbb{P}(A_1 \setminus A_n) \uparrow \mathbb{P}(A_1 \setminus A)$.
Since $A \subseteq A_1$ we have $\mathbb{P}(A_1) = \mathbb{P}(A) + \mathbb{P}(A_1 \setminus A)$
or $\mathbb{P}(A_1 \setminus A) = \mathbb{P}(A_1) - \mathbb{P}(A)$ and likewise
$\mathbb{P}(A_1 \setminus A_n) = \mathbb{P}(A_1) - \mathbb{P}(A_n)$ conclude $\mathbb{P}(A_n) \downarrow \mathbb{P}(A)$. ∎

Conditional Probability

Let $A, B \in \mathcal{F}$ and assume $\mathbb{P}(B) \neq 0$.
The conditional probability of A given B is

$$\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Immediately yields $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\,\mathbb{P}(B)$
but also

Bayes's rule: $$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A)\,\mathbb{P}(B \mid A)}{\mathbb{P}(B)}$$

Bayes's Theorem:
If $A_1, A_2, \ldots$ disjoint and $\bigcup_{n=1}^{\infty} A_n = \Omega$ then

$$\mathbb{P}(A_j \mid B) = \frac{\mathbb{P}(A_j)\,\mathbb{P}(B \mid A_j)}{\sum_{n=1}^{\infty} \mathbb{P}(A_n)\,\mathbb{P}(B \mid A_n)}, \quad j \in \mathbb{N}$$

Random Variables

The Borel σ-algebra on $\mathbb{R}$ is the smallest σ-algebra containing all the open sets.

A <u>random variable</u> (r.v. for short) X is a function $X : \Omega \to \mathbb{R}$ such that $X^{-1}(B) \in \mathcal{F}$ for any B in the Borel σ-algebra on $\mathbb{R}$.

[Real Analysis language: X is measurable]

<u>Facts</u>: If $X, Y, X_1, X_2, \ldots$ are r.v. then so are

$aX + bY$ for constants $a, b \in \mathbb{R}$

XY

X^k for any $k \in \mathbb{N}$

$\inf_n X_n$, $\sup_n X_n$, $\limsup_n X_n$, $\liminf_n X_n$

<u>Notation</u>: $\{X \in B\} := \{\omega \in \Omega : X(\omega) \in B\} = X^{-1}(B)$

The <u>distribution</u> of the r.v. X is a probability measure μ on $\mathbb{R}$ defined for any set B in the Borel σ-algebra on $\mathbb{R}$ by

$$\mu(B) := \mathbb{P}(X \in B) = \mathbb{P} \circ X^{-1}(B)$$

To see μ is a probability measure

$$\mu(\emptyset) = \mathbb{P}(X \in \emptyset) = \mathbb{P}(\emptyset) = 0$$

μ maps into $[0,1]$

$$\mu(\mathbb{R}) = \mathbb{P}(X \in \mathbb{R}) = \mathbb{P}(\Omega) = 1$$

if $A_1, A_2, \ldots$ disjoint then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}\left(X \in \bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} \{X \in A_n\}\right)$$

$\{X \in A_1\}, \{X \in A_2\}, \ldots$ disjoint!

$$= \sum_{n=1}^{\infty} \mathbb{P}(X \in A_n) = \sum_{n=1}^{\infty} \mu(A_n)$$

The <u>distribution function</u> is defined by

$$F(x) := \mathbb{P}(X \leq x) := \mathbb{P}(X \in (-\infty, x]) = \mu((-\infty, x])$$

<u>Properties</u>

(i) F is nondecreasing Idea: $F(x) = \mathbb{P}(X \leq x) \overset{\text{if } x \leq y}{\leq} \mathbb{P}(X \leq y) = F(y)$

(ii) $\lim_{x \to \infty} F(x) = 1$, $\lim_{x \to -\infty} F(x) = 0$ Idea: $x \uparrow \infty$ then $\{X \leq x\} \uparrow \Omega$; $x \downarrow -\infty$ then $\{X \leq x\} \downarrow \emptyset$

(iii) F is right continuous, i.e. $\lim_{y \downarrow x} F(y) = F(x)$

Idea: $y \downarrow x$ then $\{X \leq y\} \downarrow \{X \leq x\}$

(iv) If $F(x^-) = \lim_{y \uparrow x} F(y)$ then $F(x^-) = \mathbb{P}(X < x)$ Idea: If $y \uparrow x$ $\{X \leq y\} \uparrow \{X < x\}$

(v) $\mathbb{P}(X = x) = F(x) - F(x^-)$ Idea: $\mathbb{P}(X = x) = \mathbb{P}(X \leq x) - \mathbb{P}(X < x)$

Fact: If F satisfies properties (i), (ii), (iii) it is the distribution function of some r.v.

Idea: $\Omega = (0,1)$, $\mathcal{F}$ Borel sigma algebra on $(0,1)$ and $\mathbb{P}$ the (Lebesgue) uniform measure. If $\omega \in (0,1)$ let
$$X(\omega) := \sup\{y : F(y) < \omega\}.$$

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

$\supseteq$ bec. if $\omega \leq F(x)$ then $x \notin \{y : F(y) < \omega\}$ so $X(\omega) \leq x$.

$\subseteq$ shown with $\{\omega : X(\omega) \leq x\}^c \supseteq \{\omega : \omega \leq F(x)\}^c$.
Take $\omega > F(x)$ then by (iii) there is $\varepsilon > 0$ s.t. $F(x+\varepsilon) < \omega$ and $X(\omega) \geq x + \varepsilon > x$. $\boxed{?}$

Note: Useful because computers good at generating random variables with uniform distribution. Construction of X allows for random variables with different distribution functions.

Notation: If X and Y r.v. with same distribution μ say X and Y equal in distribution and write $X \stackrel{d}{=} Y$.

<u>Fact</u>: $X \overset{d}{=} Y \iff X$ and Y have same distribution function F

Discrete distributions

$$\underbrace{\Omega = \{\omega_1, \omega_2, \dots\}}_{\text{finite or countable.}}$$

Let $X(\omega_j) = x_j$.

Define the <u>expected value</u> or <u>mean</u> of X

$$\mathbb{E}X = \sum_{j=1}^{\infty} x_j \, \mathbb{P}(X = x_j)$$

In general

$$\mathbb{E}f(X) = \sum_{j=1}^{\infty} f(x_j)\, \mathbb{P}(X = x_j)$$

Important examples are the <u>p-th moments</u> of the distribution

$$\mathbb{E}X^p = \sum_{j=1}^{\infty} x_j^p \, \mathbb{P}(X = x_j)$$

<u>Variance</u> is also important

$$\mathrm{Var}(X) = \overset{\mathbb{E}(X-\mathbb{E}X)^2}{\cancel{(\mathbb{E}X^2 - (\mathbb{E}X)^2)}} = \sum_{j=1}^{\infty} (x_j - \mathbb{E}X)^2 \, \mathbb{P}(X = x_j)$$

Examples

① Bernoulli distribution

$$\mathbb{P}(X=j) = \begin{cases} p & , j=1 \\ q & , j=0 \end{cases}$$

where $p+q=1$ and $p,q \geq 0$. This is a fair coin if $p=q=\frac{1}{2}$. Quickly see

$$\mathbb{E}X = p, \quad Var(X) = pq$$

② Binomial distribution $B(n,p)$

$$\mathbb{P}(X=k) = \binom{n}{k} p^k q^{n-k}, \quad k=0,1,\dots,n$$

where $p+q=1$ and $p,q \geq 0$. This is n independent flips of a fair coin if $p=q=\frac{1}{2}$. See

$$\mathbb{E}X = np \quad Var(X) = npq$$

③ Poisson distribution $\mathcal{P}(\lambda)$

$$\mathbb{P}(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k=0,1,\dots$$

Can see $\mathbb{E}X = \lambda$, $Var(X) = \lambda$

[models number of events in a time interval or number of points that fall in a given set. Limit of binomial $\binom{n}{k} p^k q^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda}$ $\begin{matrix} n\to\infty \\ p\to 0 \\ np\to\lambda \end{matrix}$]

Continuous distributions

Define the <u>expected value</u> or <u>mean</u> in general as

$$\mathbb{E}[X] = \int_\Omega X \, d\mathbb{P}$$

which always makes sense if $X \geq 0$, else if $X^+ = \max\{X, 0\}$, $X^- = \max\{-X, 0\}$ need the subtraction in $\mathbb{E}X := \mathbb{E}X^+ - \mathbb{E}X^-$ to make sense, so either $\mathbb{E}X^+ < \infty$ or $\mathbb{E}X^- < \infty$.

Can generalize to

$$\mathbb{E} f(X) = \int_\Omega f(X) \, d\mathbb{P}$$

Change of variables formula

$$\mathbb{E} f(X) = \int_{\mathbb{R}} f(x) \, d\mu(x)$$

Idea: $\int_\Omega f(X(\omega)) \, d\mathbb{P}(\omega) = \int_{\mathbb{R}} f(x) \, d(\underbrace{\mathbb{P} \circ X^{-1}}_{\mu})(x)$

if $f = 1_B$ $\quad \mathbb{E}\, 1_B(X) = \int_\Omega 1_B(X) \, d\mathbb{P} = \mathbb{P}(X \in B) = \mu(B) = \int_{\mathbb{R}} 1_B(x) \, d\mu(x)$

Etc.

If there exists an integrable function $\rho(x)$ such that

$$\mu(B) = \int_B \rho(x)dx$$

for any B in the Borel σ-algebra of $\mathbb{R}$ then ρ is called the <u>probability density function</u> of X. Then can also calculate

$$\mathbb{E}f(X) = \int_{\mathbb{R}} f(x)\,\rho(x)\,dx$$

<u>Warning</u>: This is not always possible, for example for $\delta_0(B) = \begin{cases} 1 & \text{if } 0 \in B \\ 0 & \text{else} \end{cases}$

<u>Note</u>: $\frac{dF}{dx} = \rho$

<u>Properties</u>: Suppose $X, Y \geq 0$ or $\mathbb{E}|X|, \mathbb{E}|Y| < \infty$

(a) $\mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y$

(b) $\mathbb{E}(aX+b) = a\mathbb{E}X + b$, $a, b \in \mathbb{R}$

(c) If $X \geq Y$ then $\mathbb{E}X \geq \mathbb{E}Y$

<u>Variance</u>

$$\mathrm{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2$$

Useful formula

$$\begin{aligned} Var(X) &= \mathbb{E}(X - \mathbb{E}X)^2 \\ &= \mathbb{E}\left(X^2 - 2(\mathbb{E}X)X + (\mathbb{E}X)^2\right) \\ &= \mathbb{E}X^2 - 2(\mathbb{E}X)^2 + (\mathbb{E}X)^2 \\ &= \mathbb{E}X^2 - (\mathbb{E}X)^2 \end{aligned}$$

immediately conclude $Var(X) \leq \mathbb{E}X^2$ and makes sense if $\mathbb{E}X^2 < \infty$.

As before can discuss p-th moments

$$\mathbb{E}X^p = \int_\Omega X^p \, d\mathbb{P} = \int_{\mathbb{R}} x^p \, d\mu(x)$$

<u>L^p spaces</u>

For r.v. X let

$$\|X\|_p = (\mathbb{E}|X|^p)^{1/p}, \quad p \geq 1$$

and think of $L^p(\Omega)$ as roughly the r.v. X s.t. $\|X\|_p < \infty$. (Hence they will have p^{th}-order moment well defined and finite)

Classic inequalities

(i) Minkowski's inequality

$$\|X+Y\|_p \le \|X\|_p + \|Y\|_p$$

Consequence: $\|\cdot\|_p$ is a norm, $p \ge 1$.

$L^p(\Omega)$ is a Banach space

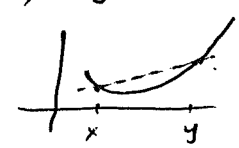
$L^2(\Omega)$ is a Hilbert space with inner product $\langle X, Y \rangle_{L^2(\Omega)} := \mathbb{E}(XY)$

(ii) Hölder's inequality

$$\underbrace{\mathbb{E}|XY|}_{\|XY\|_1} \le \|X\|_p \|Y\|_q \text{ where } \frac{1}{p}+\frac{1}{q}=1, \ p,q \ge 1$$

(iii) Jensen's inequality

$\lambda\phi(x) + (1-\lambda)\phi(y) \ge \phi(\lambda x + (1-\lambda)y)$ for all $0 \le \lambda \le 1$, $x, y \in \mathbb{R}$



Suppose $\phi: \mathbb{R} \to \mathbb{R}$ is $\boxed{\text{convex}}$. Then

$$\mathbb{E}\,\phi(X) \ge \phi(\mathbb{E}X)$$

provided $\mathbb{E}|X|, \mathbb{E}|\phi(X)| < \infty$.

Consequence: $|\mathbb{E}X| \le \mathbb{E}|X|$ and $(\mathbb{E}X)^2 \le \mathbb{E}(X^2)$

Note: By Hölder's or Jensen's inequality get

$$L^p(\Omega) \subseteq L^q(\Omega) \quad \text{if} \quad p \geq q \geq 1.$$

(iv) (Chebyshev's)/Markov's inequality

$$\mathbb{P}(|X| \geq \lambda) \leq \frac{1}{\lambda^p} \mathbb{E}|X|^p$$

for any $\lambda > 0$, $p \geq 1$.

Proof: $\mathbb{E}|X|^p = \int_{\mathbb{R}} |x|^p \, d\mu(x) \geq \int_{|x| \geq \lambda} |x|^p \, d\mu(x) \geq \lambda^p \underbrace{\int_{|x| \geq \lambda} d\mu(x)}_{\mathbb{P}(|X| \geq \lambda)}$ $\square$

Examples

① Uniform distribution on $[0,1]$

Given by probability density function

$$\rho(x) = \begin{cases} 1 & \text{if } x \in [0,1] \\ 0 & \text{else} \end{cases}$$

Quichly see

$$\mathbb{E}X = \frac{1}{2}, \quad Var(X) = \frac{1}{12}$$

② Exponential distribution

$$\rho(x) = \begin{cases} 0 & \text{if } x<0 \\ \lambda e^{-\lambda x} & \text{if } x \geq 0 \end{cases}$$

Can see $\mathbb{E}X = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$

③ Normal distribution

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

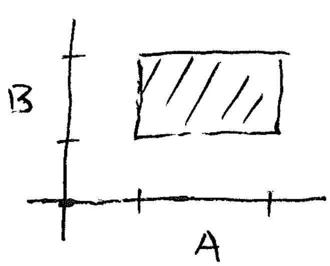
Can see $\mathbb{E}X = \mu$, $Var(X) = \sigma^2$

Product Measures

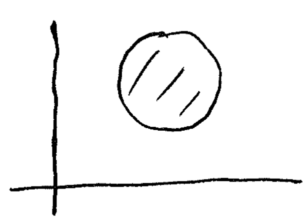
$(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$, $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ probability spaces.

Want to put a probability measure $\mathbb{P}$ on the measurable space $(\Omega, \mathcal{F})$ where

$$\Omega = \Omega_1 \times \Omega_2 \quad \text{and} \quad \mathcal{F} = \sigma(\underbrace{\{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\}}_{\text{"rectangles"}})$$



"rectangle"



$\mathcal{F}$ contains many more sets than "rectangles"

Fact: There is a unique measure $\mathbb{P}$ with

$$\mathbb{P}(A \times B) = \mathbb{P}_1(A)\,\mathbb{P}_2(B) \qquad A \in \mathcal{F}_1,\ B \in \mathcal{F}_2$$

Fubini's Theorem

If $X \geq 0$ or $X \in L^1(\Omega)$ then

$$\int_{\Omega_1}\int_{\Omega_2} X(\omega_1,\omega_2)\, d\mathbb{P}_2(\omega_2)\, d\mathbb{P}_1(\omega_1)$$

$$= \int_{\Omega_1 \times \Omega_2} X(\omega_1,\omega_2)\, d\mathbb{P}(\omega_1,\omega_2)$$

$$= \int_{\Omega_2}\int_{\Omega_1} X(\omega_1,\omega_2)\, d\mathbb{P}_1(\omega_1)\, d\mathbb{P}_2(\omega_2)$$

Warning: Hypothesis important

$\Omega_1 = \Omega_2 = \{1,2,\dots\}$

$\mathcal{F}_1 = \mathcal{F}_2 =$ all subsets

counting measure

not a prob. meas. but compelling still

$X(\omega_1,\omega_2)$

$$\begin{array}{c|ccccc}
3 & 0 & 0 & 1 & -1 & \cdots \\
2 & 0 & 1 & -1 & 0 & \cdots \\
1 & 1 & -1 & 0 & 0 & \cdots \\
\hline
\omega_2 \diagup \omega_1 & 1 & 2 & 3 & 4 &
\end{array}$$

$$\sum_{\omega_1}\sum_{\omega_2} X(\omega_1,\omega_2) = 1$$

but

$$\sum_{\omega_2}\sum_{\omega_1} X(\omega_1,\omega_2) = 0$$

Random vectors

$\mathcal{R}$ Borel σ-algebra on $\mathbb{R}$

$\mathcal{R}^d$ — " ——— $\mathbb{R}^d$

A random vector X is a function $X : \Omega \to \mathbb{R}^d$ s.t. $X^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{R}^d$

Fact: If $X_1, \dots, X_n$ are r.v. and $f : \mathbb{R}^n \to \mathbb{R}$ is $\underbrace{\text{measurable}}_{f^{-1}(B) \in \mathcal{R}^d \text{ all } B \in \mathcal{R}}$ then $f(X_1, \dots, X_n)$ is a r.v.

Ex: If $X_1, \dots, X_n$ are r.v. then so is $X_1 + \dots + X_n$.

Independence

$A, B \in \mathcal{F}$ are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$

X, Y r.v. are independent if

$$\mathbb{P}(\underbrace{X \in C, Y \in D}_{X^{-1}(C) \cap Y^{-1}(D)}) = \mathbb{P}(X \in C)\,\mathbb{P}(Y \in D) \quad \text{all } C, D \in \mathcal{R}$$

Theorem

$$\underbrace{A, B}_{\text{events}} \text{ independent } \Leftrightarrow \underbrace{1_A, 1_B}_{\text{r.v.}} \text{ independent}$$

Idea:

"$\Leftarrow$": $\mathbb{P}(A \cap B) = \mathbb{P}(1_A \in \{1\}, 1_B \in \{1\})$

$$\overset{\text{indep.}}{=} \mathbb{P}(1_A \in \{1\})\mathbb{P}(1_B \in \{1\}) = \mathbb{P}(A)\mathbb{P}(B)$$

"$\Rightarrow$": Take $C, D \in \mathbb{R}$. Then $\{1_A \in C\}$ is $\emptyset, A, A^c$ or Ω and likewise $\{1_B \in D\}$ is $\emptyset, B, B^c$ or Ω.
If $\emptyset$ or Ω involved then trivial, same if the sets are A and B. Let's show how this works if we have A and B^c.

$$\mathbb{P}(A \cap B^c) \overset{\text{disjoint}}{=} \mathbb{P}(A) - \mathbb{P}(A \cap B) \overset{\text{indep.}}{=} \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B)$$

$$= \mathbb{P}(A)\underbrace{(1 - \mathbb{P}(B))}_{\mathbb{P}(B^c)} = \mathbb{P}(A)\mathbb{P}(B^c)$$

$\boxed{?}$

Conclude: First def is a special case of second def.

$X_1, \ldots, X_n$ r.v. are <u>independent</u> if

$$\mathbb{P}(\underbrace{X_1 \in B_1, \ldots, X_n \in B_n}_{\bigcap_{i=1}^{n} \{X_i \in B_i\}}) = \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i) \quad \text{all } B_1, \ldots, B_n \in \mathcal{R}$$

$A_1, \ldots, A_n \in \mathcal{F}$ are <u>independent</u> if

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i) \quad \text{all } I \subseteq \{1, 2, \ldots, n\}$$

<u>Fact</u>: $A_1, \ldots, A_n$ independent $\Leftrightarrow$ $1_{A_1}, \ldots, 1_{A_n}$ independent

<u>Warning</u>: For the second definition it is not enough to have $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\,\mathbb{P}(A_j)$ all $i \neq j$ which is called pairwise independence.

<u>Ex</u>: X_1, X_2, X_3 independent r.v. with $\mathbb{P}(X_i = 0) = \mathbb{P}(X_i = 1) = \frac{1}{2}$ $i = 1, 2, 3$. Let $A_1 = \{X_2 = X_3\}$, $A_2 = \{X_3 = X_1\}$ and $A_3 = \{X_1 = X_2\}$. Then

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(X_1 = X_2 = X_3) = \frac{1}{4} = \mathbb{P}(A_i)\mathbb{P}(A_j)$$

if $i \neq j$ but

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \frac{1}{4} \neq \frac{1}{8} = \mathbb{P}(A_1)\,\mathbb{P}(A_2)\,\mathbb{P}(A_3)$$

Say $X_1, X_2, \ldots$ r.v. are independent if every finite subcollection is and similar for

$A_1, A_2, \ldots \in \mathcal{F}$ being independent.

<u>Note</u>: If start with independent r.v. $X_1, \ldots, X_n$ then all sorts of combinations, using measurable functions, will stay independent.

<u>Ex</u>: X_1 and $X_2 \ldots X_n$ are independent.

<u>Theorem</u>

Suppose X_1, X_2 are independent r.v. with distributions μ_1 and μ_2, then (X_1, X_2) has distribution $\mu_1 \times \mu_2$

<u>Idea</u>: $$\mathbb{P}((X_1, X_2) \in A_1 \times A_2) \overset{\text{indep.}}{=} \mathbb{P}(X_1 \in A_1)\mathbb{P}(X_2 \in A_2)$$
$$= \mu_1(A_1)\mu_2(A_2) =: \mu_1 \times \mu_2(A_1 \times A_2)$$

since know the distribution on all "rectangles" its extension is unique.

<u>Note</u>: Often call the distribution of (X_1, X_2) their <u>joint distribution</u>.

Consequences

$$\mathbb{E} f(X_1, X_2) = \iint f(x_1, x_2)\, d\mu_1(x_1)\, d\mu_2(x_2)$$

and thus

$$\mathbb{E}\big(f_1(X_1)\, f_2(X_2)\big) = \mathbb{E} f_1(X_1)\ \mathbb{E} f_2(X_2)$$

if f, f_1, f_2 measurable and $f, f_1, f_2 \geq 0$ or

$\mathbb{E}|f(X_1, X_2)|, \mathbb{E}|f_1(X_1)|, \mathbb{E}|f_2(X_2)| < \infty$,

where X_1, X_2 independent r.v.

In particular $\quad \mathbb{E} X_1 X_2 = \mathbb{E} X_1\ \mathbb{E} X_2$

This all generalizes to $X_1, \dots, X_n$.

<u>Warning</u>: Can have $\mathbb{E} X_1 X_2 = \mathbb{E} X_1\ \mathbb{E} X_2$ without X_1 and X_2 being independent

X_1 \ X_2	1	0	−1
1	0	a	0
0	b	c	b
−1	0	a	0

where $a, b > 0$, $c \geq 0$ and $2a + 2b + c = 1$

← joint distribution

Then $\mathbb{E} X_1 = a \cdot 1 + (2b + c) \cdot 0 + a \cdot (-1) = 0$, $\mathbb{E} X_2 = b + 0 - b = 0$

and $\mathbb{E} X_1 X_2 = 1 \cdot 0 + (-1) \cdot 0 + 0 \cdot 1 = 0$ but

$\mathbb{P}(X_1 = 1, X_2 = 1) = 0 < ab = \mathbb{P}(X_1 = 1)\,\mathbb{P}(X_2 = 1)$.

<u>Notation</u>: Say r.v. X_1, X_2 with $\mathbb{E}X_1^2, \mathbb{E}X_2^2 < \infty$ that have $\mathbb{E}X_1X_2 = \mathbb{E}X_1\,\mathbb{E}X_2$ are <u>uncorrelated</u>.

<u>Recall</u>: Say X r.v. with prob. density func. ρ then its distribution function is

$$F(z) = \mathbb{P}(X \le z) = \int_{-\infty}^{z} \rho(x)\,dx$$

and vice versa if F given that way it determines ρ.

<u>Theorem</u>: If X_1 and X_2 r.v. are independent with prob. dens. funct. ρ_1 and ρ_2 then the prob. dens. funct. of $X_1 + X_2$ is given by

$$\underbrace{(\rho_1 * \rho_2)}_{\text{convolution}}(x) = \int_{-\infty}^{\infty} \rho_1(x-y)\rho_2(y)\,dy$$

<u>Idea</u>:

$$\mathbb{P}(X_1 + X_2 \le z) = \mathbb{E}\,1_{X+Y\le z}$$

$$= \iint 1_{x+y\le z}\, \rho_1(x)\rho_2(y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} \rho_1(x)\, \rho_2(y)\, dx\, dy$$

$x^{new} = x^{old} + y$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{z} \rho_1(x-y)\, \rho_2(y)\, dx\, dy$$

$$= \int_{-\infty}^{z} \underbrace{\left(\int_{-\infty}^{\infty} \rho_1(x-y)\, \rho_2(y)\, dy \right)}_{\text{prob. dens. func.}} dx$$

$\square$

<u>Recall</u>: Say $X \sim N(\mu, \sigma^2)$ if it has prob. dens. funct.

normal

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

<u>Fact</u>: If $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ are independent then $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

<u>Idea</u>: $X_1 + X_2$ has prob. dens. funct.

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(x-y-\mu_1)^2} \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2\sigma_2^2}(y-\mu_2)^2} dy$$

Calculus

$$= \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}} e^{-\frac{1}{2(\sigma_1^2+\sigma_2^2)}(x-\mu_1-\mu_2)^2}$$

Do independent r.v. exist?

Easy for $X_1, \dots, X_n$.

For $X_1, X_2, \dots$ need Kolmogorov's extension theorem.

Modes of convergence

(a.s.) Almost sure convergence

$X_n \to X$ a.s. if $\mathbb{P}(\overbrace{X_n \to X}^{\{\omega \in \Omega : X_n(\omega) \to X(\omega)\}}) = 1$

(L^p) Convergence in L^p

$X_n \to X$ in L^p if $\underbrace{\mathbb{E}|X_n - X|^p}_{\|X_n - X\|^p_{L^p}} \to 0$ as $n \to \infty$.

(prob) Convergence in probability

$X_n \to X$ in prob. if for any $\varepsilon > 0$

$\mathbb{P}(|X_n - X| > \varepsilon) \to 0$ as $n \to \infty$

distr. Convergence in distribution

$X_n \to X$ in distr. also written as $X_n \xrightarrow{d} X$, $\mu_n \xrightarrow{d} \mu$, $\mu_n \Rightarrow \mu$

if $\mathbb{E} f(X_n) \to \mathbb{E} f(X)$ for any bounded continuous function f on $\mathbb{R}^d$.

Limits and expected values

Fatou's Lemma:

If $X_n \geq 0$ then

$$\liminf_{n\to\infty} \mathbb{E} X_n \geq \mathbb{E} \liminf_{n\to\infty} X_n$$

Monotone Convergence Theorem (MCT)

If $0 \leq X_n \uparrow X$ then $\mathbb{E} X_n \uparrow \mathbb{E} X$

Dominated Convergence Theorem (DCT)

If $X_n \to X$ a.s., $|X_n| \leq Y$ for all n and $\mathbb{E} Y < \infty$ then $\mathbb{E} X_n \to \mathbb{E} X$.

Ex: Uniform probability on $[0,1]$

Set $X_n(\omega) := n\,\mathbb{1}_{[\frac{1}{n},\frac{2}{n}]}$, $n \geq 2$.

Then $X_n(\omega) \to 0 := X(\omega)$ pointwise but

$$\int_0^1 X_n(\omega)\,d\omega = 1 \quad \text{while} \quad \int_0^1 X(\omega)\,d\omega = 0$$

So

$$\lim_{n\to\infty} \mathbb{E} X_n = 1 \quad \text{while} \quad \mathbb{E} \lim_{n\to\infty} X_n = 0$$

Modes of convergence revisited

$$\begin{tikzcd}
\text{a.s.} \arrow[rr, dashed, bend left, "\text{if have MCT, DCT}"] \arrow[dr, Rightarrow] & & L^p \arrow[dl, Rightarrow] \\
& \text{prob} \arrow[ul, dashed, bend left, "\text{along a subseq.}"] \arrow[d, Rightarrow] & \\
& \text{distr.} &
\end{tikzcd}$$

Ideas:

a.s. $\Rightarrow$ prob:

$$\mathbb{P}(|X_n - X| > \varepsilon) = \int_\Omega \underbrace{\mathbb{1}_{\{|X_n - X| > \varepsilon\}}}_{\leq 1 \text{ dominated}}(\omega)\,d\mathbb{P}(\omega) \xrightarrow[n\to\infty]{DCT} 0$$

$L^p \Rightarrow$ prob:

$$\mathbb{P}(|X_n - X| > \varepsilon) \overset{\text{Markov}}{\le} \frac{1}{\varepsilon^p}\mathbb{E}|X_n - X|^p \xrightarrow[n\to\infty]{} 0$$

prob $\Rightarrow$ distr.

Fix a bounded continuous function f.

Say $|f(x)| \le M$ for all $x \in \mathbb{R}^d$. Let $\varepsilon > 0$. Then

$$|\mathbb{E}f(X_n) - \mathbb{E}f(X)|$$

$$\le \mathbb{E}|f(X_n) - f(X)|$$

$$\le \mathbb{E}\Big(\underbrace{|f(X_n) - f(X)|}_{\le \frac{\varepsilon}{2}} 1_{\{|f(X_n)-f(X)| \le \frac{\varepsilon}{2}\}}\Big)$$

$$+ \mathbb{E}\Big(\underbrace{|f(X_n) - f(X)|}_{\le 2M} 1_{\{|f(X_n)-f(X)| > \frac{\varepsilon}{2}\}}\Big)$$

$$\le \frac{\varepsilon}{2}\cdot 1 + 2M\underbrace{\mathbb{P}\Big(|f(X_n) - f(X)| > \frac{\varepsilon}{2}\Big)}_{\le \frac{\varepsilon}{2}\cdot\frac{1}{2M} \text{ if } n \text{ large enough}} \le \varepsilon$$

because

$$\mathbb{P}\Big(|f(X_n) - f(X)| > \frac{\varepsilon}{2}\Big) \le \underbrace{\mathbb{P}(|X_n - X| \ge \delta)}_{\to 0 \text{ bec. conv. in prob}}$$

$$+ \mathbb{P}(X \in B_\delta) \longrightarrow 0$$

where

$$B_\delta = \{x \in \mathbb{R}^d \mid \exists y \in \mathbb{R}^d : |x-y| < \delta, |f(x)-f(y)| > \tfrac{\varepsilon}{2}\}$$

but $B_\delta \downarrow \emptyset$ as $\delta \to 0$ so conclusion follows. $\square$

Ex: Typewriter sequence.

Uniform probability on $[0,1]$

Set $X_n(\omega) = \mathbb{1}_{\left[\frac{n-2^k}{2^k}, \frac{n-2^k+1}{2^k}\right]}$, $k \geq 0$, $2^k \leq n < 2^{k+1}$

Then $\underbrace{X_n \xrightarrow[n\to\infty]{L^1} 0}$ but $X_n \overset{a.s.}{\underset{n\to\infty}{\not\to}} 0$

$\Rightarrow X_n \xrightarrow[n\to\infty]{prob.} 0$

Fact: If $1 \leq p < q$ then $X_n \xrightarrow[n\to\infty]{L^q} X \Rightarrow X_n \xrightarrow[n\to\infty]{L^p} X$

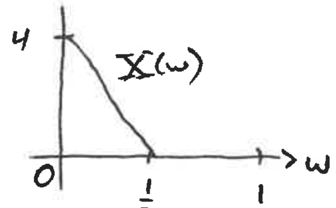
Idea:

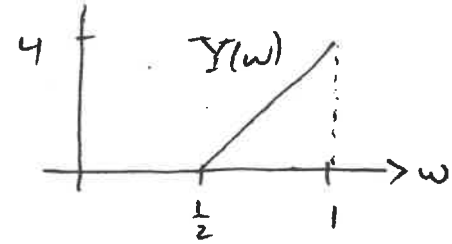
$$\mathbb{E}|X_n - X|^p \leq (\mathbb{E}|X_n - X|^q)^{p/q} \longrightarrow 0$$

$\square$

Note on "in distribution":

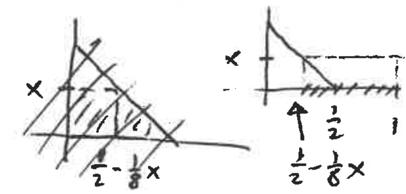
Uniform probability on $[0,1]$





$$\underline{X}(\omega) = \begin{cases} -8\omega + 4 & \text{if } 0 \le \omega \le \frac{1}{2} \\ 0 & \text{else} \end{cases} \qquad \underline{Y}(\omega) = \begin{cases} 0 & \text{else} \\ 8\omega - 4 & \text{if } \frac{1}{2} \le \omega \le 1 \end{cases}$$

$$F_{\underline{X}}(x) = \mathbb{P}(\underline{X} \le x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{2} + \frac{1}{8}x \; \cancel{\frac{1}{2}x - \frac{1}{16}x^2} & \text{if } 0 \le x \le 4 \\ 1 & \text{if } x > 4 \end{cases}$$



but also have $F_{\underline{Y}}(x) = F_{\underline{X}}(x)$ so $\underline{X}$ and $\underline{Y}$ have the same distribution, $\underline{X} \overset{d}{=} \underline{Y}$.

Now set $\underline{X}_i = \underline{X}$ for all $i \ge 1$. Then $\underline{X}_i \xrightarrow[i\to\infty]{d} \underline{X}$ but also $\underline{X}_i \xrightarrow[i\to\infty]{d} \underline{Y}$. To see the latter note that the distributions $\mu_{\underline{X}}$ and $\mu_{\underline{Y}}$ are equal so

$$\mathbb{E} f(\underline{X}_i) = \int_{\Omega} f(\underline{X}_i(\omega))\, d\mathbb{P}(\omega)$$

$$\underset{\text{change of variables}}{=} \int_{\mathbb{R}} f(x)\, d\mu_{\underline{X}}(x) = \int_{\mathbb{R}} f(x)\, d\mu_{\underline{Y}}(x) = \mathbb{E} f(\underline{Y})$$

for all f continuous and bounded.

Conclusion: Limits not unique!

$C_0(\mathbb{R}^d)$ continuous functions on $\mathbb{R}^d$ that vanis at ∞.

$C_b(\mathbb{R}^d)$ continuous and bounded functions on $\mathbb{R}^d$.

$M(\mathbb{R}^d)$ "regular Borel measures" ← good candidates for prob. meas. if also ask for prob. 1

$M(\mathbb{R}^d) = C_0(\mathbb{R}^d)^*$ In some sense $C_0(\mathbb{R}^d)$ functions generate $M(\mathbb{R}^d)$ by the Riesz Representation Theorem

Note

$$X_n \xrightarrow{d} X \iff \mathbb{E} f(X_n) \longrightarrow \mathbb{E} f(X) \quad \forall f \in C_b$$

$$\iff \int_\Omega f(X_n(\omega))\,d\mathbb{P}(\omega) \to \int f(X(\omega))\,d\mathbb{P}(\omega) \quad \forall f \in C_b$$

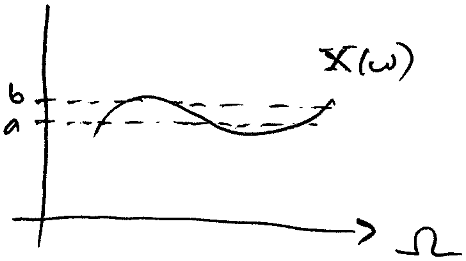
$$\iff \int f(x)\,d\mu_n(x) \to \int f(x)\,d\mu(x) \quad \forall f \in C_b$$

$$\iff \mu_n \xrightarrow{d} \mu$$

sometimes called weak convergence of measures. Here limit is unique

Why not test with $f \in C_0$ instead of $f \in C_b$?

Leads to escape to infinity.



$f \approx 1_{[a,b]}$

$$\mathbb{E} f(X) \approx \mathbb{P}(X \in [a,b]) = \mu([a,b])$$

$f \in C_b$ allows this for all intervals $[a,b]$ while $f \in C_0$ loses control if X gets too large.

Borel - Cantelli Lemma

$(\Omega, \mathcal{F}, \mathbb{P})$

$A_1, A_2, \ldots \in \mathcal{F}.$

Define

$$\lim_{n\to\infty} \overset{\inf}{\cancel{\sup}}\, A_n = \bigcup_{n=1}^{\infty} \bigcap_{j=n}^{\infty} A_j = \{\omega \text{ that are in all but finitely many } A_n\}$$

$$\subseteq$$

$$\lim_{n\to\infty} \overset{\sup}{\cancel{\inf}}\, A_n = \bigcap_{n=1}^{\infty} \bigcup_{j=n}^{\infty} A_j = \underbrace{\{\omega \text{ that are in infinitely many } A_n\}}_{\text{of special interest}}$$

Often write

$$\{A_n \text{ i.o}\} = \{\omega \in \Omega : \omega \in A_n \text{ i.o}\} := \limsup_{n\to\infty} A_n$$

Borel - Cantelli Lemma

$(\Omega, \mathcal{F}, \mathbb{P})$

$A_1, A_2, \ldots \in \mathcal{F}$

Define

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{j=n}^{\infty} A_j = \{\omega \text{ that are in all but finitely many } A_n\}$$

$$\cap\!\!\mid$$

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{j=n}^{\infty} A_j = \underbrace{\{\omega \text{ that are in infinitely many } A_n\}}_{\text{of special interest}}$$

Often write

$$\{A_n \text{ i.o.}\} = \{\omega \in \Omega : \omega \in A_n \text{ i.o.}\} := \limsup_{n\to\infty} A_n$$

Borel-Cantelli Lemma

(1) If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$ then $\mathbb{P}(A_n \text{ i.o.}) = 0$

(2) If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ and $A_1, A_2, \ldots$ are independent then $\mathbb{P}(A_n \text{ i.o.}) = 1$.

Proof:

(1) $\mathbb{P}(A_n \text{ i.o.}) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{j=n}^{\infty} A_j\right) \le \mathbb{P}\left(\bigcup_{j=n}^{\infty} A_j\right) \le \sum_{j=n}^{\infty} \mathbb{P}(A_j) \xrightarrow[n\to\infty]{} 0$.

(2) $$\mathbb{P}\left(\bigcup_{j=n}^{N} A_j\right) = 1 - \mathbb{P}\left(\bigcap_{j=n}^{N} A_j^c\right)$$

indep.
$$= 1 - \prod_{j=n}^{N} \mathbb{P}(A_j^c)$$

$$= 1 - \prod_{j=n}^{N} (1 - \mathbb{P}(A_j))$$

$1-x \le e^{-x}$
$$\ge 1 - \prod_{j=n}^{N} e^{-\mathbb{P}(A_j)}$$

$$= 1 - e^{-\sum_{j=n}^{N} \mathbb{P}(A_j)} \xrightarrow[N\to\infty]{} 1 - 0 = 1 \text{ for all } n$$

so $\mathbb{P}\left(\bigcup_{j=n}^{\infty} A_j\right) = 1$ all n

and since $\bigcup_{j=n}^{\infty} A_j \downarrow \{A_n \text{ i.o.}\}$ get $\mathbb{P}(A_n \text{ i.o.}) = 1$. ◻

Ex: Uniform probability on $[0,1]$.

$A_n = (0, a_n)$ where $a_n \to 0$.

Then $\limsup_{n\to\infty} A_n = \emptyset$ but if $a_n \ge \frac{1}{n}$ have $\sum_{n=1}^{\infty} a_n = \infty$.

Lemma: Let $X_1, X_2, \ldots$ be identically distributed r.v., such that $\mathbb{E}|X_n| < \infty$. Then

$$\lim_{n\to\infty} \frac{X_n}{n} = 0 \quad \text{a.s.}$$

Proof: Let $\varepsilon > 0$ and define $A_n^\varepsilon = \{\omega \in \Omega : |\frac{X_n(\omega)}{n}| > \varepsilon\}$.

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbb{P}(A_n^\varepsilon) &= \sum_{n=1}^{\infty} \mathbb{P}(|X_n| > n\varepsilon) \\
&= \sum_{n=1}^{\infty} \mathbb{P}(|X_1| > n\varepsilon) \\
&= \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} \mathbb{P}(k\varepsilon < |X_1| \le (k+1)\varepsilon) \\
&= \sum_{k=1}^{\infty} k\, \mathbb{P}(k\varepsilon < |X_1| \le (k+1)\varepsilon) \\
&= \sum_{k=1}^{\infty} k \int_{k\varepsilon < |X_1| \le (k+1)\varepsilon} d\mathbb{P}(\omega) \\
&\le \frac{1}{\varepsilon} \sum_{k=1}^{\infty} \int_{k\varepsilon < |X_1| \le (k+1)\varepsilon} |X_1|\, d\mathbb{P}(\omega) \\
&= \frac{1}{\varepsilon} \int_{\varepsilon < |X_1|} |X_1|\, d\mathbb{P}(\omega) \le \frac{1}{\varepsilon} \mathbb{E}|X_1| < \infty
\end{aligned}$$

Define $B_\varepsilon = \{A_n^\varepsilon \text{ i.o.}\}$. Then $\mathbb{P}(B_\varepsilon) = 0$.

Let $B = \bigcup_{n=1}^{\infty} B_{\frac{1}{n}}$. Then $\mathbb{P}(B) = 0$ and

$$\lim_{n\to\infty} \frac{X_n(\omega)}{n} = 0 \quad \text{if} \quad \omega \notin B. \quad \square$$

<u>Theorem</u>

Convergence in probability implies a.s. convergence along a subsequence.

<u>Proof</u>: WLOG assume $X_n \xrightarrow{\text{prob}} 0$.

Take $n_k \in \mathbb{N}$, increasing with k, s.t.

$$\mathbb{P}\left(|X_{n_k}| \geq \frac{1}{k}\right) \leq \frac{1}{2^k} \qquad k \in \mathbb{N}$$

which is possible because $X_n \xrightarrow{\text{prob}} 0$.

For any fixed $\varepsilon > 0$ there is $k_0 \in \mathbb{N}$ s.t. $\frac{1}{k_0} \leq \varepsilon$

and

$$\sum_{k=1}^{\infty} \mathbb{P}(|X_{n_k}| \geq \varepsilon) \leq \sum_{k=1}^{k_0} \mathbb{P}(|X_{n_k}| \geq \varepsilon) + \sum_{k=k_0+1}^{\infty} \mathbb{P}\left(|X_{n_k}| \geq \frac{1}{k}\right)$$

$$\leq \sum_{k=1}^{k_0} \mathbb{P}(|X_{n_k}| \geq \varepsilon) + \sum_{k=k_0+1}^{\infty} \frac{1}{2^k} < \infty$$

By Borel-Cantelli lemma

$$\mathbb{P}(|X_{n_k}| \geq \varepsilon \ \text{ i.o.}) = 0.$$

Thus $X_{n_k} \xrightarrow[k\to\infty]{} 0$ a.s. ∎

Lemma:

If $Y \geq 0$ and $p > 0$ then

$$\mathbb{E}(Y^p) = \int_0^\infty p y^{p-1} \mathbb{P}(Y > y)\, dy$$

Idea: $\int_0^\infty p y^{p-1} \mathbb{P}(Y > y)\, dy$

$$= \int_0^\infty \int_\Omega p y^{p-1} 1_{Y > y}\, d\mathbb{P}\, dy$$

Fubini

$$= \int_\Omega \int_0^\infty p y^{p-1} 1_{Y > y}\, dy\, d\mathbb{P}$$

$$= \int_\Omega \int_0^Y p y^{p-1}\, dy\, d\mathbb{P} = \int_\Omega Y^p\, d\mathbb{P} = \mathbb{E}(Y^p) \quad \boxtimes$$

Notation:

i.i.d. = independent and identically distributed.

Proposition:

If $X_1, X_2, \ldots$ are i.i.d with $\mathbb{E}|X_i| = \infty$ then

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

Idea:

$$\infty = \mathbb{E}|X_1| = \int_0^\infty \mathbb{P}(|X_1| > x)\,dx \le \sum_{n=0}^{\infty} \mathbb{P}(|X_1| > n) \overset{\text{i.d.}}{=} \sum_{n=0}^{\infty} \mathbb{P}(|X_n| > n)$$

thus by Borel-Cantelli $\mathbb{P}(|X_n| \ge n \text{ i.o.}) = 1$
where we use $A_n = \{|X_n| > n\}$ are independent. ☐

Proposition

If $X_1, X_2, \dots$ are i.i.d with $\mathbb{E}|X_i| = \infty$ and $S_n = X_1 + \dots + X_n$
then

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{S_n}{n} \text{ exists in } \mathbb{R}\right) = 0$$

Idea:

Let $C = \{\omega \in \Omega : \lim_{n\to\infty} \frac{S_n}{n} \text{ exists in } \mathbb{R}\}$.

For any $\omega \in C$ then $\lim_{n\to\infty} \frac{S_n(\omega)}{n(n+1)} = 0$.

Assume $C \cap \{\omega \in \Omega : |X_n| \ge n \text{ i.o}\} \neq \emptyset$

Take ω in the intersection, then

$$\left|\frac{S_n}{n} - \frac{S_{n+1}}{n+1}\right| = \left|\frac{S_n}{n} - \frac{S_n + X_{n+1}}{n+1}\right| = \Big|\underbrace{\frac{S_n}{n(n+1)}}_{\downarrow\, 0} - \underbrace{\frac{X_{n+1}}{n+1}}_{\ge 1 \text{ i.o.}}\Big|$$

so $\left|\frac{S_n}{n} - \frac{S_{n+1}}{n+1}\right| > \frac{2}{3}$ i.o. which contradicts $\omega \in C$.

Conclude $C \cap \{|X_n| \ge n \text{ i.o.}\} = \emptyset$ and since $\mathbb{P}(|X_n| \ge n \text{ i.o.}) = 1$
it follows that $\mathbb{P}(C) = 0$.

Ex: "Head runs"

$X_1, X_2, \ldots$ i.i.d. with $\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = \frac{1}{2}$.

$$l_n = \max\{m : X_{n-m+1} = \ldots = X_n = 1\}$$

$$L_n = \max_{1 \le m \le n} l_m$$

Claim $\quad \dfrac{L_n}{\log_2 n} \longrightarrow 1$ a.s.

Idea: $\mathbb{P}(l_n = k) = \frac{1}{2}\cdot\left(\frac{1}{2}\right)^k = \frac{1}{2^{k+1}}$

$$\mathbb{P}(l_n \ge k) = \frac{1}{2^{k+1}} + \frac{1}{2^{k+2}} + \ldots = \frac{1}{2^k}$$

Thus $\mathbb{P}(l_n \ge (1+\varepsilon)\log_2 n) \le n^{-(1+\varepsilon)}$ so $\sum_{n=1}^{\infty} \mathbb{P}(l_n \ge (1+\varepsilon)\log_2 n) < \infty$

and by Borel-Cantelli $l_n \le (1+\varepsilon)\log_2 n$ a.s. for $n \ge N_\varepsilon$.

~~Since~~ Then if n large enough also have

$L_n \le (1+\varepsilon)\log_2 n$ a.s. Since $\varepsilon > 0$ arbitrary

conclude $\quad \limsup_{n\to\infty} \dfrac{L_n}{\log_2 n} \le 1$ a.s.

Break the first n trials into disjoint blocks of length $\lfloor (1-\varepsilon)\log_2 n \rfloor + 1$. Get all 1's on a block with probability

$$2^{-(\lfloor (1-\varepsilon)\log_2 n \rfloor + 1)} \ge n^{-(1-\varepsilon)}/2$$

Choose n large enough s.t.

$$\frac{n}{\lfloor (1-\varepsilon)\log_2 n \rfloor + 1} \geq \frac{n}{\log_2 n}$$

Then

$$\mathbb{P}(L_n \leq (1-\varepsilon)\log_2 n)$$

$$\leq \left(1 - n^{-(1-\varepsilon)}/2\right)^{\frac{n}{\log_2 n}}$$

$$\overset{1-x\leq e^{-x}}{\leq} e^{-\frac{n^{\varepsilon}}{2\cancel{n}}\cdot\frac{\cancel{n}}{\log_2 n}} = \underbrace{e^{-\frac{n^{\varepsilon}}{2\log_2 n}}}_{\text{summable in } n}$$

Thus by Borel-Cantelli $L_n \geq (1-\varepsilon)\log_2 n$ a.s.

Since $\varepsilon > 0$ arbitrary conclude $\liminf_{n\to\infty} \frac{L_n}{\log_2 n} \geq 1$ a.s.

Laws of Large Numbers

Def: Family of r.v. $\{X_i\}_{i\in I}$ with $\mathbb{E}X_i^2<\infty$ is said to be uncorrelated if

$$\mathbb{E}(X_iX_j)=\mathbb{E}X_i\,\mathbb{E}X_j \text{ whenver } i\neq j.$$

Lemma: Let $X_1,\dots,X_n$ have $\mathbb{E}X_i^2<\infty$, $i=1,\dots,n$, and be uncorrelated. Then

$$\mathrm{var}(X_1+\dots+X_n)=\mathrm{var}(X_1)+\dots+\mathrm{var}(X_n)$$

Proof: Write $\mu_i=\mathbb{E}X_i$, $i=1,\dots,n$, $S_n=\sum_{i=1}^n X_i$.

$$\begin{aligned}\mathrm{var}(S_n) &= \mathbb{E}\big(S_n-\underbrace{\mathbb{E}(S_n)}_{\mu_1+\dots+\mu_n}\big)^2\\
&=\mathbb{E}\Big(\sum_{i=1}^n(X_i-\mu_i)\Big)^2\\
&=\mathbb{E}\Big(\sum_{i=1}^n\sum_{j=1}^n(X_i-\mu_i)(X_j-\mu_j)\Big)\\
&=\underbrace{\sum_{i=1}^n\mathbb{E}(X_i-\mu_i)^2}_{\mathrm{var}(X_1)+\dots+\mathrm{var}(X_n)}+\underbrace{2\sum_{i=1}^n\sum_{j=1}^{i-1}\underbrace{\mathbb{E}\big((X_i-\mu_i)(X_j-\mu_j)\big)}_{=0}}_{=0}\end{aligned}$$

because

$$\mathbb{E}\big((X_i-\mu_i)(X_j-\mu_j)\big)=\underbrace{\mathbb{E}X_iX_j}_{=\mathbb{E}X_i\,\mathbb{E}X_j=\mu_i\mu_j\text{ bec. uncorrelated}}-\mu_i\overbrace{\mathbb{E}X_j}^{\mu_j}-\mu_j\overbrace{\mathbb{E}X_i}^{\mu_i}+\mu_i\mu_j=0$$

$\square$

Theorem (L^2 Weak Law of Large Numbers)
(L^2 WLLN)

Let $X_1, X_2, \ldots$ be uncorrelated r.v. with $\mathbb{E} X_i = \mu$ and $\mathrm{var}(X_i) \leq C < \infty$. If $S_n = X_1 + \ldots + X_n$ then

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{L^2} \mu$$

Note: Immediatoly implies also $\frac{S_n}{n} \xrightarrow[n\to\infty]{\text{prob}} \mu$.

Proof:

$$\mathbb{E}\left(\frac{S_n}{n} - \underset{\substack{\uparrow \\ = \mathbb{E}\left(\frac{S_n}{n}\right)}}{\mu}\right)^2 = \mathrm{var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\left(\underbrace{\mathrm{var}(X_1)}_{\leq C} + \ldots + \underbrace{\mathrm{var}(X_n)}_{\leq C}\right) \leq \frac{Cn}{n^2} \xrightarrow[n\to\infty]{} 0$$

∎

Note: Immediately works for $X_1, X_2, \ldots$ i.i.d with $\mathbb{E} X_i^2 < \infty$.

Triangular Arrays

$$\begin{array}{lll} X_{1,1} & & \\ X_{2,1} & X_{2,2} & \\ X_{3,1} & X_{3,2} & X_{3,3} \\ \vdots & \vdots & \ddots \end{array} \qquad \begin{array}{l} S_1 = X_{1,1} \\ S_2 = X_{2,1} + X_{2,2} \\ S_3 = X_{3,1} + X_{3,2} + X_{3,3} \\ \vdots \end{array}$$

<u>Prop</u>: Let $\mu_n = \mathbb{E} S_n$, $\sigma_n^2 = \mathrm{var}(S_n)$. If $\frac{\sigma_n^2}{b_n^2} \xrightarrow[n\to\infty]{} 0$ then

$$\frac{S_n - \mu_n}{b_n} \xrightarrow[n\to\infty]{L^2} 0$$

and thus in prob. too.

<u>Proof</u>:

$$\mathbb{E}\left(\frac{S_n - \mu_n}{b_n}\right)^2 = \mathrm{var}\left(\frac{S_n}{b_n}\right) = \frac{1}{b_n^2}\mathrm{var}(S_n) = \frac{\sigma_n^2}{b_n^2} \xrightarrow[n\to\infty]{} 0. \quad \blacksquare$$

<u>Ex</u>: An occupancy problem

Put r balls at random in n boxes.

A_i = i-th box is empty

N_n = number of empty boxes. $N_n = 1_{A_1} + \dots + 1_{A_n}$

$$\mathbb{P}(A_i) = \left(1 - \frac{1}{n}\right)^r$$

$$\mathbb{E} N_n = n\left(1 - \frac{1}{n}\right)^r$$

If $\frac{r}{n} \xrightarrow[n\to\infty]{} c$ then

$$\frac{\mathbb{E} N_n}{n} = \left(1 - \frac{1}{n}\right)^r \xrightarrow[n\to\infty]{} e^{-c}$$

and

$$\frac{\operatorname{var}(N_n)}{n^2} = \frac{1}{n^2}\left(\mathbb{E} N_n^2 - (\mathbb{E} N_n)^2 \right)$$

$$= \frac{1}{n^2}\left(\mathbb{E}\left(\sum_{m=1}^{n} 1_{A_m}\right)^2 - \left(\mathbb{E} \sum_{m=1}^{n} 1_{A_m}\right)^2 \right)$$

$$= \frac{1}{n^2} \sum_{1 \le k, m \le n} \left(\mathbb{P}(A_k \cap A_m) - \mathbb{P}(A_k)\,\mathbb{P}(A_m) \right)$$

$$= \frac{1}{n^2}\left(\underbrace{n(n-1)}_{\substack{k \,\&\, m \\ \text{distinct}}} \underbrace{\left(\underbrace{\left(1-\frac{2}{n}\right)^r}_{\hookrightarrow e^{-2c}} - \underbrace{\left(1-\frac{1}{n}\right)^{2r}}_{\hookrightarrow e^{-2c}} \right)}_{\hookrightarrow 0} + \underbrace{n}_{k=m} \underbrace{\left(\underbrace{\left(1-\frac{1}{n}\right)^r}_{\hookrightarrow e^{-c}} - \underbrace{\left(1-\frac{1}{n}\right)^{2r}}_{e^{-2c}} \right)}_{\hookrightarrow e^{-c} - e^{-2c}} \right)$$

$$\xrightarrow[n\to\infty]{} 0$$

Thus by previous Prop $\frac{N_n - \mathbb{E} N_n}{n} \xrightarrow{L^2} 0$

or by re-arranging terms

$$\frac{N_n}{n} \xrightarrow{L^2} e^{-c}$$

Prop: [Weak law for triangular arrays]

For each n let $X_{n,k}$, $1 \le k \le n$, be independent.

Let $b_n > 0$ with $b_n \to \infty$ and let

$\widetilde{X}_{n,k} = X_{n,k} \mathbb{1}_{|X_{n,k}| \le b_n}$. Suppose as $n \to \infty$

(i) $\sum_{k=1}^{n} \mathbb{P}(|X_{n,k}| > b_n) \to 0$

and

(ii) $$\frac{\sum_{k=1}^{n} \mathbb{E}(\widetilde{X}_{n,k}^2)}{b_n^2} \longrightarrow 0$$

If we let $S_n = X_{n,1} + \dots + X_{n,n}$ and put

$a_n = \sum_{k=1}^{n} \mathbb{E}\widetilde{X}_{n,k}$ then

$$\frac{S_n - a_n}{b_n} \xrightarrow[n\to\infty]{\text{prob}} 0$$

Proof: Let $\widetilde{S}_n = \widetilde{X}_{n,1} + \dots + \widetilde{X}_{n,n}$. Then

$$\mathbb{P}\left(\left|\frac{S_n - a_n}{b_n}\right| > \varepsilon\right) \le \mathbb{P}(S_n \ne \widetilde{S}_n) + \mathbb{P}\left(\left|\frac{\widetilde{S}_n - a_n}{b_n}\right| > \varepsilon\right)$$

and

$$\mathbb{P}(S_n \ne \widetilde{S}_n) \le \mathbb{P}\left(\bigcup_{k=1}^{n} \{X_{n,k} \ne \widetilde{X}_{n,k}\}\right) \le \sum_{k=1}^{n} \mathbb{P}(|X_{n,k}| > b_n) \xrightarrow[n\to\infty]{\text{by (i)}} 0$$

$$\mathbb{P}\left(\left| \frac{\tilde{S}_n - a_n}{b_n} \right| > \varepsilon \right) \overset{\text{Markov}}{\leq} \frac{1}{\varepsilon^2} \mathbb{E} \left| \frac{\tilde{S}_n - a_n}{b_n} \right|^2$$

$$\overset{a_n = \mathbb{E}\tilde{S}_n}{=} \frac{1}{\varepsilon^2} \frac{1}{b_n^2} \operatorname{var}(\tilde{S}_n)$$

$$\overset{\text{independence}}{=} \frac{1}{\varepsilon^2} \frac{1}{b_n^2} \sum_{k=1}^{n} \operatorname{var}(\tilde{X}_{n,k})$$

$$\leq \frac{1}{\varepsilon^2} \frac{\sum_{k=1}^{n} \mathbb{E}(\tilde{X}_{n,k}^2)}{b_n^2} \xrightarrow[n\to\infty]{\text{by (ii)}} 0$$

which completes the proof. ☐

Corollary [A WLLN]

Let $X_1, X_2, \ldots$ be i.i.d. with

$$x\,\mathbb{P}(|X_i| > x) \longrightarrow 0 \quad \text{as } x \to \infty.$$

Let $S_n = X_1 + \ldots + X_n$ and let $\mu_n = \mathbb{E}(X_1 \mathbb{1}_{|X_1| \leq n})$.

Then $\frac{S_n}{n} - \mu_n \xrightarrow[n\to\infty]{\text{prob}} 0$.

Proof: Apply prev. Prop. with $X_{n,k} = X_k$ and $b_n = n$. Need to show assumptions (i) and (ii) in Prop. hold.

(i) $\sum_{k=1}^{n} \mathbb{P}(|X_{n,k}| > n) = n\,\mathbb{P}(|X_i| > n) \xrightarrow[n\to\infty]{} 0$

(ii) $$\frac{\sum_{k=1}^{n} \mathbb{E}(\tilde{X}_{n,k}^2)}{b_n^2} = \frac{\sum_{k=1}^{n} \mathbb{E}((X_k 1_{|X_k| \le n})^2)}{n^2}$$

$$= \frac{n \, \mathbb{E}((X_1 1_{|X_1| \le n})^2)}{n^2}$$

$$\overset{\text{Lemma}}{=} \frac{1}{n} \int_0^\infty 2y \, \mathbb{P}(|X_1| 1_{|X_1| \le n} > y) \, dy$$

$$= \frac{1}{n} \int_0^n 2y \, \mathbb{P}(|X_1| 1_{|X_1| \le n} > y) \, dy$$

$$\le \frac{1}{n} \int_0^n 2y \, \mathbb{P}(|X_1| \cancel{1_{|X_1| \le n}} > y) \, dy$$

$\boxed{\text{Let } \varepsilon > 0. \text{ There exists } t > 0 \text{ s.t. } 2y\,\mathbb{P}(|X_1| > y) < \frac{\varepsilon}{2} \text{ whenever } y > t. \text{ Then for } n > t}$

$$= \frac{1}{n} \int_0^t 2y \, \mathbb{P}(|X_1| > y) \, dy + \frac{1}{n} \int_t^n 2y \, \mathbb{P}(|X_1| > y) \, dy$$

$$\le \underbrace{\frac{2t}{n}}_{\substack{\le \frac{\varepsilon}{2} \\ \text{if } n \\ \text{large enough}}} + \underbrace{\frac{(n-t)}{n}}_{\le 1} \frac{\varepsilon}{2} \le \varepsilon$$

and since $\varepsilon > 0$ arbitrary conclude

$$\frac{\sum_{k=1}^{n} \mathbb{E}(\tilde{X}_{n,k}^2)}{b_n^2} \xrightarrow[n \to \infty]{} 0.$$

$\square$

<u>Ex</u>: Let $X \geq e$ be such that for some $\alpha \geq 0$

$$\mathbb{P}(X > x) = \frac{1}{x(\log x)^{\alpha}}, \quad \forall\, x \geq e.$$

Then

$$\mathbb{E} X^2 = \infty$$

$$\mathbb{E} X = \begin{cases} e + \frac{1}{\alpha - 1} & \text{if } \alpha > 1 \\ \infty & \text{if } 0 \leq \alpha \leq 1 \end{cases}$$

but

$$y\,\mathbb{P}(X > y) = \frac{1}{(\log y)^{\alpha}} \longrightarrow 0 \quad \text{for all } \alpha > 0.$$

<u>Theorem</u> [The Weak Law of Large Numbers]

Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}|X_i| < \infty$.
Let $S_n = X_1 + \ldots + X_n$ and let $\mu = \mathbb{E} X_1$.
Then

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{\text{prob}} \mu$$

<u>Proof</u>:

$$x\,\mathbb{P}(|X_1| > x) \overset{\text{Markov}}{\leq} \mathbb{E}\left(|X_1|\, 1_{|X_1| > x}\right) \xrightarrow{\text{DCT}} 0 \quad \text{as } x \to \infty$$

$$\mu_n = \mathbb{E}\left(X_1\, 1_{|X_1| \leq n}\right) \xrightarrow{\text{DCT}} \mathbb{E} X_1 = \mu \quad \text{as } n \to \infty$$

Finally if $\varepsilon > 0$ then first note that since $\mu_n \xrightarrow[n\to\infty]{} \mu$ there exists N s.t. $n \geq N \Rightarrow |\mu_n - \mu| < \frac{\varepsilon}{2}$ and thus for $n \geq N$

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \leq \mathbb{P}\left(\left|\frac{S_n}{n} - \mu_n\right| + \underbrace{|\mu_n - \mu|}_{\leq \varepsilon/2} > \varepsilon\right)$$

$$\leq \mathbb{P}\left(\left|\frac{S_n}{n} - \mu_n\right| > \frac{\varepsilon}{2}\right) \xrightarrow[n\to\infty]{\text{prev. Corollary}} 0$$

$\boxtimes$

Ex: [A high-dimensional cube is almost the boundary of a ball]

Let $X_1, X_2, \ldots$ be i.i.d. on $(-1,1)$. Let $Y_i = X_i^2$ and note $Y_1, Y_2, \ldots$ i.i.d., $\mathbb{E}Y_i = \frac{1}{3} < \infty$ so by the WLLN

$$\frac{X_1^2 + \ldots + X_n^2}{n} \xrightarrow[n\to\infty]{\text{prob}} \frac{1}{3}$$

Thus for $\varepsilon > 0$

$$\mathbb{P}\left(\left|\frac{X_1^2 + \ldots + X_n^2}{n} - \frac{1}{3}\right| > \frac{\varepsilon}{3}\right) \xrightarrow[n\to\infty]{} 1$$

or

$$\mathbb{P}\left(\sqrt{1-\varepsilon}\sqrt{\tfrac{n}{3}} < \sqrt{X_1^2 + \ldots + X_n^2} < \sqrt{1+\varepsilon}\sqrt{\tfrac{n}{3}}\right) \xrightarrow[n\to\infty]{} 1$$

which implies

$$\mathbb{P}\left((1-\varepsilon)\sqrt{\tfrac{n}{3}} < \sqrt{X_1^2 + \ldots + X_n^2} < (1+\varepsilon)\sqrt{\tfrac{n}{3}}\right) \xrightarrow[n\to\infty]{} 1$$

Let $A_{n,\varepsilon} = \{x \in \mathbb{R}^n : (1-\varepsilon)\sqrt{\frac{n}{3}} < |x| < (1+\varepsilon)\sqrt{\frac{n}{3}}\}$

Conclude, due to uniform probability

measure of cube on "boundary of ball"

$$\frac{|A_{n,\varepsilon} \cap (-1,1)^n|}{2^n} \xrightarrow[n\to\infty]{} 1$$

measure of $(-1,1)^n$

Theorem [Warm-up $\overbrace{\text{Strong Law of Large Numbers}}^{\text{SLLN}}$]

Let $X_1, X_2, \dots$ be i.i.d. with $\mathbb{E} X_i = \mu$ and $\mathbb{E} X_i^4 < \infty$. If $S_n = X_1 + \dots + X_n$ then

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{a.s.} \mu$$

Proof: WLOG $\mu = 0$, else study $X_i' = X_i - \mu$.

$$\mathbb{E} S_n^4 = \mathbb{E} \left(\sum_{i=1}^n X_i\right)^4$$

$$= \mathbb{E} \sum_{1 \le i,j,k,l \le n} X_i X_j X_k X_l$$

$\mathbb{E} X_i = 0$ so
$\mathbb{E}(X_i^3 X_j) = 0$
$\mathbb{E}(X_i^2 X_j X_k) = 0$
$\mathbb{E}(X_i X_j X_k X_l) = 0$
if i,j,k,l distinct

$$= n\, \mathbb{E} X_1^4 + 6 \cdot \frac{n(n-1)}{2} \left(\mathbb{E} X_1^2\right)^2 \le C n^2$$

constant $<\infty$

bec. $\mathbb{E} X_1^4 < \infty$ and thus $\mathbb{E} X_1^2 < \infty$

For all $\varepsilon > 0$ Markov's inequality yields

$$\mathbb{P}(|S_n| > n\varepsilon) \le \mathbb{E}(S_n^4)/(n\varepsilon)^4 \le \frac{c}{\varepsilon^4}\frac{1}{n^2}$$

This is summable in n so by Borel-Cantelli

$$\mathbb{P}\left(\frac{|S_n - 0|}{n} > \varepsilon \text{ i.o.}\right) = 0$$

Since $\varepsilon > 0$ arbitrary proof is complete. $\square$

<u>Theorem</u> [Strong Law of Large Numbers]
Let $X_1, X_2, \ldots$ be pairwise independent, identically distributed random variables with $\mathbb{E}|X_i| < \infty$. Let $\mathbb{E}X_i = \mu$ and $S_n = X_1 + \ldots + X_n$. Then

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{a.s.} \mu$$

<u>Proof</u>:

Step 1: Truncation

Let $Y_n = X_n 1_{|X_n| \le n}$ and $T_n = Y_1 + \ldots + Y_n$

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > n) \le \int_0^{\infty} \mathbb{P}(|X_1| > t)\,dt = \mathbb{E}|X_1| < \infty$$

so by Borel-Cantelli $\mathbb{P}(X_n \neq Y_n \text{ i.o.}) = 0$.

This shows $|S_n(\omega) - T_n(\omega)| \le R(\omega) < \infty$ a.s. for all n so if $\frac{T_n}{n} \to \mu$ a.s. then $\frac{S_n}{n} \to \mu$ a.s.

Step 2 : Subsequence

For $\alpha > 1$ let $k(n) = \lfloor \alpha^n \rfloor$. By Markov's inequality

$$\sum_{n=1}^{\infty} \mathbb{P}(|T_{k(n)} - \mathbb{E}T_{k(n)}| > \varepsilon k(n))$$

$$\le \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{\operatorname{var}(T_{k(n)})}{k(n)^2}$$

indep. & $\mathbb{E}|Y_i|^2 < \infty$

$$= \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{1}{k(n)^2} \sum_{i=1}^{k(n)} \operatorname{var}(Y_i)$$

$$= \frac{1}{\varepsilon^2} \sum_{i=1}^{\infty} \operatorname{var}(Y_i) \underbrace{\sum_{\substack{n \text{ s.t.} \\ k(n) \ge i}} \frac{1}{(k(n))^2}}$$

$k(n) = \lfloor \alpha^n \rfloor \ge \alpha^n/2$, $n \ge 1$

$$\le 4 \sum_{\substack{n \text{ s.t.} \\ \alpha^n \ge i}} \frac{1}{(\alpha^2)^n} \le 4 \cdot \underbrace{\frac{1}{1 - \frac{1}{\alpha^2}}}_{1 + \frac{1}{\alpha^2} + \frac{1}{\alpha^4} + \cdots} \cdot \underbrace{\frac{1}{i^2}}_{\text{dominates first term in geom. ser.}}$$

$$\le \frac{4}{1 - \frac{1}{\alpha^2}} \frac{1}{\varepsilon^2} \underbrace{\sum_{i=1}^{\infty} \frac{\operatorname{var}(Y_i)}{i^2}}_{< \infty \text{ Lemma}} < \infty$$

Thus since $\varepsilon > 0$ arbitrary by Borel-Cantelli

$$\frac{T_{k(n)} - \mathbb{E}\, T_{k(n)}}{k(n)} \xrightarrow[n\to\infty]{\text{a.s.}} 0$$

By DCT $\mathbb{E}\, Y_k \to \mathbb{E}\, X_1 = \mu$ as $k \to \infty$ so

$$\frac{T_{k(n)}}{k(n)} \xrightarrow[n\to\infty]{\text{a.s.}} \mu$$

Step 3: Sandwiching

Writing $X_n = \underbrace{X_n^+}_{\text{positive part}} - \underbrace{X_n^-}_{\text{negative part}}$ and because linearity and X_n^+, X_n^- fulfilling the conditions of the Theorem can WLOG assume $X_n \geq 0$.

Then for $k(n) \leq m < k(n+1)$

$$\frac{T_{k(n)}}{k(n+1)} \leq \frac{T_m}{m} \leq \frac{T_{k(n+1)}}{k(n)}$$

and using $\frac{k(n+1)}{k(n)} \xrightarrow[n\to\infty]{} \alpha$ we get

$$\frac{1}{\alpha}\mu \leq \liminf_m \frac{T_m}{m} \leq \limsup_m \frac{T_m}{m} \leq \alpha\mu$$

but since $\alpha > 1$ arbitrary get $\frac{T_m}{m} \xrightarrow{\text{a.s.}} \mu$. $\square$

<u>Lemma</u>: $\displaystyle\sum_{i=1}^{\infty} \frac{\mathrm{var}(Y_i)}{i^2} < \infty$

<u>Proof</u>: $\displaystyle\sum_{i=1}^{\infty} \frac{\mathrm{var}(Y_i)}{i^2} \le \sum_{i=1}^{\infty} \frac{\mathbb{E} Y_i^2}{i^2}$

$$= \sum_{i=1}^{\infty} \frac{1}{i^2} \int_0^{\infty} 2y\, \mathbb{P}(|Y_i| > y)\, dy$$

truncated

$$= \sum_{i=1}^{\infty} \frac{1}{i^2} \int_0^{\infty} 1_{\{y \le i\}}\, 2y\, \mathbb{P}(|Y_i| > y)\, dy$$

Fubini

$$= \int_0^{\infty} \underbrace{\left(2y \sum_{i=1}^{\infty} \frac{1}{i^2} 1_{\{y \le i\}}\right)}_{\le 4} \mathbb{P}(|Y_i| > y)\, dy$$

$$\le 4 \underbrace{\int_0^{\infty} \mathbb{P}(|Y_i| > y)\, dy}_{\mathbb{E}|Y_i|} \le 4\, \mathbb{E}|X_1| < \infty$$

because if $1 \le y$ then

$$2y \sum_{i=1}^{\infty} \frac{1}{i^2} 1_{\{y \le i\}} = 2y \sum_{i=\lfloor y \rfloor + 1}^{\infty} \frac{1}{i^2} \le 2y \int_{\lfloor y \rfloor}^{\infty} \frac{1}{x^2}\, dx = \frac{2y}{\lfloor y \rfloor} \le 4$$

and if $0 \le y < 1$ then

$$2y \sum_{i=1}^{\infty} \frac{1}{i^2} 1_{\{y < i\}} \le 2 \cdot \underbrace{\sum_{i=1}^{\infty} \frac{1}{i^2}}_{= \frac{\pi^2}{6} \approx 1.6\ldots} \le 4.$$

$\boxtimes$

Ex: Renewal theory

X_i lifetime of i-th light bulp

$T_n = X_1 + \dots + X_n$ time when n-th light bulp burns out

Assume $X_1, X_2, \dots$ i.i.d. with $0 < X_i < \infty$ and $\mathbb{E} X_i = \mu$. Then if

$$\underbrace{N_t = \sup\{n : T_n \leq t\}}_{\text{number of light bulps burnt out by time } t}$$

we have

$$\frac{N_t}{t} \xrightarrow[t\to\infty]{a.s.} \frac{1}{\mu}$$

Idea: SLLN yields $\frac{T_n}{n} \xrightarrow[n\to\infty]{a.s.} \mu$.

By definition

$$T_{N_t} \leq t < T_{N_t + 1}$$

and note $N_t(\omega) \uparrow \infty$ for all $\omega \in \Omega$. Thus

$$\underbrace{\frac{T_{N_t(\omega)}(\omega)}{N_t(\omega)}}_{\hookrightarrow \mu \text{ a.s.}} \leq \frac{t}{N_t(\omega)} \leq \underbrace{\frac{T_{N_t(\omega)+1}(\omega)}{N_t(\omega)+1}}_{\hookrightarrow \mu \text{ a.s.}} \cdot \underbrace{\frac{N_t(\omega)+1}{N_t(\omega)}}_{\hookrightarrow 1}$$

and conclude $\frac{N_t}{t} \xrightarrow[t\to\infty]{a.s.} \frac{1}{\mu}$

Note: General result

If $X_n \xrightarrow[n\to\infty]{a.s.} X_\infty$ and $N_n \xrightarrow[n\to\infty]{a.s.} \infty$ then

$$X_{N_n} \xrightarrow[n\to\infty]{a.s.} X_\infty$$

Warning: Not true if replace a.s. by prob.

Ex: $X_1, X_2, \ldots$ independent with

$$\mathbb{P}(X_n = 1) = a_n \quad \& \quad \mathbb{P}(X_n = 0) = 1 - a_n$$

where $a_n > 0$ s.t. $\sum_{n=1}^{\infty} a_n = \infty$ and $a_n \xrightarrow[n\to\infty]{} 0$

then $X_n \xrightarrow[n\to\infty]{prob} 0$ but if

$$\cancel{N(n)} \; N_n := \inf\{m \geq n : X_m = 1\}$$

then $X_{N_n} = 1$ a.s. using Borel-Cantelli

Ex: [Empirical distribution function]

$X_1, X_2, \ldots$ i.i.d. with distribution function F.

Let
$$F_n(x) = \frac{1}{n} \sum_{m=1}^{n} 1_{X_m \leq x}$$

Then by SLLN $\quad F_n(x) \xrightarrow[n\to\infty]{a.s.} F(x)$

Note: Using right continuity and boundedness of F can even show

$$\sup_x |F_n(x) - F(x)| \xrightarrow[n\to\infty]{a.s.} 0 \quad \text{for all } x$$

This is the Glivenko-Cantelli Theorem.

Ex: X with distribution function F.

Then $X_n = X + \frac{1}{n}$ has distribution function

$$F_n(x) := \mathbb{P}(X + \tfrac{1}{n} \le x) = F(x - \tfrac{1}{n})$$

Thus $F_n(x) \xrightarrow[n\to\infty]{} F(x^-)$ so get convergence to $F(x)$ only at continuity points of F.

Recall:

$$X_n \xrightarrow{d} X \iff \mathbb{E} f(X_n) \to \mathbb{E} f(X) \quad \forall f \in C_b$$

$$\iff \int f(x)\, d\mu_n(x) \to \int f(x)\, d\mu(x) \quad \forall f \in C_b$$

new

$$\iff F_n(x) \longrightarrow F(x) \quad \text{for all } x \text{ that are continuity points of } F$$

Fact: F has at most countably many discontinuities.

Idea: Set $A_n = \{x : F(x) - F(x^-) \geq \frac{1}{n}\}$

Since F non-decreasing, $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$ must have $|A_n| \leq n$. Thus $A = \bigcup_{n=1}^{\infty} A_n$ countable but A is precisely the set of discontinuities of F.

Theorem: If $F_n(x) \longrightarrow F(x)$ for all x that are continuity points of F then there are r.v. Y_n with distribution F_n s.t. $Y_n \xrightarrow[n\to\infty]{a.s.} Y$ where Y has distribution F.

Idea: Consider uniform probability on $(0,1)$.

Let $Y_n(x) := \sup\{y : F_n(y) < x\}$ and similar for Y.

By an early fact Y_n has distribution F_n.

If $y < Y(x)$ then $F(y) < x$ so if n is large enough and y is a point of continuity of F then $F_n(y) < x$ and therefore $Y_n(x) \geq y$. Conclude $\liminf_n Y_n(x) \geq Y(x)$. Similarly $\limsup_n Y_n(x) \leq Y(x)$. ☐

Idea behind $X_n \xrightarrow{d} X \Leftrightarrow F_n(x) \to F(x)$ for all x that are continuity points of F

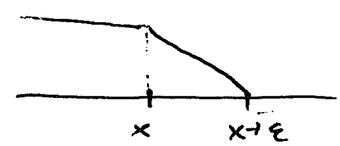
If $F_n(x) \to F(x)$ there exist r.v. Y_n with distribution F_n s.t. $Y_n \xrightarrow{a.s.} Y$ where Y has distribution F. Then

$$\mathbb{E} f(X_n) = \mathbb{E} f(Y_n) \xrightarrow[n\to\infty]{DCT} \mathbb{E} f(Y) = \mathbb{E} f(X) \quad \forall f \in C_b$$

so $X_n \xrightarrow{d} X$.

If $X_n \xrightarrow{d} X$ let

$$g_{x,\varepsilon}(y) = \begin{cases} 1 & y \le x \\ 0 & y \ge x+\varepsilon \\ \text{linear} & x \le y \le x+\varepsilon \end{cases}$$



Then $g_{x,\varepsilon}$ continuous and bounded

$$\limsup_{n\to\infty} \mathbb{P}(X_n \le x) \le \limsup_{n\to\infty} \mathbb{E}\, g_{x,\varepsilon}(X_n)$$
$$= \mathbb{E}\, g_{x,\varepsilon}(X) \le \mathbb{P}(X \le x+\varepsilon)$$

Letting $\varepsilon \to 0$ gives

$$\limsup_{n\to\infty} \mathbb{P}(X_n \le x) \le \mathbb{P}(X \le x)$$

Similarly

$$\liminf_{n\to\infty} \mathbb{P}(X_n \le x) \ge \liminf_{n\to\infty} \mathbb{E}\, g_{x-\varepsilon,\varepsilon}(X_n)$$

$$= \mathbb{E}\, g_{x-\varepsilon,\varepsilon}(X) \ge \mathbb{P}(X \le x-\varepsilon)$$

Letting $\varepsilon \to 0$ gives

$$\liminf_{n\to\infty} \mathbb{P}(X_n \le x) \ge \mathbb{P}(X < x)$$

At points of continuity for F get

$\mathbb{P}(X \le x) = \mathbb{P}(X < x)$. $\boxtimes$

Fourier Transform vs. Characteristic Function

Fourier Transform of r.v. X

$$\hat{X}(\xi) := \mathbb{E}\, e^{-2\pi i \xi X} = \int_{\mathbb{R}} \underbrace{e^{-2\pi i \xi x}}_{\cos(2\pi\xi x) - i\sin(2\pi\xi x)} d\mu(x) =: \hat{\mu}(\xi) \quad , \xi \in \mathbb{R}$$

Characteristic function of a r.v. X

$$\phi(t) = \mathbb{E}\, e^{itX} = \int e^{itx} d\mu(x) \quad , t \in \mathbb{R}$$

Immediately see

$$\hat{X}(\xi) = \phi(-2\pi\xi)$$

so same up to scaling.

Many properties

1) $\hat{X}(0) = 1$

2) $\hat{X}(-\xi) = \overline{\hat{X}(\xi)}$

3) $|\hat{X}(\xi)| \leq \mathbb{E}|e^{-2\pi i \xi X}| \leq 1$

4) $|\hat{X}(\xi+h) - \hat{X}(\xi)| \leq \mathbb{E}|e^{-2\pi i h X} - 1|$

so $\hat{X}$ is uniformly continuous

5) $\mathbb{E}\, e^{-2\pi i (aX+b)} =$

Theorem

If X_1 and X_2 are independent then

$$\widehat{X_1 + X_2} = \hat{X}_1 \hat{X}_2$$

We choose characteristic function (ch.f.)

$$\phi(t) = \mathbb{E}\, e^{itX} = \int_{\mathbb{R}} e^{itx}\, d\mu(x) = \hat{\mu}\left(-\tfrac{t}{2\pi}\right)$$

$\phi(0) = 1$

$\phi(-t) = \overline{\phi(t)}$

$|\phi(t)| \le |\mathbb{E}\, e^{itX}| \le 1$

$\phi(t)$ is uniformly continuous

$\mathbb{E}\, e^{it(aX+b)} = e^{itb}\phi(at)$

Thm: If X_1 and X_2 are independent and have ch. f. ϕ_1 and ϕ_2 then $X_1 + X_2$ has ch. f. $\phi_1(t)\phi_2(t)$

Proof: $\mathbb{E}\, e^{it(X_1+X_2)} = \mathbb{E}\left(e^{itX_1} e^{itX_2}\right) \overset{\text{indep.}}{=} \underbrace{\mathbb{E}\, e^{itX_1}}_{\phi_1(t)}\ \underbrace{\mathbb{E}\, e^{itX_2}}_{\phi_2(t)}$ ☒

Ex: Bernoulli distribution/Coin Flips

If $\mathbb{P}(X=1) = \mathbb{P}(X=-1) = \frac{1}{2}$ then

$$\mathbb{E}\, e^{itX} = e^{it}\cdot\tfrac{1}{2} + e^{-it}\cdot\tfrac{1}{2} = \cos(t)$$

Ex: Normal distribution

If $X \sim N(0,1)$ it has density function

$$\rho(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad \text{[Gaussian]}$$

Then

$$\phi(t) = \mathbb{E}\, e^{itX}$$

$$= \int e^{itx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

$$= \int_{\mathbb{R}} \cos(tx) \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx + \underbrace{i \int_{\mathbb{R}} \underbrace{\sin(tx) \frac{1}{\sqrt{2\pi}} e^{-x^2/2}}_{\text{odd function}}\, dx}_{=0}$$

so

$$\phi'(t) = \int (-x) \sin(tx) \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

int. by parts

$$= \left[\sin(tx) \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right]_{-\infty}^{\infty} - \int t \cos(tx) \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

0

$$= -t\phi(t)$$

Thus $\frac{d}{dt}\left(\phi(t) e^{t^2/2}\right) = \phi'(t) e^{t^2/2} + t\phi(t) e^{t^2/2} = 0$

and hence $\phi(t) e^{t^2/2} = \phi(0) = 1$ so conclude

$$\phi(t) = e^{-t^2/2} \qquad \text{[Gaussian]}$$

In general if $X \sim N(\mu, \sigma^2)$

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

and

$$\phi(t) = e^{i\mu t} e^{-\sigma^2 t^2/2}$$

<u>Note</u>: Recall $\phi(t) = \hat{\mu}(-\frac{t}{2\pi})$.

Can we recover μ from ϕ?

Maybe hope

$$\frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\phi(t)\,dt \overset{???}{=} \underbrace{\mu(x)}_{\text{should be } \mu(E)} \;???$$

<u>Theorem</u>

If $a<b$ then

$$\lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \underbrace{\frac{e^{-ita}-e^{-itb}}{it}}_{=\int_a^b e^{-itx}dx} \phi(t)\,dt = \mu((a,b)) + \frac{1}{2}\mu(\{a,b\})$$

<u>Ex</u>: If $\mu = \delta_0$ so $\mu(E) = \begin{cases} 1 & \text{if } 0 \in E \\ 0 & \text{else} \end{cases}$ then $\phi(t) \equiv 1$

and if $a=-1, b=1$ have $\lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{2\sin(t)}{t}\cdot 1\,dt = \delta_0((-1,1)) + \frac{1}{2}\delta_0(\{-1,1\})$

Theorem:

If $\int |\phi(t)|\,dt < \infty$ then μ has bounded continuous density

$$\rho(x) = \frac{1}{2\pi} \int e^{-itx} \phi(t)\,dt$$

Proof: By prev. Thm

$$\mu((a,b)) + \frac{1}{2}\mu(\{a,b\}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \underbrace{\underbrace{\frac{e^{-ita} - e^{-itb}}{it}}_{\int_a^b e^{-itx}\,dx}}_{\leq |b-a|} \phi(t)\,dt$$

$$\underbrace{\leq \frac{|b-a|}{2\pi} \underbrace{\int_{-\infty}^{\infty} |\phi(t)|\,dt}_{<\infty}}_{\mapsto 0 \text{ as } a \text{ and } b \text{ converge}}$$

Conclude μ has no point masses. Thus

$$\mu((x,x+h)) = \frac{1}{2\pi} \int \frac{e^{-itx} - e^{-it(x+h)}}{it} \phi(t)\,dt$$

$$= \frac{1}{2\pi} \int \left(\int_x^{x+h} e^{-ity}\,dy \right) \phi(t)\,dt$$

$$\overset{\text{Fubini}}{=} \int_x^{x+h} \underbrace{\left(\frac{1}{2\pi}\int e^{-ity}\phi(t)\,dt\right)}_{\text{density which is clearly bounded and continuous.}} dy$$

$\square$

Corollary

If r.v. X and Y have same ch.f. then $X \overset{d}{=} Y$.

Idea: Prev. Thm. if $\int|\phi(t)|dt < \infty$ else use Thm before that which gives μ in terms of ϕ. $\square$

Theorem:

Let $\mu, \mu_1, \mu_2, \dots$ be prob. meas. with ch.f. $\phi, \phi_1, \phi_2, \dots$

(i) If $\mu_n \xrightarrow{d} \mu$ then $\phi_n(t) \underset{n\to\infty}{\longrightarrow} \phi(t)$ for all t.

(ii) If $\phi_n(t) \underset{n\to\infty}{\longrightarrow} \phi(t)$ for all t and ϕ is continuous at 0 then $\mu_n \xrightarrow{d} \mu$

Idea:

(i) $\mu_n \xrightarrow{d} \mu$ means $\int f(x)d\mu_n(x) \to \int f(x)d\mu(x)$ $\forall f \in C_b$ but in particular $e^{itx} \in C_b$ all t so

$\phi_n(t) = \int e^{itx}d\mu_n(x) \longrightarrow \int e^{itx}d\mu(x) = \phi(t)$ all t.

(ii) Most of the effort is spent on showing that probability does not leak to ∞ by showing

$$\mu_n\{x : |x| > \tfrac{2}{u}\} \le \frac{1}{u}\int_{-u}^{u} (\cancel{\phi}\, 1 - \phi_n(t))\, dt$$

$$\downarrow$$

$$\frac{1}{u}\int_{-u}^{u} (1 - \phi(t))\, dt \xrightarrow[u \to 0+]{\text{bec. } \phi \text{ cont. around } 0 \text{ and } \phi(0)=1} 0$$

Then can exploit convergence along subsequences. $\boxed{???}$

Thm: If $\int |x|^n d\mu(x) < \infty$ then its ch. f. ϕ has a continuous derivative of order n

$$\phi^{(n)}(t) = \int (ix)^n e^{itx} d\mu(x)$$

Proof: Integrate by parts. $\boxed{??}$

Consequence:

If $\mathbb{E}|X|^{n+1} < \infty$ then since $\phi^{(n)}(0) = \int (ix)^n d\mu(x) = \mathbb{E}(iX)^n$ we get by Taylor expanding

$$\phi(t) = \sum_{m=0}^{n} \frac{\mathbb{E}(itX)^m}{m!} + \overbrace{o(t^n)}$$

~~quantity that $\to 0$ as $t \to 0$ even when divided by t^n~~

Lemma:

$$\left| e^{ix} - \sum_{m=0}^{n} \frac{(ix)^m}{m!} \right| \leq \min \Bigg\{ \underbrace{\frac{|x|^{n+1}}{(n+1)!}}_{\text{usual Taylor bound}} \, , \quad \frac{2|x|^n}{n!} \Bigg\}$$

new
better for big x

Idea: Integration by parts gives

$$(*) \qquad \int_0^x (x-s)^n e^{is}\,ds = \frac{x^{n+1}}{n+1} + \frac{i}{n+1} \int_0^x (x-s)^{n+1} e^{is}\,ds$$

Thus

$$\frac{e^{ix}-1}{i} = \int_0^x e^{is}\,ds = x + i \int_0^x (x-s)\,e^{is}\,ds$$

or

$$e^{ix} = 1 + ix + i^2 \underbrace{\int_0^x (x-s)\,e^{is}\,ds}_{\text{apply } (*) \text{ with } n=1 \text{ etc.}}$$

eventually end up with

$$e^{ix} - \sum_{m=0}^{n} \frac{(ix)^m}{m!} = \frac{i^{n+1}}{n!} \int_0^x (x-s)^n e^{is}\,ds .$$

First observe $\left| \frac{i^{n+1}}{n!} \int_0^x (x-s)^n e^{is}\,ds \right| \cancel{\leq \frac{1}{n!} |x|\,|x|^n \leq \frac{1}{n!} \int_0^x x^n\,ds}$

$$\leq \frac{1}{n!} \int_0^x (x-s)^n\,ds = \frac{1}{(n+1)!}\, x^{n+1} \quad \text{if} \quad x \geq 0$$

and similar if $x < 0$.

Second observe

$$\frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is}\,ds = \frac{i^n}{(n-1)!}\underbrace{\frac{i}{n}\int_0^x (x-s)^n e^{is}\,ds}$$

$$\overset{\text{int by parts}}{=} -\underbrace{\frac{x^n}{n}}_{=\int_0^x (x-s)^{n-1}ds} + \int_0^x (x-s)^{n-1} e^{is}\,ds$$

$$= \int_0^x (x-s)^{n-1}(e^{is}-1)\,ds$$

$$= \frac{i^n}{(n-1)!}\int_0^x (x-s)^{n-1}(e^{is}-1)\,ds$$

Thus

$$\left|\frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is}\,ds\right| \le \frac{2}{(n-1)!}\int_0^x (x-s)^{n-1}\,ds = \frac{2x^n}{n!} \quad \text{if } x>0$$

and similar if $x<0$. $\square$

Consequence:

$$\Big|\underbrace{\mathbb{E}e^{itX}}_{\phi(t)} - \sum_{m=0}^{n}\frac{\mathbb{E}(itX)^m}{m!}\Big|$$

$$\le \mathbb{E}\left|e^{itX} - \sum_{m=0}^{n}\frac{(itX)^m}{m!}\right|$$

$$\overset{\text{Lemma}}{\le} \mathbb{E}\min\left\{\frac{|tX|^{n+1}}{(n+1)!},\ \frac{2|tX|^n}{n!}\right\}$$

Theorem:

If $\mathbb{E}|X|^2 < \infty$ then

$$\phi(t) = 1 + it\,\mathbb{E}X - \frac{t^2}{2}\mathbb{E}X^2 + \underbrace{o(t^2)}_{\substack{\text{quantity } g(t) \text{ s.t.} \\ \frac{g(t)}{t^2} \to 0 \text{ as } t \to 0.}}$$

Proof: The error term is bounded by

$$t^2\,\underbrace{\mathbb{E}\min\left\{ |t|\frac{|X|^3}{3!}, |X|^2 \right\}}_{\text{bounded by } |X|^2 \text{ \& goes to } 0 \text{ as } t\to 0}$$

$\square$

Theorem [Central Limit Theorem]

Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}X_i = \mu$ and $\sigma^2 = \mathrm{var}(X_i) \in (0,\infty)$. If $S_n = X_1 + \ldots + X_n$ then

$$\frac{S_n - n\mu}{\sigma n^{1/2}} \xrightarrow{d} X$$

where $X \sim \mathcal{N}(0,1)$.

Proof: By considering $X_i' = X_i - \mu$ it suffices to prove the result when $\mu = 0$.
By prev. Thm.

$$\phi(t) = \mathbb{E}\, e^{itX_1} = 1 - \frac{\sigma^2 t^2}{2} + o(t^2)$$

and thus

$$\mathbb{E}\, e^{it\frac{S_n}{\sigma n^{1/2}}} = \mathbb{E}\left(e^{it\frac{X_1}{\sigma n^{1/2}}} \cdots e^{it\frac{X_n}{\sigma n^{1/2}}}\right)$$

$$\overset{\text{i.i.d.}}{=} \Big(1 - \frac{t^2}{2n} + \underbrace{o\left(\frac{t^2}{\sigma^2 n}\right)}_{\xrightarrow[n\to\infty]{} 0}\Big)^n$$

$$\xrightarrow[n\to\infty]{} \underbrace{e^{-t^2/2}}_{\text{ch.f. for } N(0,1)}$$

∎

Note: $$\left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{\sigma^2 n}\right)\right)^n = \left(1 - \frac{t^2/2 + \overbrace{o\left(\frac{t^2}{\sigma^2}\right)}^{\to 0 \text{ as } n\to\infty}}{n}\right)^n$$

General result is that if $c_n \to 0$, $a_n \to \infty$ and $a_n c_n \to \lambda$ then $(1+c_n)^{a_n} \to e^{\lambda}$
because $\frac{\log(1+x)}{x} \longrightarrow 1$ as $x \to 0$ so

$$(1+c_n)^{a_n} = e^{a_n \log(1+c_n)} = e^{\overbrace{a_n c_n}^{\to\lambda} \overbrace{\frac{\log(1+c_n)}{c_n}}^{\to 1}} \longrightarrow e^{\lambda}$$

In our case $a_n = n$ and $c_n = -\frac{t^2}{2} + o\left(\frac{t^2}{\sigma^2}\right)$ ← but this might be in $\mathbb{C}$ ....

Ex: Roulette

Wheel with slots $\underbrace{1, 2, \ldots, 36}_{18 \text{ red}, 18 \text{ black}}, \underbrace{0, 00}_{\text{green}}$

Bet \$1 that ball lands in a red or black slot and win \$1 if it does.

X_i winnings on i-th play.

Then $X_1, X_2, \ldots$ i.i.d. with $\mathbb{P}(X_i = 1) = \frac{18}{38}$, $\mathbb{P}(X_i = -1) = \frac{20}{38}$

$\mu = \mathbb{E} X_i = -\frac{1}{19}$ and $\sigma^2 = \mathrm{var}(X_i) = \mathbb{E} X_i^2 - (\mathbb{E} X_i)^2 = 1 - \left(\frac{1}{19}\right)^2$

Let $S_n = X_1 + \ldots + X_n$. Then

$$\mathbb{P}(S_n \geq 0) = \mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \geq \frac{-n\mu}{\sigma\sqrt{n}}\right)$$

CLT
$n = 361 = 19^2$

$$\simeq \mathbb{P}\left(\underset{\substack{\uparrow \\ N(0,1)}}{X} \geq \frac{-361 \cdot (-1/19)}{\sqrt{1 - \left(\frac{1}{19}\right)^2} \cdot 19}\right)$$

Look up in a table

$$\simeq 1 - 0.8413 = 0.1587$$

So despite $\mathbb{E} S_{361} = -19$ then still 15.87% chance you are still ahead.

Fact: [Rate of convergence]

Let $X_1, X_2, \dots$ be i.i.d with $\mathbb{E} X_i = 0$, $\mathbb{E} X_i^2 = \sigma^2$ and $\mathbb{E}|X_i|^3 = \rho < \infty$. If $F_n(x)$ is the distribution function of $\frac{X_1 + \dots + X_n}{\sigma\sqrt{n}}$ and $F_{N(0,1)}(x)$ is the distribution function for $N(0,1)$ then

$$|F_n(x) - F_{N(0,1)}(x)| \leq \frac{3\rho}{\sigma^3\sqrt{n}}$$

Note: This is the reason why most Monte Carlo methods have a rate of convergence of $O(\frac{1}{\sqrt{n}})$ where n is the sample size.

Note: Trivial if $n \leq 9$ because $\rho \geq \sigma^3$.

Note: Can't get better convergence than in terms of $\frac{1}{\sqrt{n}}$

Ex: $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$ yields

$$F_{2n}(0) = \frac{1}{2}\left(1 + \frac{1}{\sqrt{\pi n}}\right) + o\left(\frac{1}{\sqrt{n}}\right)$$

Extensions of CLT:

- Triangular arrays (still need independence)
- unbounded variance (stable laws)

Random walks

Def: A *stochastic process* (SP) is a collection $\{X_t\}_{t\in T}$ of r.v. on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where T is called the index set. For a fixed $\omega \in \Omega$, $\{X_t(\omega) : t \in T\}$ is called a *sample path*.

Ex: Get discrete-time SP if $T = \mathbb{N}$

(i) $X_1, X_2, \dots$ i.i.d. r.v.

(ii) $\{S_n\}_{n\geq 1}$ where $S_n = X_1 + \dots + X_n$

Here let $\mathcal{F}_n = \underbrace{\sigma(X_1, \dots, X_n)}_{\text{information known up to time } n}$

Def: A *random walk* (RW) on $\mathbb{R}^d$ is a SP of the form

$$S_n = S_0 + X_1 + \dots + X_n$$

where $X_1, X_2, \dots$ i.i.d. and independent of S_0.

Ex: When $d=1$ recall

SLLN $\quad \frac{S_n}{n} \longrightarrow \mathbb{E}X_1$ when $\mathbb{E}|X_1| < \infty$

CLT $\quad \frac{S_n - n\mathbb{E}X_1}{\sqrt{n \operatorname{Var}(X_1)}} \Longrightarrow X \sim N(0,1)$ when $\mathbb{E}|X_1|^2 < \infty$

These are limit theorems.

Now we are also interested in sample path properties involving $S_1(\omega), S_2(\omega), \ldots$

For example for $A \subseteq \mathbb{R}^d$ can we find

$\mathbb{P}(S_n \in A \text{ for some } n \geq 1)$?

$\mathbb{P}(S_n \in A \ \text{ i.o.})$?

$\mathbb{E}\, T_A$ where $T_A = \inf\{n \geq 1 : S_n \in A\}$?

Def: A r.v. $T : \Omega \longrightarrow \{\cancel{0}, 1, 2, \ldots, +\infty\}$ is called a stopping time if

$$\{T \leq n\} \in \mathcal{F}_n \quad \forall n \in \{1, 2, \ldots, \infty\}$$

or equivalently

$$\{T = n\} \in \mathcal{F}_n \quad \forall n \in \{1, 2, \ldots, \infty\}$$

$\boxed{\text{whether to stop at time } n \text{ only depends on the history up to time } n}$

Note: $\{T \leq n\} = \bigcup_{i=1}^{n} \{T = i\}$ and $\{T = n\} = \{T \leq n\} \setminus \{T \leq n-1\}$

Ex: Let $\{S_n\}$ be a RW and $B \in \mathcal{F}$. Then

$$T := \inf\{n \geq 1 : S_n \in B\}$$

is a stopping time. This is often called the hitting time of B.

Thm: [Wald's first identity]

$X_1, X_2, \ldots$ i.i.d. with $\mathbb{E}|X_i| < \infty$, $S_0 = 0$ and T a stopping time with $\mathbb{E}|T| < \infty$. Then

$$\mathbb{E} S_T = \mathbb{E} X_1 \, \mathbb{E} T$$

Proof: Let $U_T = \sum_{i=1}^{T} |X_i|$. Observe

$$\mathbb{E} U_T = \mathbb{E}\left(\sum_{n=1}^{\infty} 1_{\{T=n\}} \sum_{m=1}^{n} |X_m|\right)$$

$$= \sum_{m=1}^{\infty} \sum_{n=m}^{\infty} \mathbb{E}\left(|X_m| 1_{\{T=n\}}\right)$$

$$= \sum_{m=1}^{\infty} \mathbb{E}\left(|X_m| 1_{\{T \geq m\}}\right)$$

$$= \sum_{m=1}^{\infty} \mathbb{E}\left(|X_m| \underbrace{1_{\{T \leq m-1\}^c}}_{\in \mathcal{F}_{m-1} = \sigma(X_1, \ldots, X_{m-1})}\right)$$

indep.

$$= \sum_{m=1}^{\infty} \mathbb{E}|X_m| \, \mathbb{E} 1_{\{T \geq m\}}$$

$$\overset{\text{i.d.}}{=} \mathbb{E}|X_1| \underbrace{\sum_{m=1}^{\infty} \mathbb{P}(T \geq m)}_{= \mathbb{E}T}$$

This completes the proof if $X_i \geq 0$ for all i and justifies using Fubini for same calculations with X_i instead of $|X_i|$. $\blacksquare$

<u>Thm</u>: [Wald's second identity]
$X_1, X_2, \dots$ i.i.d. with $\mathbb{E}|X_i|^2 < \infty$, $\mathbb{E}X_i = 0$, $S_0 = 0$, $Var(X_i) = \sigma^2$ and T a stopping time with $\mathbb{E}|T| < \infty$.
Then
$$\mathbb{E}S_T^2 = \sigma^2 \mathbb{E}T$$

<u>Proof</u>: Similar to the first. $\boxed{?}$

<u>Ex</u>: Simple Random Walk on $\mathbb{Z}$

Let $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{2}$

Let $S_0 = 0$ and take $a < 0 < b$ and define
$$T = \inf\{n \geq 1 : S_n \notin (a,b)\}$$

Then $\mathbb{E}T < \infty$ because

$$\mathbb{P}(T > b-a) \le 1 - \underbrace{\tfrac{1}{2}\cdot\tfrac{1}{2}\cdot\ldots\cdot\tfrac{1}{2}}_{b-a \text{ times}} = 1 - 2^{-(b-a)}$$

and thus

$$\mathbb{P}(T > n(b-a)) \overset{\text{indep.}}{\le} \left(1 - 2^{-(b-a)}\right)^n$$

so

$$\mathbb{E}T = \sum_{k=0}^{\infty} \mathbb{P}(T>k) \overset{\text{monotonicity}}{\le} \sum_{n=0} (b-a)\big(\underbrace{1-2^{-(b-a)}}_{<1}\big)^n < \infty$$

In particular conclude $T < \infty$ a.s. [Escape any interval]

By Wald's first identity

$$a\mathbb{P}(S_T = a) + b\underbrace{\mathbb{P}(S_T = b)}_{1-\mathbb{P}(S_T = a)} = \mathbb{E}S_T = \underbrace{\mathbb{E}X_1}_{=0}\mathbb{E}T = 0$$

so $\mathbb{P}(S_T = a) = \dfrac{b}{b-a}$, $\mathbb{P}(S_T = b) = \dfrac{-a}{b-a}$

Now let

$$T_x := \inf\{n \ge 1 : S_n = x\}$$

Then

$$\mathbb{P}(T_a < \infty) \ge \mathbb{P}(T_a < T_b) = \frac{b}{b-a} \xrightarrow[b\to\infty]{} 1$$

In fact $\mathbb{P}(T_x < \infty) = 1$ for any x.

We come back to where we started a.s.

This property is called <u>recurrence</u>.

Wald's second identity yields

$$\underbrace{\sigma^2}_{=1} \mathbb{E} T = \mathbb{E} S_T^2 = \frac{b}{b-a} a^2 + \frac{-a}{b-a} b^2 = -ab$$

so $\mathbb{E} T = -ab$.

<u>Def</u>: For a stopping time T define the σ-algebra

$$\mathcal{F}_T = \{A : A \cap \{T=n\} \in \mathcal{F}_n \ \forall n\}$$

"information known up to time T."

<u>Theorem</u> [Strong Markov property]

$X_1, X_2, \ldots$ i.i.d. and T stopping time with $\mathbb{P}(T<\infty)>0$. On $\{T<\infty\}$ the sequence $\{X_{T+n}\}_{n\geq 1}$ is independent of $\mathcal{F}_T$ and has the same distribution as the original series.

<u>Idea</u>: For $A \in \mathcal{F}_T$ want to show

$$\mathbb{P}(A, T<\infty, X_{T+j} \in B_j, 1 \le j \le k) = \mathbb{P}(A, T<\infty) \prod_{j=1}^{k} \mathbb{P}(X_j \in B_j).$$

Get this by observing

$$\mathbb{P}(A, T=n, X_{T+j} \in B_j, 1 \le j \le k)$$

$$= \mathbb{P}(\underbrace{A, T=n}_{\in \sigma(X_1,\dots,X_n)}, \underbrace{X_{n+j}}_{\text{independent from } X_1,\dots,X_n} \in B_j, 1 \le j \le k)$$

$$= \mathbb{P}(A, T=n) \prod_{j=1}^{k} \underbrace{\mathbb{P}(X_{n+j} \in B_j)}_{\overset{\text{i.d.}}{=} \mathbb{P}(X_j \in B_j)}$$

and sum in n. $\boxed{?}$

<u>Ex</u>: Simple random walk on $\mathbb{Z}$ is recurrent.

Let $S_0 = 0$ and define

$$T_0^{(1)} = T_0 = \inf\{m > 0 : S_m = 0\}$$

$$T_0^{(n)} = \inf\{m > T_0^{(n-1)} : S_m = 0\} \text{ for } n \ge 2.$$

Note

$$\sum_m \mathbb{P}(S_m = 0) = \mathbb{E}\left(\sum_m 1_{\{S_m = 0\}}\right)$$

$$= \mathbb{E}\Big(\sum_n \mathbb{1}_{\{T_0^{(n)} < \infty\}}\Big)$$

$$= \sum_n \mathbb{P}\big(T_0^{(n)} < \infty\big)$$

Strong Markov prop.

$$= \sum_n \mathbb{P}(T_0 < \infty)^n = \frac{1}{1 - \mathbb{P}(T_0 < \infty)}$$

but also

parity

$$\sum_m \mathbb{P}(S_m = 0) = \sum_n \mathbb{P}(S_{2n} = 0)$$

$$= \sum_n \binom{2n}{n} 2^{-2n}$$

$$= \sum_n \frac{(2n)!}{n!\,n!} 2^{-2n}$$

$$\geq \sum_n \frac{\sqrt{2\pi}\sqrt{2n}\left(\frac{2n}{e}\right)^{2n}}{e^2\, n \left(\frac{n}{e}\right)^{2n}} 2^{-2n}$$

$$= \sum_n \frac{2\sqrt{\pi}}{e^2} \frac{1}{\sqrt{n}} = \infty$$

Stirling's approx
$\sqrt{2\pi}\sqrt{n}\left(\frac{n}{e}\right)^n \leq n! \leq e\sqrt{n}\left(\frac{n}{e}\right)^n$
for all $n \geq 1$

Conclude $\mathbb{P}(T_0 < \infty) = 1$.

<u>Ex</u>: Simple random walk on $\mathbb{Z}^2$ is recurrent.

Same calculations as in $\mathbb{Z}$. Only need to calculate $\underbrace{\mathbb{P}(S_{2n} = (0,0))}_{\text{can think of this as two independent 1-D random walks}} \approx \underset{\uparrow \text{ constant}}{C} \cdot \frac{1}{n}$ } has divergent sum

<u>Ex</u>: Simple random walk on $\mathbb{Z}^3$ is transient (= not recurrent)

Can show $\mathbb{P}(S_{2n} = (0,0,0)) \approx C \cdot \frac{1}{n^{3/2}}$ which has a finite sum. Thus S_n visits $(0,0,0)$ only finitely many times with probability 1.

<u>Ex</u>: Simple random walk on $\mathbb{Z}^d$, $d \geq 3$, is transient

Proof: Note, since the number of steps in opposite directions has to be equal,

$$\begin{aligned}
\mathbb{P}[S_{2n}=0] &= 6^{-2n}\sum_{j,k}\frac{(2n)!}{(j!k!(n-k-j)!)^2} \\
&= 2^{-2n}\binom{2n}{n}\sum_{j,k}\left(3^{-n}\frac{n!}{j!k!(n-k-j)!}\right)^2 \\
&\le 2^{-2n}\binom{2n}{n}\max_{j,k} 3^{-n}\frac{n!}{j!k!(n-k-j)!},
\end{aligned}$$

where we used that $\sum_{j,k} a_{j,k}^2 \le \max_{j,k} a_{j,k} \equiv a^*$ if $\sum_{j,k} a_{j,k} = 1$ and $a_{j,k} \ge 0$. Note that if $j < n/3$ and $k > n/3$ then

$$\frac{(j+1)!(k-1)!}{j!k!} = \frac{j+1}{k} \le 1.$$

That implies that the term in the max is maximized when $j, k, (n-k-j)$ are roughly $n/3$. Using Stirling

$$\frac{n!}{j!k!(n-k-j)!} \sim \frac{n^n}{j^j k^k (n-k-j)^{n-k-j}}\sqrt{\frac{n}{jk(n-k-j)}}\frac{1}{2\pi} \sim Cn^{-1}3^n,$$

if j, k are close to $n/3$. Hence $\mathbb{P}[S_{2n}=0] \sim Cn^{-3/2}$ which is summable and $\mathbb{P}[T_0 < \infty] < 1$. Note that it implies that S_n visits 0 only finitelym any times with probability 1 as the expectation of the number of visits to 0 is $\sum_m \mathbb{P}[S_m = 0]$ (which is then finite). ■

Ex: A note on σ-algebras.

Consider

$$\underbrace{\underline{X} = (\underline{X}_1, \underline{X}_2, \ldots)}_{\text{i.i.d with } \mathbb{P}(\underline{X}_i=-1)=\mathbb{P}(\underline{X}_i=1)=\frac{1}{2}} \in \{-1,1\}^{\mathbb{N}}$$

what is this probability space?

Warning: Pick an event $k = (k_1, k_2, \ldots) \in \{-1,1\}^{\mathbb{N}}$ then $\mathbb{P}(\underline{X}=k) = \lim_{n\to\infty} (\frac{1}{2})^n = 0$. How to proceed?

Say $\Omega = \{H,T\}^{\mathbb{N}}$ and consider events of the form

$$C = \{\omega : \omega_{i_k} = a_k \in \{H,T\},\ k=1,2,\ldots,m\}$$ cylinder set

where $1 \le i_1 < i_2 < \ldots < i_m$, $m \in \mathbb{N}$.

The Kolmogorov extension Thm gives a σ-algebra $\mathcal{F}$ that contains all the cylinder sets and a probability measure $\mathbb{P}$ s.t.

$$\mathbb{P}(C) = \frac{1}{2^m}$$

as expected.

Conditional Expectation

$(\Omega, \mathcal{F}, \mathbb{P})$ probability space

X r.v. with $\mathbb{E}|X| < \infty$

$\mathcal{G} \subseteq \mathcal{F}$ a σ-algebra.

The conditional expectation of X given $\mathcal{G}$, denoted $\mathbb{E}(X|\mathcal{G})$ is any r.v. Y s.t.

(i) Y is $\mathcal{G}$ meas.

(ii) for all $A \in \mathcal{G}$ $\quad \int_A X \, d\mathbb{P} = \int_A Y \, d\mathbb{P}$

Fact: $\mathbb{E}(X|\mathcal{G})$ $\underbrace{\text{exists}}_{\text{Radon-Nikodym}}$ and is $\underbrace{\text{unique}}_{\text{up to a.s.}}$.

Idea: $\mathbb{E}(X|\mathcal{G})$ is the "best guess" of X given the information $\mathcal{G}$.

Ex: $\mathbb{E}(X|\mathcal{F}) = X$.

Ex: Suppose X is independent of $\mathcal{G}$ meaning

$$\mathbb{P}(\{X \in B\} \cap A) = \mathbb{P}(X \in B)\,\mathbb{P}(A) \quad \text{all } A \in \mathcal{G} \text{ and Borel } B$$

then $\mathbb{E}(X \mid \mathcal{G}) = \mathbb{E}X$. Note, for all $A \in \mathcal{G}$

$$\int_A X \, d\mathbb{P} = \mathbb{E} X 1_A \overset{\text{indep.}}{=} \mathbb{E}X \, \mathbb{E} 1_A = \int_A \mathbb{E}X \, d\mathbb{P}.$$

Ex: $\mathbb{E}(X \mid \{\emptyset, \Omega\}) = \mathbb{E}X$

Ex: Suppose

$$\Omega = \underbrace{\Omega_1 \cup \Omega_2 \cup \dots}_{\text{disjoint \& each with positive prob.}}$$

and let $\mathcal{G} = \sigma(\Omega_1, \Omega_2, \dots)$, the σ-algebra generated by $\Omega_1, \Omega_2, \dots$

Then

$$\mathbb{E}(X \mid \mathcal{G}) = \frac{\mathbb{E}(X 1_{\Omega_i})}{\mathbb{P}(\Omega_i)} \quad \text{on } \Omega_i$$

To verify (ii) from def only need to test on $A = \Omega_i$

$$\int_{\Omega_i} \frac{\mathbb{E}(X 1_{\Omega_i})}{\mathbb{P}(\Omega_i)} \, d\mathbb{P} = \mathbb{E}(X 1_{\Omega_i}) = \int_{\Omega_i} X \, d\mathbb{P} \quad \checkmark$$

Further if we define

$$\mathbb{P}(A \mid \mathcal{G}) := \mathbb{E}(1_A \mid \mathcal{G})$$

then see on Ω_i

$$\mathbb{P}(A \mid \mathcal{G}) = \frac{\mathbb{E}(1_A 1_{\Omega_i})}{\mathbb{P}(\Omega_i)} = \frac{\mathbb{P}(A \cap \Omega_i)}{\mathbb{P}(\Omega_i)} = \mathbb{P}(A \mid \Omega_i)$$

Notation: $\mathbb{E}(X \mid Y) := \mathbb{E}(X \mid \underbrace{\sigma(Y)}_{\sigma\text{-algebra generated by } Y})$

Ex: Suppose r.v. X and Y have joint density $f(x,y)$ so

$$\mathbb{P}((X,Y) \in B) = \int_B f(x,y) \, dx\, dy \quad \text{all } B.$$

and suppose for simplicity $\int f(x,y)\,dx > 0$ for all y.

Claim: If $\mathbb{E}|g(X)| < \infty$ then $\mathbb{E}(g(X) \mid Y) = h(Y)$ where $h(y) = \dfrac{\int g(x) f(x,y)\,dx}{\int f(x,y)\,dx}$

Idea: $\mathbb{P}(X = x \mid Y = y) = \dfrac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} = \dfrac{f(x,y)}{\int f(x,y)\,dx}$

and $\mathbb{E}(g(X) \mid Y = y) = \int g(x)\,\mathbb{P}(X = x \mid Y = y)\,dx$

Correct argument:

$h(Y) \in \sigma(Y)$ so (i) holds.

For (ii) observe if $A \in \sigma(Y)$ then

$A = \{\omega : Y(\omega) \in B\}$ for some B so

$$\int_A h(Y)\,d\mathbb{P} = \mathbb{E}(h(Y)\,1_B(Y)) = \iint_B h(y) f(x,y)\,dx\,dy$$
$$= \iint_B g(x) f(x,y)\,dx\,dy$$
$$= \mathbb{E}(g(X)\,1_B(Y))$$
$$= \int_A g(X)\,d\mathbb{P}$$

Ex: Suppose X and Y are independent. Let ϕ be a function with $\mathbb{E}|\phi(X,Y)| < \infty$ and let $g(x) = \mathbb{E}(\phi(x,Y))$. Then

$$\mathbb{E}(\phi(X,Y) \mid X) = g(X)$$

Sol: $g(X) \in \sigma(X)$ so (i) holds

For (ii) observe if $A \in \sigma(X)$ then

$A = \{X \in C\}$ so if μ distribution of X

and ν distribution of Y then

$$\int_A \phi(X,Y)\, d\mathbb{P} = \mathbb{E}\left(\phi(X,Y) 1_c(X)\right)$$

$$= \iint \phi(x,y) 1_c(x)\, d\nu(y)\, d\mu(x)$$

$$= \int 1_c(x) g(x)\, d\mu(x)$$

$$= \int_A g(X)\, d\mathbb{P}$$

Various Properties

Linearity

$$\mathbb{E}(aX + Y \mid \mathcal{F}) = a\,\mathbb{E}(X \mid \mathcal{F}) + \mathbb{E}(Y \mid \mathcal{F})$$

Idea: RHS clearly $\mathcal{F}$ meas so (i) holds

$$\int_A \left(a\mathbb{E}(X \mid \mathcal{F}) + \mathbb{E}(Y \mid \mathcal{F})\right) d\mathbb{P}$$

$$= a \int_A \mathbb{E}(X \mid \mathcal{F})\, d\mathbb{P} + \int_A \mathbb{E}(Y \mid \mathcal{F})\, d\mathbb{P}$$

$$= a\int_A X\,d\mathbb{P} + \int_A Y\,d\mathbb{P}$$

$$= \int_A (aX+Y)\,d\mathbb{P}$$

Expectation

$$\mathbb{E}(\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(X)$$

Idea:

$$\int_{\substack{\Omega\\ \in\mathcal{G}}} \mathbb{E}(X|\mathcal{G})\,d\mathbb{P} = \int_\Omega X\,d\mathbb{P} \quad \checkmark$$

Monotonicity

If $X \le Y$ a.s. then $\mathbb{E}(X|\mathcal{G}) \le \mathbb{E}(Y|\mathcal{G})$

Idea: For $A \in \mathcal{G}$

$$\int_A \mathbb{E}(X|\mathcal{G})\,d\mathbb{P} = \int_A X\,d\mathbb{P} \le \int_A Y\,d\mathbb{P} = \int_A \mathbb{E}(Y|\mathcal{G})\,d\mathbb{P}$$

In particular for $A = \{\mathbb{E}(X|\mathcal{G}) - \mathbb{E}(Y|\mathcal{G}) \ge \varepsilon\} \in \mathcal{G}$ where $\varepsilon > 0$ would get

$$0 \ge \int_A \big(\mathbb{E}(X|\mathcal{G}) - \mathbb{E}(Y|\mathcal{G})\big)\,d\mathbb{P} \ge \varepsilon\,\mathbb{P}(A) \ge 0$$

so $\mathbb{P}(A)$ for all $\varepsilon > 0$. ✓

MCT

If $X_n \geq 0$ and $X_n \uparrow X$ then $\mathbb{E}(X_n | \mathcal{G}) \uparrow \mathbb{E}(X|\mathcal{G})$ a.s.

Idea:

Set $Y_n = \mathbb{E}(X_n | \mathcal{G})$. Then $0 \leq Y_n$ and $Y_1 \leq Y_2 \leq \ldots$

so can define $Y = \lim_{n\to\infty} Y_n = \lim_{n\to\infty} \mathbb{E}(X_n | \mathcal{G})$ and

$$\int_A X \, d\mathbb{P} \overset{MCT}{=} \lim_{n\to\infty} \int_A X_n \, d\mathbb{P} = \lim_{n\to\infty} \int_A \underbrace{\mathbb{E}(X_n | \mathcal{G})}_{Y_n} \, d\mathbb{P}$$

$$\overset{MCT}{=} \int_A Y \, d\mathbb{P}$$

for all $A \in \mathcal{G}$ and conclude $Y = \mathbb{E}(X|\mathcal{G})$ a.s.

Similarly have Fatou and DCT.

Jenson's inequality

If ϕ is convex and $\mathbb{E}|X|, \mathbb{E}|\phi(X)| < \infty$ then

$$\phi(\mathbb{E}(X|\mathcal{G})) \leq \mathbb{E}(\phi(X)|\mathcal{G})$$

Consequences

$$|\mathbb{E}(X|\mathcal{G})| \leq \mathbb{E}(|X| \,|\, \mathcal{G})$$

$p \geq 1$: $|\mathbb{E}(X|\mathcal{G})|^p \leq \mathbb{E}(|X|^p | \mathcal{G})$ $\quad$ [$\Rightarrow$ conditional expectation is a contraction in L^p]

$\mathbb{E}(|\mathbb{E}(X|\mathcal{G})|^p) \leq \mathbb{E}(\mathbb{E}(|X|^p|\mathcal{G})) = \mathbb{E}|X|^p$

"Smallest σ-algebra wins"

Say $\mathcal{H} \subseteq \mathcal{G}$ σ-algebras.

(i) If $\mathbb{E}(X \mid \mathcal{G}) \in \mathcal{H}$ then $\mathbb{E}(X \mid \mathcal{G}) = \mathbb{E}(X \mid \mathcal{H})$

(ii) $\mathbb{E}(\mathbb{E}(X \mid \mathcal{H}) \mid \mathcal{G}) = \mathbb{E}(X \mid \mathcal{H})$

(iii) $\mathbb{E}(\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H}) = \mathbb{E}(X \mid \mathcal{H})$

Ideas:

(i) By assumption $\mathbb{E}(X \mid \mathcal{G})$ ~~$\in$~~ is $\mathcal{H}$ meas and for any $A \in \mathcal{H} \subseteq \mathcal{G}$ have

$$\int_A X \, d\mathbb{P} = \int \mathbb{E}(X \mid \mathcal{G}) \, d\mathbb{P}.$$

(ii) $\mathbb{E}(X \mid \mathcal{H})$ ~~$\in$~~ is $\mathcal{H} \subseteq \mathcal{G}$ meas.

(iii) $\mathbb{E}(X \mid \mathcal{H})$ is $\mathcal{H}$ meas. and if $A \in \mathcal{H} \subseteq \mathcal{G}$ then

$$\int_A \mathbb{E}(X \mid \mathcal{H}) \, d\mathbb{P} = \int_A X \, d\mathbb{P} = \int_A \mathbb{E}(X \mid \mathcal{G}) \, d\mathbb{P}$$

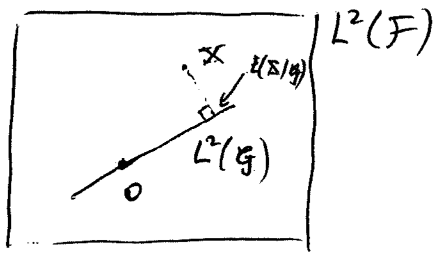
<u>Thm</u>: [Taking out what is known]

If X is $\mathcal{G}$ meas. and $\mathbb{E}|Y|, \mathbb{E}|XY| < \infty$ then

$$\mathbb{E}(XY \mid \mathcal{G}) = X \, \mathbb{E}(Y \mid \mathcal{G})$$

Thm:

Suppose $\mathbb{E}X^2 < \infty$. $\mathbb{E}(X|\mathcal{G})$ is the $\mathcal{G}$ meas. variable Y that minimizes the "mean square error" $\mathbb{E}(X-Y)^2$.



Idea: If $Z \in L^2(\mathcal{G})$ then

$$\mathbb{E}(Z\,\mathbb{E}(X|\mathcal{G})) \overset{\text{prev. Thm}}{=} \mathbb{E}(\mathbb{E}(ZX|\mathcal{G})) = \mathbb{E}(ZX)$$

so

$$\mathbb{E}(Z(X-\mathbb{E}(X|\mathcal{G}))) = 0. \quad \cancel{\text{for}} \quad (*)$$

If $Y \in L^2(\mathcal{G})$ and we set $Z := \mathbb{E}(X|\mathcal{G}) - Y \in L^2(\mathcal{G})$ then

$$\mathbb{E}(X-Y)^2 = \mathbb{E}(\underline{X-\mathbb{E}(X|\mathcal{G})} + \underline{Z})^2$$

$$\overset{(*)}{=} \mathbb{E}(X-\mathbb{E}(X|\mathcal{G}))^2 + \mathbb{E}Z^2$$

so $\mathbb{E}(X-Y)^2$ minimized when $Z=0$. ∎

Markov Processes

Consider a stochastic process $\underbrace{X_0, X_1, X_2, \dots}_{\text{r.v. on } (\Omega, \mathcal{F}, \mathbb{P})}$

Notation

$$\mathcal{F}_n^m = \sigma(X_n, \dots, X_m) \text{ for } n \le m.$$

$$\mathcal{F}_n = \sigma(X_n) = \mathcal{F}_n^n$$

Def: A process $X_0, X_1, X_2, \dots$ is Markov if for every $n > 0$ and every measurable bounded function $f: \Omega \to \mathbb{R}$ one has

$$\mathbb{E}(f(X_n) \mid \mathcal{F}_0^{n-1}) = \mathbb{E}(f(X_n) \mid \mathcal{F}_{n-1}) \quad \text{a.s.}$$

Ex: The simple random walk on $\mathbb{Z}^d$, $S_0, S_1, \dots$ is a Markov process.

Prop: Let $X_0, X_1, \dots$ be a Markov process and let $l \le m \le n$. Then for every meas. and bounded function f

$$\mathbb{E}(f(X_n) \mid \sigma(\mathcal{F}_l, \mathcal{F}_m)) = \mathbb{E}(f(X_n) \mid \mathcal{F}_m)$$

Idea: Fix l and m and do a proof by induction on n.

If $n=m$ we have

$$\mathbb{E}(f(X_m)|\sigma(\mathcal{F}_l,\mathcal{F}_m)) = f(X_m)$$
$$\shortparallel$$
$$\mathbb{E}(f(X_m)|\mathcal{F}_m) = f(X_m)$$

Assume the claim holds for $n=k-1$ with $k>m$. Then

$\sigma(\mathcal{F}_l,\mathcal{F}_m) \subseteq \mathcal{F}_0^{k-1}$ and smallest σ algebra wins

$$\mathbb{E}(f(X_k)|\sigma(\mathcal{F}_l,\mathcal{F}_m)) = \mathbb{E}(\mathbb{E}(f(X_k)|\mathcal{F}_0^{k-1})|\sigma(\mathcal{F}_l,\mathcal{F}_m))$$

$$\overset{\text{Markov}}{=} \mathbb{E}(\underbrace{\mathbb{E}(f(X_k)|\mathcal{F}_{k-1})}|\sigma(\mathcal{F}_l,\mathcal{F}_m))$$

$= g(X_{k-1})$ bec. conditional expectations exist and here they are a r.v. only determined by $\mathcal{F}_{k-1} = \sigma(X_{k-1})$.

$$= \mathbb{E}(g(X_{k-1})|\sigma(\mathcal{F}_l,\mathcal{F}_m))$$

$$\overset{\text{induction assumption}}{=} \mathbb{E}(g(X_{k-1})|\mathcal{F}_m)$$

$$\overset{\text{def of } g}{=} \mathbb{E}(\mathbb{E}(f(X_k)|\mathcal{F}_{k-1})|\mathcal{F}_m)$$

$$\overset{\text{Markov}}{=} \mathbb{E}(\mathbb{E}(f(X_k)|\mathcal{F}_0^{k-1})|\mathcal{F}_m)$$

$$\overset{\mathcal{F}_m \subseteq \mathcal{F}_0^{k-1} \text{ and smallest wins}}{=} \mathbb{E}(f(X_k)|\mathcal{F}_m). \qquad \boxtimes$$

<u>Thm</u>: Given a stochastic process $X_0, X_1, \ldots$ and indices $l < m < n$ the following are equivalent

(i) $\mathbb{E}(f(X_n) \mid \sigma(\mathcal{F}_l, \mathcal{F}_m)) = \mathbb{E}(f(X_n) \mid \mathcal{F}_m)$

(ii) $\mathbb{E}(g(X_l) \mid \sigma(\mathcal{F}_m, \mathcal{F}_n)) = \mathbb{E}(g(X_l) \mid \mathcal{F}_m)$

(iii) $\mathbb{E}(f(X_n) g(X_l) \mid \mathcal{F}_m) = \mathbb{E}(f(X_n) \mid \mathcal{F}_m)\, \mathbb{E}(g(X_l) \mid \mathcal{F}_m)$

<u>Ideas</u>: Show (i) <=> (iii). ~~(ii) <=> (iii) follows by symmetry.~~ (i)

(ii) ~~(i)~~ $\Rightarrow$ (iii):

$$\mathbb{E}(f(X_n) g(X_l) \mid \mathcal{F}_m)$$

$$= \mathbb{E}(\mathbb{E}(f(X_n) g(X_l) \mid \sigma(\mathcal{F}_m, \mathcal{F}_n)) \mid \mathcal{F}_m)$$

$$= \mathbb{E}(f(X_n)\, \mathbb{E}(g(X_l) \mid \sigma(\mathcal{F}_m, \mathcal{F}_n)) \mid \mathcal{F}_m)$$

$$\overset{(i)\Leftrightarrow(ii)}{=} \mathbb{E}(f(X_n)\, \underbrace{\mathbb{E}(g(X_l) \mid \mathcal{F}_m)} \mid \mathcal{F}_m)$$

$$= \mathbb{E}(g(X_l) \mid \mathcal{F}_m)\, \mathbb{E}(f(X_n) \mid \mathcal{F}_m)$$

(i) $\Rightarrow$ (iii): $\mathbb{E}(f(X_n) g(X_l) \mid \mathcal{F}_m) \overset{\text{smallest}}{=} \mathbb{E}(\mathbb{E}(f(X_n) g(X_l) \mid \sigma(\mathcal{F}_l, \mathcal{F}_m)) \mid \mathcal{F}_m)$

$$= \mathbb{E}(g(X_l)\, \underbrace{\mathbb{E}(f(X_n) \mid \sigma(\mathcal{F}_l, \mathcal{F}_n))}_{\mathbb{E}(f(X_n) \mid \mathcal{F}_m) \text{ by (i)}} \mid \mathcal{F}_m)$$

Done

(iii) $\Rightarrow$ (i): f, g, h arbitrary functions

$$\mathbb{E}\Big(\underbrace{g(X_l)\,h(X_m)}\,\mathbb{E}\big(f(X_n)\,|\,\sigma(\mathcal{F}_l, \mathcal{F}_n)\big)\Big)$$

$$= \mathbb{E}\big(g(X_l)\,h(X_m)\,f(X_n)\big)$$

$$= \mathbb{E}\big(h(X_m)\,\mathbb{E}(g(X_l)\,f(X_n)\,|\,\mathcal{F}_m)\big)$$

$$\overset{(iii)}{=} \mathbb{E}\big(h(X_m)\,\mathbb{E}(g(X_l)\,|\,\mathcal{F}_m)\,\mathbb{E}(f(X_n)\,|\,\mathcal{F}_m)\big)$$

$$= \mathbb{E}\Big(\mathbb{E}\big(h(X_m)\,g(X_l)\,\mathbb{E}(f(X_n)\,|\,\mathcal{F}_m)\,|\,\mathcal{F}_m\big)\Big)$$

$$= \mathbb{E}\big(h(X_m)\,g(X_l)\,\mathbb{E}(f(X_n)\,|\,\mathcal{F}_m)\big)$$

and conclusion follows since g and h arbitrary. ▫

<u>Note</u>: A Markov process satisfies (i), (ii) and (iii)

<u>Warning</u>: There exist processes that satisfy (i), (ii) and (iii) that are not Markov.

<u>Def</u>: A Markov process is <u>time-homogenous</u> if there exists a measurable map P from $\mathbb{R}$ into the space of probability measures on $\mathbb{R}$ s.t.

$$\mathbb{P}(X_n \in A \mid X_{n-1} = a) = (P(a))(A)$$

for all ~~meas.~~ $A \in F$, $a \in \mathbb{R}$ a.s. and every $n > 0$.

Write $(P(a))(A) = P(a, A) = P_a(A)$ and call P the <u>transition probabilities</u> for the process.

<u>Ex</u>: $X_0, X_1, \ldots$ i.i.d. and normal, $N(0,1)$.

For $\alpha, \beta \in \mathbb{R}$ define

$$S_0 = X_0 \quad \text{and} \quad S_{n+1} = \alpha S_n + \beta X_{n+1}$$

Then $S_0, S_1, \ldots$ is Markov and its transition probabilities are given by

$$P(x, dy) = dP_x(y) = \frac{1}{\sqrt{2\pi}\beta} e^{-\frac{(y-\alpha x)^2}{2\beta^2}} dy$$

[If $(\alpha,\beta) = (1,0)$ or $(0,1)$ the distribution of S_n is independent of n]

Theorem

Let $X_0, X_1, \ldots$ be a time-homogeneous Markov process with transition probabilities P.

Then

$$\mathbb{P}(X_n \in A \mid X_0 = a) = P_a^n(A) \; \left(= P^n(a, A) \right)$$

where P^n is defined recursively by

$$\underbrace{P^1 = P, \quad P_a^n(A) = \int_{\mathbb{R}} P_x(A) \, dP_a^{n-1}(x)}_{\text{Chapman-Kolmogorov equation}}$$

Proof: True for $n=1$ by def.

Assume true for $n = k \geq 1$. Then

$$\begin{aligned}
\mathbb{P}(X_{k+1} \in A \mid \mathcal{F}_0) &= \mathbb{E}(1_A(X_{k+1}) \mid \mathcal{F}_0) \\
&= \mathbb{E}(\mathbb{E}(1_A(X_{k+1}) \mid \sigma(\mathcal{F}_0, \mathcal{F}_k)) \mid \mathcal{F}_0) \\
&= \mathbb{E}(\mathbb{E}(1_A(X_{k+1}) \mid \mathcal{F}_k) \mid \mathcal{F}_0) \\
&= \mathbb{E}(P(X_k, A) \mid \mathcal{F}_0)
\end{aligned}$$

and result follows as assumed the distribution of X_k conditioned on $X_0 = a$ is $P^k(a, \cdot)$ $\square$

Recall: $T : \Omega \to \{0,1,2,\dots,\infty\}$ is a stopping time if

$$\{T \le n\} \in \mathcal{F}_0^n \ \forall n \iff \{T = n\} \in \mathcal{F}_0^n \ \forall n$$

Introduce the stopped process

$$X_{\min\{n,T\}} = \begin{cases} X_n & \text{if } n \le T \\ X_T & \text{otherwise} \end{cases}$$

Denote $\mathcal{F}_T = \mathcal{F}_T^T$ σ-algebra generated by X_T and by $\mathcal{F}_m^T$ the σ-algebra generated by $\{X_{\min(n,T)}\}_{n \ge m}$

Theorem [Strong Markov Property]
Let $X_0, X_1, X_2, \dots$ be a time-homogeneous Markov process with transition probabilities P. Let T be a stopping time. Then on $\{T < \infty\}$ the process $X_T, X_{T+1}, X_{T+2}, \dots$ is also Markov with transition probabilities P and

$$\mathbb{E}(f(X_T, X_{T+1}, \dots) \mid \mathcal{F}_0^T) = \mathbb{E}(f(X_T, X_{T+1}, \dots) \mid \mathcal{F}_T)$$

for every meas. f.

Idea: Take $A \in \mathcal{F}_0^T$. Then

$$\int_{A \cap \{T<\infty\}} \mathbb{E}\left(f(X_T, X_{T+1}, \dots) \mid \mathcal{F}_0^T\right) d\mathbb{P}$$

$$= \sum_{n=0}^{\infty} \int_{A \cap \{T=n\}} \mathbb{E}\left(f(X_n, X_{n+1}, \dots) \mid \mathcal{F}_0^n\right) d\mathbb{P}$$

Markov

$$= \sum_{n=0}^{\infty} \int_{A \cap \{T=n\}} \mathbb{E}\left(f(\underline{X}_n, \underline{X}_{n+1}, \dots) \mid \mathcal{F}_n\right) d\mathbb{P}$$

$$= \int_{A \cap \{T<\infty\}} \mathbb{E}\left(f(\underline{X}_T, \underline{X}_{T+1}, \dots) \mid \mathcal{F}_T\right) d\mathbb{P}. \qquad \boxtimes$$

A Markov Chain on a countable discrete probability space

$\Omega = \{\omega_1, \omega_2, \dots\}$ countable

$\mathcal{F} = 2^{\Omega}$ discrete σ-algebra (of all subsets of Ω)

$\mathbb{P}$.

For $X_0, X_1, \dots$ time-homogenous Markov Chain write the transition probabilities as

$$p(i,j) \quad \left(= p(\{\omega_i\}, \{\omega_j\})\right)$$

where then have

$$p(i, A) = \sum_{j \in A} p(i,j)$$

and $p(i, \cdot)$ is a probability measure on Ω. or really $\mathbb{N}$

If ~~μ_0~~ $X_0 \sim \mu$ then

$$\mathbb{P}_\mu(X_n = j) := \mathbb{P}(X_n = j \mid \overset{\mathcal{F}_0}{X_0}) = \sum_{i \in \Omega} \mu(i)\, p^n(i,j)$$

where p^n is the n-th matrix power of p

$$p^n(i,j) = \sum_{k_1, \dots, k_{n-1}} p(i,k_1)\, p(k_1,k_2) \cdots p(k_{n-1}, j)$$

Notation:

$X_0, X_1, \ldots$ Markov Chain

Ω countable

For $x \in \Omega$ let

$$T_x^0 = 0 \quad \text{and} \quad T_x^k = \inf\{n > T_x^{k-1} : X_n = x\},\ k \geq 1$$

"k-th return time"

Fact: T_x^k is a stopping time.

Notation: $\rho_{xy} := \mathbb{P}_x(T_y^1 < \infty)$ [probability of ever reaching y if start at x]

$\uparrow$ start at x

Lemma [Probability of k-th return]

$X_0, X_1, \ldots$ Markov Chain with Ω countable (and transition probability P). For all $x, y \in \Omega$

$$\mathbb{P}_x(T_y^k < \infty) = \rho_{xy}\, \rho_{yy}^{k-1}$$

Idea: True by def. for $k=1$.

Assume true for $k-1$. Then

$$\mathbb{P}_x(T_y^h < \infty) = \int \mathbb{1}_{\{T_y^h < \infty\}} \, d\mathbb{P}_x$$

$$= \int_{\{T_y^{h-1} < \infty\}} \underbrace{\mathbb{1}_{\{T_y^h < \infty\}}}_{=0 \text{ outside}} \, d\mathbb{P}_x$$

$$= \int_{\{T_y^{h-1} < \infty\}} \mathbb{E}_x\left(\mathbb{1}_{\{T_y^h < \infty\}} \,\middle|\, \mathcal{F}_{T_y^{h-1}}\right) d\mathbb{P}_x$$

Strong Markov prop

$$= \int_{\{T_y^{h-1} < \infty\}} \underbrace{\mathbb{E}_y\left(\mathbb{1}_{\{T_y < \infty\}}\right)} \, d\mathbb{P}_x$$

induction assump.

$$= \mathbb{P}_y(T_y < \infty)\, \mathbb{P}_x(T_y^{h-1} < \infty) = \rho_{yy}\, \rho_{xy}\, \rho_{yy}^{h-2}. \quad \square$$

<u>Def</u>: A state $x \in \Omega$ is <u>recurrent</u> if $\rho_{xx} = 1$.
Otherwise it is <u>transient</u>.

<u>Theorem</u>: If y is recurrent then

$$\mathbb{P}_y(X_n = y \text{ i.o.}) = 1$$

<u>Idea</u>: $\mathbb{P}_y(X_n = y \text{ i.o.}) = \mathbb{P}_y\left(\bigcap_{h \geq 1} \{T_y^h < \infty\}\right) \overset{\text{MCT}}{=} \lim_{h \to \infty} \underbrace{\mathbb{P}_y(T_y^h < \infty)}_{= \rho_{yy}^h = 1^h = 1} = 1.$ $\square$

Notation:

$$N(x) = \sum_{n\geq 1} \mathbb{1}_{\{X_n = x\}} \qquad \text{"number of visits to } x\text{"}$$

Theorem: If y is transient then for any x

$$\mathbb{E}_x(N(y)) < \infty$$

Idea:

$$\begin{aligned}
\mathbb{E}_x(N(y)) &= \sum_{k=1}^{\infty} \mathbb{P}_x(N(y) \geq k) \\
&= \sum_{k=1}^{\infty} \mathbb{P}_x(T_y^k < \infty) \\
&= \sum_{k\geq 1}^{\infty} \rho_{xy}\rho_{yy}^{k-1} = \frac{\rho_{xy}}{1-\rho_{yy}} < \infty \text{ if } \rho_{yy} < 1. \quad \square
\end{aligned}$$

Corollary:

x is recurrent

if and only if

$$\mathbb{E}_x(N(x)) = \sum_{n=1}^{\infty} p^n(x,x) = \infty$$

Fact: For $x \neq y$ in Ω the following are equivalent

(i) $\rho_{xy} > 0$

(ii) $p^n(x,y) > 0$ for some $n \geq 1$

(iii) there exist $i_0 = x, i_1, \ldots, i_{n-1}, i_n = y$ s.t.

$$p(i_{r-1}, i_r) > 0 \text{ for all } r = 1, \ldots, n$$

Notation

Define

$$T_x = \inf\{n \geq 0 : X_n = x\}$$

If start at $y \neq x$ then $T_x = T_x'$ from before.
If start at x then $T_x = 0$ while $T_x' > 0$.

If $\mathbb{P}_x(T_y < \infty) > 0$ write $x \to y$. If $x \to y$ and $y \to x$ write $x \leftrightarrow y$ and say x communicates with y.

Def: A subset $C \subseteq \Omega$ is irreducible if for all $x, y \in C$ we have $x \leftrightarrow y$. A Markov chain on Ω is irreducible if the full space Ω is irreducible.

Note:

$$x \longleftrightarrow x \quad \text{bec. } \mathbb{P}_x(\underbrace{\overset{=0}{T_x} < \infty}_{\Omega}) = 1 > 0$$

$$x \leftrightarrow y \Rightarrow y \longleftrightarrow x \quad \text{by def.}$$

$$x \longleftrightarrow y \text{ and } y \longleftrightarrow z \Rightarrow x \longleftrightarrow z$$

uses the equivalent representations of $\rho_{xy} > 0$ and $\rho_{yz} > 0$ etc.

Thus $\longleftrightarrow$ is an equivalence relation and its equivalence classes partition Ω and are called communicating classes.

Lemma [Recurrence is contagious]

If x is recurrent and $\rho_{xy} > 0$ then y is recurrent and $\rho_{yx} = 1$.

Idea: Since $\rho_{xy} > 0$ there exist $i_0 = x, i_1, \dots, i_K = y$ in Ω s.t. $p(i_{r-1}, i_r) > 0$ all $r = 1, \dots, K$ and we can take the smallest such K so that $i_1, \dots, i_{K-1} \neq x$.

Then

$$\underset{\substack{\uparrow \\ x \text{ recurrent}}}{0} = \mathbb{P}_x(T_x' = \infty) \overset{\text{Markov}}{\geq} \underbrace{\prod_{r=1}^{K} \overbrace{p(i_{r-1}, i_r)}^{>0}}_{\text{walk to } y \text{ not through } x} \cdot \underbrace{(1 - \rho_{yx})}_{\text{don't visit } x \text{ after } y} \Rightarrow \rho_{yx} = 1$$

Since $\rho_{yx}=1$ there exists $L \geq 0$ s.t. $p^L(y,x)>0$.
Thus

$$\sum_{n\geq 1} p^{L+n+K}(y,y) \overset{\text{Chapman-Kolmogorov}}{\geq} \sum_{n\geq 1} p^L(y,x)\,p^n(x,x)\,p^K(x,y)$$

$$= \underbrace{p^L(y,x)\,p^K(x,y)}_{>0} \underbrace{\sum_{n\geq 1} p^n(x,x)}_{=\infty \text{ bec. } x \text{ recurrent}} = \infty$$

so y is recurrent. ☒

<u>Thm</u>: If $X_0, X_1, \ldots$ is a Markov chain on a countable Ω and $C \subseteq \Omega$ is irreducible then either all states in C are recurrent or all are transient.

<u>Def</u>: $C \subseteq \Omega$ is <u>closed</u> if for all $x \in C$ and $y \notin C$ it holds that $\rho_{xy}=0$.

<u>Thm</u>: If $C \subseteq \Omega$ is a recurrent communicating class then C is closed

Thm: If $C \subseteq \Omega$ is finite and closed then it contains a recurrent state.

Proof: Else all $y \in C$ transient and then

$$\infty > \sum_{y \in C} \mathbb{E}_x(N(y)) = \sum_{y \in C} \sum_{n \geq 1} p^n(x,y) = \sum_{n \geq 1} \underbrace{\sum_{y \in C} p^n(x,y)}_{=1 \text{ bec. } \rho_{xy}=0 \text{ for } y \notin C} = \infty \quad \blacksquare$$

Thm [Decomposition Thm]

If $X_0, X_1, \ldots$ Markov chain on a finite set Ω then all closed communicating classes are recurrent. All other communicating classes are transient.

Def: Let $X_0, X_1, \ldots$ be a Markov chain on a countable set Ω with transition probability p. A non-zero measure μ on Ω is stationary if

$$\sum_{i \in \Omega} \mu(i) p(i,j) = \mu(j)$$

If in addition μ is a probability measure then we say μ is a stationary distribution.

Lemma:
If μ is a stationary distribution then

$$\mathbb{P}_\mu(X_n = j) = \mu(j) \quad \text{for all } n \geq 0.$$

Idea:

$$\begin{aligned}
\mathbb{P}_\mu(X_n = j) &= \sum_{i \in \Omega} \mu(i) p^n(i,j) \\
&= \sum_{i \in \Omega} \mu(i) \sum_{k_1, \ldots, k_{n-1}} p(i,k_1) p(k_1,k_2) \ldots p(k_{n-1}, j) \\
&= \sum_{k_1, \ldots, k_{n-1}} \underbrace{\sum_{i \in \Omega} \mu(i) p(i,k_1)} p(k_1,k_2) \ldots p(k_{n-1}, j) \\
&= \sum_{k_1, \ldots, k_{n-1}} \mu(k_1) p(k_1,k_2) \ldots p(k_{n-1}, j) \overset{\text{same argument}}{=} \ldots = \mu(j)
\end{aligned}$$

$\square$

Ex: Biased simple random walk on $\mathbb{Z}$

$X_1, X_2, \ldots$ i.i.d. with $\mathbb{P}(X_1 = 1) = q > \frac{1}{2}$, $\mathbb{P}(X_1 = -1) = 1-q$.

Then $S_0 = 0$, $S_1 = X_1, \ldots, S_n = X_1 + \ldots + X_n, \ldots$ is a Markov chain on $\Omega = \mathbb{Z}$ with transition probability

$$p(i,j) = \begin{cases} q & \text{if } j = i+1 \\ 1-q & \text{if } j = i-1 \\ 0 & \text{otherwise} \end{cases}$$

Then $\mu(i) = \left(\frac{q}{1-q}\right)^i$ is a stationary measure

$$\sum_{i \in \mathbb{Z}} \left(\frac{q}{1-q}\right)^i p(i,j) = \left(\frac{q}{1-q}\right)^{j-1} \cdot q + \left(\frac{q}{1-q}\right)^{j+1} (1-q)$$

$$= \left(\frac{q}{1-q}\right)^j \left((1-q) + q\right) = \left(\frac{q}{1-q}\right)^j = \mu(j)$$

However μ is not a stationary distribution

$$\sum_{i \in \mathbb{Z}} \mu(i) = \sum_{i \in \mathbb{Z}} \Big(\underbrace{\frac{q}{1-q}}_{>1}\Big)^i = \infty$$

so can't even be normalized into one.

Ex: Consider a Markov chain on $\{0,1,2,\dots\}$ with transition probabilities

$$p(0,1)=1$$

$$p(i,i+1)=q_i \quad \text{and} \quad p(i,0)=1-q_i \ , \ i\geq 1$$

where $\sum_{i=1}^{\infty}(1-q_i) < 1$ $\left(\text{ex: } q_i = 1-\frac{1}{3^i}\right)$.

If μ were a stationary measure it would satisfy

$$\mu(0)=\sum_{j=1}^{\infty}\mu(j)(1-q_j)$$

$$\mu(i)=\mu(i-1)\,q_{i-1}=\dots=\mu(0)\prod_{j=1}^{i-1}q_j$$

and thus

$$\mu(0)=\mu(0)\sum_{j=1}^{\infty}(1-q_j)\prod_{k=1}^{j-1}q_k$$

If $0<\mu(0)<\infty$

$$\mu(0)\leq\mu(0)\sum_{j=1}^{\infty}(1-q_j) < \mu(0) \quad \text{contradiction}$$

Thus $\mu(0)=0$ and $\mu(i)=0$ for all $i\geq 1$ so no stationary measure.

Notation:

For $x, y \in \Omega$ let

$$\gamma_x(y) = \mathbb{E}_x\left(\sum_{n=0}^{T_x'-1} 1_{\{X_n = y\}}\right) = \sum_{n=0}^{\infty} \mathbb{P}_x\left(X_n = y,\ n \le T_x' - 1\right)$$

"Expected number of visits to y before first return to x"

By definition $\gamma_x(x) = 1$.

Theorem [Existence of stationary measure]
$X_0, X_1, \ldots$ Markov chain on countable Ω.
Let $x \in \Omega$ be recurrent. Then γ_x is a stationary measure. In addition
$\rho_{xy} = 0 \Rightarrow \gamma_x(y) = 0$ and $\rho_{xy} > 0 \Rightarrow 0 < \gamma_x(y) < \infty$

Idea:

$$\gamma_x(y) = \mathbb{E}_x\left(\sum_{n=0}^{T_x'-1} 1_{\{X_n = y\}}\right)$$

$T_x' < \infty$ a.s. $X_0 = X_{T_x'} = x$

$$= \mathbb{E}_x\left(\sum_{n=1}^{T_x'} 1_{\{X_n = y\}}\right)$$

$$= \mathbb{E}_x \left(\sum_{n=1}^{\infty} 1_{\{X_n = y,\ n \le T_x'\}} \right)$$

$$= \sum_{n=1}^{\infty} \mathbb{P}_x \left(X_n = y,\ n \le T_x' \right)$$

$$= \sum_{n=1}^{\infty} \sum_{z \in \Omega} \mathbb{P}_x \left(X_{n-1} = z,\ X_n = y,\ n \le T_x' \right)$$

$$= \sum_{n=1}^{\infty} \sum_{z \in \Omega} p(z, \overset{y}{\cancel{x}})\, \mathbb{P}_x \left(X_{n-1} = z,\ n \le T_x' \right)$$

$m = n-1$

$$= \sum_{z \in \Omega} p(z, \overset{y}{\cancel{x}}) \underbrace{\sum_{m=0}^{\infty} \mathbb{P}_x \left(X_m = z,\ m \le T_x' - 1 \right)}_{\gamma_x(z)}$$

Thus γ_x stationary.

If $\rho_{xy} = 0$ then $p^n(x,y) = 0$ for all $n \ge 0$ so

$$\gamma_x(y) = \sum_{n=0}^{\infty} \mathbb{P}_x \left(X_n = y,\ n \le T_x' - 1 \right) \le \sum_{n=0}^{\infty} p^n(x,y) = 0$$

If $\rho_{xy} > 0$ there exists K s.t. $p^K(x,y) > 0$. Further since recurrence is contagious $\rho_{yx} > 0$ so exists L s.t. $p^L(y,x) > 0$. Thus

$$\gamma_x(y) = \sum_{z \in \Omega} \gamma_x(z)\, p^K(z,y) \ge \gamma_x(x)\, p^K(x,y) > 0$$

and similarly $1 = \gamma_x(x) \ge \gamma_x(y)\, p^L(y,x)$ so $\gamma_x(y) < \infty$. $\square$

<u>Fact</u>: If μ_1 and μ_2 are stationary measures then so is $\nu = \alpha\mu_1 + \beta\mu_2$ for any α, β s.t. ν is non-negative and not identically zero.

<u>Theorem</u> [Uniqueness of stationary measure]

$X_0, X_1, \dots$ Markov chain on countable Ω.

If μ is a stationary measure with $\mu(x) = 1$ for some $x \in \Omega$ then $\mu \geq \gamma_x$.

If moreover the chain is irreducible and recurrent then $\mu = \gamma_x$.

<u>Idea</u>: First note $\mu(x) = 1 = \gamma_x(x)$. For $y \neq x$

$$\mu(y) = \sum_{z_0 \in \Omega} \mu(z_0)\, p(z_0, y)$$

$$\overset{\mu(x)=1}{=} p(x,y) + \sum_{z_0 \neq x} \mu(z_0)\, p(z_0, y)$$

$$= \mathbb{P}_x(X_1 = y,\ 1 \leq T_x' - 1) + \sum_{z_0 \neq x} \Big(\sum_{z_1 \in \Omega} \mu(z_1)\, p(z_1, z_0) \Big) p(z_0, y)$$

$$= \mathbb{P}_x(X_1 = y,\ 1 \leq T_x' - 1) + \underbrace{\sum_{z_0 \neq x} p(x, z_0)\, p(z_0, y)}_{\mathbb{P}_x(X_2 = y,\ 2 \leq T_x' - 1)} + \sum_{z_0, z_1 \neq x} \mu(z_1)\, p(z_1, z_0)\, p(z_0, y)$$

By induction

$$\mu(y) \geq \sum_{m=1}^{n} \mathbb{P}_x(X_n = y, n \leq T_x' - 1) \uparrow \gamma_x(y)$$

Now assume the chain is irreducible and recurrent. By the Existence Thm γ_x is stationary and thus $\nu = \mu - \gamma_x$ also stationary and $\nu(x) = \underbrace{\mu(x)}_{=1} - \underbrace{\gamma_x(x)}_{=1} = 0$.

Since the chain is irreducible then for any $y \in \Omega \setminus \{x\}$ there is $K > 0$ s.t. $p^K(y,x) > 0$. Thus

$$0 = \nu(x) = \sum_{z \in \Omega} \nu(z) p^K(z,x) \geq \nu(y) p^K(y,x) \geq 0$$

so conclude $\nu(y) = 0$ for all $y \in \Omega$ and thus $\mu = \gamma_x$. $\square$

Note:

$$\sum_{z\in\Omega} \gamma_x(z) = \sum_{z\in\Omega} \mathbb{E}_x\left(\sum_{n=0}^{T_x'-1} 1_{\{X_n=z\}}\right)$$

$$= \mathbb{E}_x\left(\sum_{n=0}^{T_x'-1} \underbrace{\sum_{z\in\Omega} 1_{\{X_n=z\}}}_{=1}\right)$$

$$= \mathbb{E}_x(T_x')$$

Thus if x is recurrent and $\mathbb{E}_x(T_x')<\infty$ then $\frac{\gamma_x}{\mathbb{E}_x(T_x')}$ is a stationary distribution

Def: A recurrent state $x\in\Omega$ is positive recurrent if $\mathbb{E}_x(T_x')<\infty$. Otherwise it is null recurrent.

$\boxed{\text{recurrent if } \mathbb{P}_x(T_x'<\infty)=1}$

Ex: Every state in the simple random walk on $\mathbb{Z}$ is null recurrent.

$\boxed{\text{Ex: } X \text{ r.v. s.t. } X=2^n \text{ with prob. } 2^{-n}. \text{ Then } \mathbb{P}(X<\infty)=1 \text{ but } \mathbb{E}(X)=\infty}$

Fact: If Ω is finite then any recurrent state has to be positive recurrent.

Theorem

$X_0, X_1, \ldots$ irreducible Markov chain on a countable Ω. Then the following are equivalent

(i) every state is positive recurrent

(ii) some state is positive recurrent

(iii) there exists a stationary distribution

Moreover, when any of the conditions above holds, the unique stationary distribution is given by

$$\pi(x) = \frac{1}{\mathbb{E}_x(T_x')}$$

Idea: (i) $\overset{\text{trivial}}{\Rightarrow}$ (ii) $\overset{\text{prev note}}{\Rightarrow}$ (iii)

(iii) $\Rightarrow$ (i):

Let π be a stationary distribution.

Since $\sum_{z \in \Omega} \pi(z) = 1$ there exists $z \in \Omega$ with $\pi(z) > 0$.

Let $y \in \Omega \setminus \{z\}$. By irreducibility there exists $K > 0$ s.t. $p^K(z,y) > 0$. Thus

$$\pi(y) = \sum_{w \in \Omega} \pi(w) p^K(w,y) \geq \pi(z) p^K(z,y) > 0$$

Conclude π strictly positive on Ω.

Fix $x \in \Omega$. Let $\mu(w) = \frac{1}{\pi(x)} \pi(w)$ for all $w \in \Omega$.

Then μ is a stationary measure and $\mu(x) = 1$.

By Uniqueness Thm $\mu \geq \gamma_x$ so

$$\mathbb{E}_x(T_x') = \sum_{w \in \Omega} \gamma_x(w) \leq \sum_{w \in \Omega} \frac{\pi(w)}{\pi(x)} \overset{\pi \text{ prob meas}}{=} \frac{1}{\pi(x)} < \infty$$

and thus x positive recurrent. Since x arbitrary (iii)$\Rightarrow$(i). $\boxtimes$

<u>Question</u>: Does $p^n(x,y) \xrightarrow{n \to \infty} \pi(y)$ if π is a stationary distribution?

<u>Ex</u>: $\Omega = \{1,2\}$ $\quad P = \begin{bmatrix} p(1,1) & p(1,2) \\ p(2,1) & p(2,2) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

Then $P^m = \begin{cases} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} & \text{if } m \text{ odd} \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \text{if } m \text{ even} \end{cases}$

Thus for example $p^m(1,1)$ does not converge. This happens despite $\pi = \begin{bmatrix} \pi(1) \\ \pi(2) \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$ being a stationary distribution.

Def: A Markov chain $X_0, X_1, \ldots$ on a countable Ω with transition probability p is aperiodic if for all $x \in \Omega$ we have

$p^n(x,x) > 0$ for all n large enough.

Lemma [Criterion for aperiodicity]
For an irreducible Markov chain $X_0, X_1, \ldots$ to be aperiodic it suffices there exists a state $x \in \Omega$ and an integer K s.t. $p^K(x,x) > 0$ and $p^{K+1}(x,x) > 0$. In particular this is immediate if $p(x,x) > 0$ for some x.

Idea: For any $n \geq K^2$ write

$$n - K^2 = mK + r \quad , \quad 0 \leq r < K$$

Then

$$\begin{aligned} p^n(x,x) &= p^{K^2 + mK + r}(x,x) \\ &= p^{r(K+1) + (K+m-r)K}(x,x) \\ &\geq \left(p^{K+1}(x,x)\right)^r \left(p^K(x,x)\right)^{K+m-r} > 0. \end{aligned}$$

Finish using irreducibility

<u>Theorem</u> [Convergence to stationary distribution]

Let $X_0, X_1, \dots$ be a Markov chain on a countable Ω with transition probability p. Assume it is irreducible, aperiodic and has stationary distribution π. Then for all $x, z \in \Omega$

$$p^n(x,z) \xrightarrow[n\to\infty]{} \pi(z)$$

<u>Note</u>: The proof uses a technique called <u>coupling</u>.

<u>Theorem</u> [Law of large numbers for a Markov chain]

Let $X_0, X_1, \dots$ be a Markov chain on a countable Ω. Assume it is irreducible and has stationary distribution π. Let $f: \Omega \to \mathbb{R}$ be a function s.t. $\sum_{z\in\Omega} |f(z)|\pi(z) < \infty$. Then for any initial distribution

$$\frac{1}{n}\sum_{m=1}^{n} f(X_n) \longrightarrow \sum_{z\in\Omega} f(z)\pi(z)$$

Brownian Motion / The Wiener process

Recall $X \sim N(0,1)$ is a standard normal/Gaussian random variable.

$$\text{density} \quad \rho(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$\text{ch.f.} \quad \phi(t) = e^{-t^2/2}$$

Then $\sigma X + \mu \sim N(\mu, \sigma^2)$.

Consider a random vector

$$\underline{X} = (X_1, \ldots, X_d)$$

Its mean is

$$\mathbb{E}\underline{X} = (\mathbb{E}X_1, \ldots, \mathbb{E}X_d)$$

What do we do about its variance?

The covariance of X is the $d \times d$ matrix $\Gamma = [\Gamma_{ij}]$ with

$$\Gamma_{ij} = \mathbb{E}\big((X_i - \underset{\mathbb{E}X_i}{\underset{\uparrow}{\mu_i}})(X_j - \underset{\mathbb{E}X_j}{\underset{\uparrow}{\mu_j}})\big) =: \mathrm{Cov}(X_i, X_j)$$

A d-dimensional <u>standard Gaussian</u> is a random vector $\underline{X} = (X_1, \dots, X_d)$ where $X_1, \dots, X_d$ are independent standard Gaussians. In particular $\underline{X}$ has mean 0 and covariance matrix I.

More generally a random vector $\underline{X} = (X_1, \dots, X_d)$ is Gaussian if there is a vector b, a $d \times r$ matrix A and an r-dimensional standard Gaussian $\underline{Y}$ such that

$$\underline{X} = A\underline{Y} + b$$

Then $\underline{X}$ has mean b and covariance matrix $\Gamma = AA^T$. The ch.f. of $\underline{X}$ is

$$\phi(\underset{\sim}{t}) = e^{i \sum_{j=1}^{d} t_j \mu_j - \frac{1}{2} \sum_{j,k=1}^{d} t_j t_k \Gamma_{jk}}$$

$t = (t_1, \dots, t_d)$

Fact: Let $X = (X_1, \ldots, X_d)$ d-dim Gaussian. Then

$$X_1, \ldots, X_d \text{ independent} \iff \underbrace{\Gamma_{ij} = 0 \text{ for all } i \neq j}_{\text{they are uncorrelated}}$$

Fact:

$X = (X_1, \ldots, X_d)$ d-dim Gaussian if and only if $a_1 X_1 + \ldots + a_d X_d$ Gaussian for any linear combination.

Def: A continuous-time stochastic process $\{X(t)\}_{t \geq 0}$ is a Gaussian process if for all $n \geq 1$ and $0 \leq t_1 < \ldots < t_n < \infty$ the random vector

$$(X(t_1), \ldots, X(t_n))$$

is n-dim Gaussian. The mean and covariance functions of X are $\mathbb{E}(X(t))$ and $\mathrm{Cov}(X(s), X(t))$.

Ex: Say $\underline{Y}, Z$ i.i.d. $N(0,1)$. Then

$$\underline{X}(t) = \underline{Y} + t Z$$

is a Gaussian process.

$$(\underline{X}(t_1), \dots, \underline{X}(t_n)) = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \\ 1 & t_n \end{bmatrix} \begin{bmatrix} \underline{Y} \\ Z \end{bmatrix}$$ Thus Gaussian

Def: $\{\underline{X}(t)\}_{t \geq 0}$ has stationary increments if the distribution of $\underline{X}(t) - \underline{X}(s)$ depends only on $t-s$ for all $0 \leq s \leq t$. It has independent increments if $\underline{X}(t_n) - \underline{X}(t_{n-1}), \dots, \underline{X}(t_2) - \underline{X}(t_1)$ are independent whenever $0 \leq t_1 < t_2 < \dots < t_n$.

Def: $\{\underline{X}(t)\}_{t \geq 0}$ is a standard Brownian motion if $\underline{X}$ has a.s. continuous paths and stationary independent increments such that $\underline{X}(s+t) - \underline{X}(s)$ is Gaussian with mean 0 and variance t. Def II

Def: $\{X(t)\}_{t\geq 0}$ is a standard Brownian motion if X is a Gaussian process with

$$\mathbb{P}(X(t) \text{ is continuous in } t) = 1$$

$$X(0) = 0$$

$$\mathbb{E}(X(t)) = 0$$

$$\mathrm{Cov}(X(s), X(t)) = \min\{s,t\} \qquad \text{Def I}$$

More generally $B = \sigma X + x$ is a Brownian motion started at x.

Fact: The two definitions are equivalent.

Note: The Kolmogorov Extension Thm guarantees existence of a unique distribution that satisfies all the conditions of Def I except possibly the continuity statement.
Major problem is that the event
$\{X(t) \text{ is continuous in } t\}$
is not in the σ-algebra given by the Kolmogorov extension!

Theorem: Standard Brownian motion exists.

Some ideas:

Construct $\{B(t)\}_{t\geq 0}$ on $[0,1]$

Set $D_n = \{\frac{k}{2^n}, 0 \leq k \leq 2^n\}$ and $\underbrace{D = \bigcup_{n=0}^{\infty} D_n}_{\text{dyadic numbers}}$

Let $\{Z_t\}_{t\in D}$ i.i.d. standard Gaussians.

Inductively define

$$B(0) = 0 \quad , \quad B(1) = Z_1$$

and for $t \in D_n \setminus D_{n-1}$

$$B(t) = \frac{B(t-2^{-n}) + B(t+2^{-n})}{2} + \frac{Z_t}{2^{(n+1)/2}}$$

Quickly see $B(d)$ is a zero mean Gaussian, independent of $\{Z_t\}_{t\in D\setminus D_n}$ and $B(d) - B(d-2^{-n})$ are independent Gaussians with variance 2^{-n}.

Finally interpolate linearly between dyadic points

$$F_0(t) = \begin{cases} Z_1 & t=1 \\ 0 & t=0 \\ \text{linearly} & \text{in between} \end{cases}, \quad F_n(t) = \begin{cases} Z_t/2^{\frac{n+1}{2}} & t\in D_n \setminus D_{n-1} \\ 0 & t \in D_{n-1} \\ \text{linearly} & \text{in between} \end{cases}$$

Then for $t \in D_n$

$$B(t) = \sum_{i=0}^{n} F_i(t) = \sum_{i=0}^{\infty} F_i(t)$$

Last step is observing that this uniformly convergent for $t \in [0,1]$ and thus $B(t)$ continuous a.s. and the Gaussian and variance properties hold in the limit as D is dense in $[0,1]$.

$\boxed{???}$

Fact [Time translation]

If $B(t)$ is a standard Brownian motion then so is $X(t) = B(t+s) - B(s)$.

Fact [Scaling invariance]

If $B(t)$ is a standard Brownian motion then so is $X(t) = a^{-1} B(a^2 t)$

Fact [Time inversion]

If $B(t)$ is a standard Brownian motion then so is $X(t) = \begin{cases} 0 & t=0 \\ tB(t^{-1}) & t>0 \end{cases}$

Fact:

If $B(t)$ is a standard Brownian motion then with probability 1

$$\limsup_{t\to\infty} \frac{B(t)}{\sqrt{t}} = \infty \quad , \quad \liminf_{t\to\infty} \frac{B(t)}{\sqrt{t}} = -\infty$$

Fact:

If $T_0 = \inf\{t>0 : B_t = 0\}$ then $P_0(T_0 = 0) = 1$

Fact:

$Z = \{t : B(t) = 0\}$ has $\dim_{\mathcal{H}} Z = \frac{1}{2}$ a.s.

Fact [Non-Monotonicity]

A.s. for all $0 < a < b < \infty$ the standard Brownian motion is not monotone on $[a,b]$

Idea: $a = a_0 < a_1 < \dots < a_{n-1} < a_n = b$

Probability that $B(a_i) - B(a_{i-1}) \geq 0$ $\forall\, i = 1, \dots, n$
≤ 0

is $2 \cdot \left(\frac{1}{2}\right)^n \xrightarrow[n\to\infty]{} 0$. ☒

Fact: A.s. the standard Brownian motion satisfies

1. The set of times at which local maxima occur is dense.
2. Every local max is strict.
3. The set of local max is countable.

But Hölder cont for any $\gamma < \frac{1}{2}$

Fact: Fix $t \geq 0$. Then a.s. the standard Brownian motion is not differentiable at t, moreover

$$\limsup_{h \downarrow 0} \frac{B(t+h) - B(t)}{h} = \infty \text{ and } \liminf_{h \downarrow 0} \frac{B(t+h) - B(t)}{h} = -\infty$$

<u>Fact</u> [Markov property]

Suppose $\{B(t)\}_{t\geq 0}$ is a Brownian Motion started at x. Then $\{B(s+t)-B(s)\}_{t\geq 0}$ is a Brownian motion started at 0 and is independent of the process $\{B(t): 0\leq t\leq s\}$.

[There is a Strong Markov property too]