Probability Theory I (Fall 2024)

Primary reference:
Varadhan, Probability Theory

December 9, 2024

1. Essentials of measure theory

1.1. Definitions

Defn. Let Ω be a set and $\mathcal{F}$ a set of subsets of Ω.

- $\mathcal{F}$ is a field or algebra if $\emptyset \in \mathcal{B}$ and

$$A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F} \qquad \text{(complements)}$$
$$A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}. \qquad \text{(finite unions)}$$

- $\mathcal{F}$ is a σ-field or σ-algebra if $\mathcal{F}$ is also closed under countable unions:

$$(A_i)_{i\in\mathbb{N}},\ A_i \in \mathcal{F} \Rightarrow \bigcup_{i\in\mathbb{N}} A_i \in \mathcal{F} \qquad \text{(countable unions)}$$

- $\sigma(\mathcal{F})$ denotes the smallest σ-field containing $\mathcal{F}$.

Defn. Let $\mathcal{F}$ be a field and let $\mu : \mathcal{F} \to [0,\infty]$ be a function. Then μ is a finitely additive measure if

$$\mu(\emptyset) = 0$$
$$\mu(A) \geq 0 \quad \forall A \in \mathcal{F}$$
$$\mu(A \cup B) = \mu(A) + \mu(B) \quad \forall A, B \in \mathcal{F} \text{ disjoint}$$

and a finitely additive probability measure if also

$$\mu(\Omega) = 1$$

Defn. Let $\mathcal{F}$ be a σ-field and $\mu: \mathcal{F} \to [0,\infty]$. Then μ is a (countably additive) measure if in addition

$$\mu\Big(\bigcup_{i\in\mathbb{N}} A_i\Big) = \sum_{i\in\mathbb{N}} \mu(A_i) \quad \forall\ (A_i)_{i\in\mathbb{N}},\ A_i \in \mathcal{F} \text{ disjoint}$$

Exercise. Countable additivity is equivalent to either of:

- $\forall (A_i),\ A_{i+1} \supseteq A_i: \ \mu\big(\underbrace{\bigcup A_i}_{\lim A_i}\big) = \lim_{i\to\infty} \mu(A_i)$
- $\forall (A_i),\ A_{i+1} \subseteq A_i: \ \mu\big(\underbrace{\bigcap A_i}_{\lim A_i}\big) = \lim_{i\to\infty} \mu(A_i)$

Defn. If $\mathcal{F}$ is a σ-field on Ω then $(\Omega, \mathcal{F})$ is called a measurable space. If μ is a (probability) measure on $(\Omega, \mathcal{F})$ then $(\Omega, \mathcal{F}, \mu)$ is a measure (probability) space.

1.2. Construction of measures

<u>Caratheodory Extension Theorem</u>. Let $\mathcal{F}$ be a field and $\mu : \mathcal{F} \to [0,\infty]$ a <u>count. additive</u> measure. Then μ extends to a measure on $\sigma(\mathcal{F})$.

<u>Proof</u>. Define the <u>outer measure</u>: For $A \subseteq \Omega$,

$$\mu^*(A) = \inf\left\{ \sum_j \mu(A_j) : (A_j)_{j\in\mathbb{N}} \subset \mathcal{F}, \ \bigcup_{j\in\mathbb{N}} A_j \supseteq A \right\}$$

($= +\infty$ if the infimum does not exist).

Define $A \subseteq \Omega$ to be <u>μ^*-measurable</u> if

$$\mu^*(B) = \mu^*(B\cap A) + \mu^*(B \cap A^c) \quad \forall B \subset \Omega$$

and $\mathcal{M}$ to be the set of μ^*-measurable A.

We will show that $\mathcal{M}$ is a σ-field and that μ^* restricts to a measure on $\mathcal{M}$ extending μ. This proves the theorem.

Step 1. μ^* is count. subadditive, i.e.,

$$\mu^*(B) \le \sum_{i\in\mathbb{N}} \mu^*(B_i) \quad \text{if} \quad B \subset \bigcup_{i\in\mathbb{N}} B_i .$$

WLOG $\mu^*(B_i) < \infty$ for all $i \in \mathbb{N}$. Then given $\varepsilon > 0$ there are $(A_{ij})_{j\in\mathbb{N}} \subset \mathcal{F}$ s.t.

$$B_i \subset \bigcup_j A_{ij}, \quad \sum_j \mu(A_{ij}) \le \mu^*(B_i) + \varepsilon 2^{-i}$$

$$\Rightarrow \quad B \subset \bigcup_i \bigcup_j A_{ij}, \quad \mu^*(B) \le \sum_i \mu^*(B_i) + \varepsilon$$

Since $\varepsilon > 0$ is arbitrary, $\mu^*(B) \le \sum_i \mu^*(B_i)$ follows.

Step 2. μ^* extends μ, i.e., $\mu^*(A) = \mu(A)$ $\forall A \in \mathcal{F}$.
Indeed, for any $A \in \mathcal{F}$ and $(A_i)_{i\in\mathbb{N}} \subset \mathcal{F}$, $\bigcup_i A_i \supseteq A$,

$$\mu(A) \underset{\substack{\uparrow \\ \text{subadd.}}}{\le} \sum_i \mu(A \cap A_i) \underset{\substack{\uparrow \\ A\cap A_i \subseteq A_i}}{\le} \sum_i \mu(A_i)$$

$$\Rightarrow \mu(A) \le \mu^*(A).$$

Since trivially $\mu^*(A) \le \mu(A)$ thus $\mu(A) = \mu^*(A)$.

Step 3. $\mathcal{M}$ contains $\mathcal{F}$:

$$A \in \mathcal{F}, \ B \subseteq \Omega \Rightarrow \mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c).$$

By subadditivity enough to show $\ge$.

WLOG $\mu^*(B) < \infty$. Given $\varepsilon > 0$ there is $(A_i)_{i\in\mathbb{N}} \subset \mathcal{F}$

$$B \subset \bigcup_i A_i, \quad \sum \mu(A_i) \le \mu^*(B) + \varepsilon$$

$$\begin{aligned}
\Rightarrow \mu^*(B) + \varepsilon &\ge \sum_i \mu(A_i) \\
&= \sum_i \left(\mu(A_i \cap A) + \mu(A_i \cap A^c)\right) \\
&\ge \mu^*(B \cap A) + \mu^*(B \cap A^c)
\end{aligned}$$

Step 4. $\mathcal{M}$ is a field.

Clearly, $\emptyset \in \mathcal{M}$ and $A^c \in \mathcal{M}$ if $A \in \mathcal{M}$.

Let $A_1, A_2 \in \mathcal{M}$ and $B \subseteq \Omega$. Then

$$\mu^*(B \cap (A_1 \cap A_2)) + \underbrace{\mu^*(B \cap (A_1 \cap A_2)^c)}$$

$$\overset{A_1 \in \mathcal{M}}{=} \mu^*(B \cap \underbrace{(A_1 \cap A_2)^c \cap A_1}_{\substack{(A_1^c \cup A_2^c) \cap A_1 \\ = A_2^c \cap A_1}}) + \mu^*(B \cap \underbrace{(A_1 \cap A_2)^c \cap A_1^c}_{A_1^c})$$

$$= \underbrace{\mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap A_1 \cap A_2^c)}_{\overset{A_2 \in \mathcal{M}}{=} \mu^*(B \cap A_1)} + \mu^*(B \cap A_1^c) \overset{A_1 \in \mathcal{M}}{=} \mu^*(B)$$

Thus $A_1 \cap A_2 \in \mathcal{M}$. Hence $\mathcal{M}$ is a field.

Step 5. $\mathcal{M}$ is a σ-field and μ^* is a measure on $\mathcal{M}$.

Let $(A_i)_{i\in\mathbb{N}} \subset \mathcal{M}$ be disjoint, $A = \bigcup_i A_i$. It suffices to show

$$A \in \mathcal{M} \quad \text{and} \quad \mu^*(A) = \sum_i \mu^*(A_i)$$

For any $B \subseteq \Omega$,

$$\begin{aligned}
\mu^*(B) &\overset{A_1\in\mathcal{M}}{=} \mu^*(B\cap A_1) + \mu^*(B\cap A_1^c) \\
&\overset{A_2\in\mathcal{M}}{=} \mu^*(B\cap A_1) + \mu^*(B\cap \underbrace{A_1^c \cap A_2}_{A_2 \text{ since } A_2\cap A_1=\emptyset}) + \mu^*(B\cap A_1^c\cap A_2^c) \\
&= \cdots \\
&= \sum_{i=1}^n \mu^*(B\cap A_i) + \underbrace{\mu^*(B\cap A_1^c \cap\cdots\cap A_n^c)}_{\geq \mu^*(B\cap A^c)}
\end{aligned}$$

$$\Rightarrow \mu^*(B) \geq \sum_{i\in\mathbb{N}} \mu^*(B\cap A_i) + \mu^*(B\cap A^c) \qquad (*)$$

$$\overset{\text{subadd.}}{\geq} \mu^*(B\cap A) + \mu^*(B\cap A^c)$$

OTOH, by subadditivity, $\leq$ holds, so $=$, i.e., $A\in\mathcal{M}$.

Take $B=A$ in $(*)$ $\Rightarrow$ $\mu^*(A) = \sum_{i\in\mathbb{N}} \mu^*(A_i)$, i.e., μ^* is a measure on $\mathcal{M}$.

1.3. Uniqueness of measures

<u>Defn.</u> Let $\mathcal{A}$ be a set of subsets of Ω.

- $\mathcal{A}$ is a <u>monotone class</u> if
  $A_i \in \mathcal{A}$, $A_{i+1} \supseteq A_i$ (or $\subseteq$) $\Rightarrow \bigcup A_i \in \mathcal{A}$ (or $\bigcap$)
- $\mathcal{A}$ is a <u>λ-system</u> if it is a monotone class, $\Omega \in \mathcal{A}$,
  $A, B \in \mathcal{A}$, $A \subseteq B \Rightarrow B \setminus A \in \mathcal{A}$.
- $\mathcal{A}$ is a <u>π-system</u> (d-system) if $\emptyset \in \mathcal{A}$ and
  $A, B \in \mathcal{A} \Rightarrow A \cap B \in \mathcal{A}$

<u>Fact.</u> $\mathcal{A}$ is a σ-field
$\Leftrightarrow$ $\mathcal{A}$ is a field & a monotone class
$\Leftrightarrow$ $\mathcal{A}$ is a π-system & a λ-system.

<u>Monotone class lemma.</u> Let $\mathcal{F}$ be a field. Then the smallest monotone class containing $\mathcal{F}$ equals $\sigma(\mathcal{F})$.

<u>Dynkin's π-system lemma.</u> Let $\mathcal{A}$ be a π-system. Then any λ-system containing $\mathcal{A}$ contains $\sigma(\mathcal{A})$.

Both statements are similar.

Proof (monotone class lemma).

Every σ-field is a monotone class, so $\sigma(\mathcal{F}) \supseteq \mathcal{M}$ if $\mathcal{M}$ denotes the smallest monotone class containg $\mathcal{F}$.

To show $\mathcal{M}$ is a σ-field, it suffices to prove it is a field (previous fact), i.e.,

$$A, B \in \mathcal{M} \Rightarrow A \setminus B,\ B \setminus A,\ A \cap B \in \mathcal{M}.$$

(Note $\Omega \in \mathcal{F}$, so $\Omega \in \mathcal{M}$). Define

$$\mathcal{D}(A) = \{ B \in \mathcal{M} : A \setminus B,\ B \setminus A,\ A \cap B \in \mathcal{M} \}.$$

It suffices to prove $\mathcal{M} \subset \mathcal{D}(A)$, so it is enough to show $\mathcal{D}(A)$ is a monotone class containing $\mathcal{F}$.

Claim: $\mathcal{D}(B)$ is a monotone class.

- If $A_i \in \mathcal{D}(B)$, $A_{i+1} \supseteq A_i$, $A = \bigcup A_i$ then $A \in \mathcal{D}(B)$:

  $B \setminus A_i = B \cap A_i^c \in \mathcal{M}$ is decreasing
  $\Rightarrow B \setminus A = \bigcap_i (B \cap A_i^c) \in \mathcal{M}$

  $A_i \setminus B = A_i \cap B^c \in \mathcal{M}$ is increasing $\Rightarrow A \setminus B \in \mathcal{M}$

  $A \cap B = \bigcup_i A_i \cap B \in \mathcal{M}$

- If $A_i \in \mathcal{D}(B)$, $A_{i+1} \subseteq A_i$, $A = \bigcap A_i$ then $A \in \mathcal{D}(B)$: similar

Claim: $\mathcal{D}(A)$ contains $\mathcal{F}$, i.e. $B \in \mathcal{F} \Rightarrow B \in \mathcal{D}(A)$

$B \in \mathcal{F} \Rightarrow \mathcal{F} \subset \mathcal{D}(B)$ since $\mathcal{F}$ is a field

$\Rightarrow \mathcal{M} \subset \mathcal{D}(B)$ since $\mathcal{D}(B)$ is a monotone class $\supset \mathcal{F}$

$\Rightarrow A \in \mathcal{D}(B)$ since $A \in \mathcal{M}$

$\Leftrightarrow B \in \mathcal{D}(A)$

<u>Cor</u> (uniqueness of extension). Let $\mathcal{F}$ be a field and let μ_1 and μ_2 be two count. add. measures on $\sigma(\mathcal{F})$, $\mu_i(\Omega) < \infty$, s.t. $\mu_1(A) = \mu_2(A)$ $\forall A \in \mathcal{F}$. Then $\mu_1(A) = \mu_2(A)$ $\forall A \in \sigma(\mathcal{F})$.

<u>Proof</u>. Let

$$\mathcal{A} = \{\mu_1(A) = \mu_2(A) : A \in \sigma(\mathcal{F})\}.$$

Then $\mathcal{A}$ is a monotone class: if $A_i \uparrow A$ then

$$\mu_1(A) = \lim_{i\to\infty} \mu_1(A_i) = \lim_{i\to\infty} \mu_2(A_i) = \mu_2(A)$$

by countable additivity.

Thus $\mathcal{A}$ is a monotone class containing $\mathcal{F}$, so contains $\sigma(\mathcal{F})$ by the monotone class lemma.

<u>Rk</u>. Same holds for σ-finite measures μ_i.

1.4. The Lebesgue measure

Defn. Let X be a topological space. The Borel σ-field $\mathcal{B}(X)$ is the smallest σ-field on X that contains all open sets of X.

Exercise. If $X = \mathbb{R}$ then $\mathcal{B}(\mathbb{R})$ is generated by

$$\begin{aligned} I_{a,b} &= \{x \in \mathbb{R} : a < x \leq b\}, \quad a \in \mathbb{R} \cup \{-\infty\},\ b \in \mathbb{R} \\ I_{a,\infty} &= \{x \in \mathbb{R} : a < x < \infty\}. \end{aligned}$$

The set of finite disjoint unions of the $I_{a,b}$ together with $\emptyset$ is a field $\mathcal{F}$.

Let $F : \mathbb{R} \to [0,1]$ be nondecreasing and satisfy

$$\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to \infty} F(x) = 1.$$

Then a finitely additive measure $P : \mathcal{F} \to [0,1]$ is defined by

$$P(I_{a,b}) = F(b) - F(a).$$

Thm. P is countably additive iff F is right-continous.

Thus the probability measures on $\mathcal{B}(\mathbb{R})$ are in one to one correspondence with such F called the distribution function of P.

Proof. ($\Rightarrow$) Suppose P is countably additive. Then

$$F(x) = P((-\infty, x]) = P\Big(\bigcap_{i=1}^{\infty} (-\infty, x_i]\Big) = \lim_{i\to\infty} P((-\infty, x_i]) = F(x_i)$$

for any sequence (x_i) with $x_i \downarrow x$.

($\Leftarrow$) Suppose F is right-continuous.

Let $A_j \in \mathcal{F}$, $A_j \downarrow \emptyset$. It suffices to show $P(A_j) \downarrow 0$.

By contradiction, assume $P(A_i) \geq \delta > 0$ for all i.

Step 1. It suffices to assume $A_j \subset [-\ell, \ell]$ for some ℓ:

$$P(A_j) - P(A_j \cap [-\ell, \ell]) \leq \underbrace{1 - F(\ell)}_{\to 0} + \underbrace{F(-\ell)}_{\to 0} \quad (\ell \to \infty)$$

Thus we can assume $P(A_j) \geq \frac{\delta}{2}$ and $A_j \subseteq [-\ell, \ell]$.

Step 2. There are $B_j \in \mathcal{F}$ s.t. $\overline{B_j} \subset A_j$, $P(A_j \setminus B_j) \leq \frac{\delta}{10} 2^{-j}$.

Indeed, since $A_j \in \mathcal{F}$, there are $k_j \in \mathbb{N}$, a_{ji}, $b_{ji} \in \mathbb{R}$ s.t.

$$A_j = \bigcup_{i=1}^{k_j} I_{a_{ji}, b_{ji}}.$$

Take $B_j = \bigcup_{i=1}^{k_j} I_{a_{ji}+\varepsilon_j, b_{ji}}$ with $\varepsilon_j > 0$ sufficient small.

$$\Rightarrow P(A_j \setminus B_j) \leq \sum_{i=1}^{k_j} \big(F(a_{ji}) - F(a_{ji} + \varepsilon_j)\big) \to 0 \text{ as } \varepsilon_j \to 0.$$

right-continuity

Step 3. Let $E_j = \bigcap_{i=1}^{j} B_i$ and $\bar{E}_j = \bigcap_{i=1}^{j} \bar{B}_i$.

Then $\bar{E}_j$ is decreasing, closed, bounded, and $\bar{E}_j \downarrow \emptyset$ since $A_j \supseteq \bar{E}_j$.

Since A_j is decreasing and $B_j \subset A_j$,

$$
\begin{aligned}
P(E_j) &= P\Big(\bigcap_{i=1}^{j} B_j\Big) \\
&= P\Big(\bigcap_{i=1}^{j} A_i \setminus (A_i \setminus B_i)\Big) \\
&= P\Big(\bigcap_{i=1}^{j} A_i\Big) - P\Big(\bigcup_{i=1}^{j} A_i \setminus B_i\Big) \\
&\geq P(A_j) - \sum_{i=1}^{j} P(A_i \setminus B_i) \geq \frac{\delta}{2} - \frac{\delta}{10} \geq \frac{4}{10}\delta.
\end{aligned}
$$

Since $P(\bar{E}_j) \geq P(E_j)$ thus each $\bar{E}_j \neq \emptyset$.

This contradicts $\bar{E}_j \downarrow \emptyset$: any intersection of nonempty, bounded, closed, decreasing intervals is nonempty.

Thus the assumption $P(A_i) \geq \delta$ was false and $P(A_j) \downarrow 0$ (note $P(A_j)$ is nonincreasing).

Rk. Can drop condition $\lim_{x\to\infty} F(x) = 1$. This would lead to Borel measures on $\mathbb{R}$ that are finite on all bounded Borel sets.

1.5. Integration

Defn. Let (Ω, Σ) be a measurable space. Then a measurable function or random variable is a map $f : \Omega \to \mathbb{R}$ s.t. $f^{-1}(B) \in \Sigma$ for every $B \in \mathcal{B}(\mathbb{R})$.

Fact. For any $A \in \Sigma$ the indicator function

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

is measurable.

Fact. Sums, products, limits, and compositions of measurable functions are measurable.

Simple functions. For any finite collection of disjoint sets $A_j \in \Sigma$ and $a_j \in \mathbb{R}$, the function

$$f = \sum a_j \mathbb{1}_{A_j} \qquad (*)$$

is called a simple function (and measurable).

Defn. For a simple function f and a measure μ, set

$$\int f \, d\mu = \mu(f) = \sum_j a_j \mu(A_j).$$

Rk. The repr. $(*)$ is not unique, but the int. well-defined.

Fact. For f and g simple, so are $af+bg$, $a,b\in\mathbb{R}$, and $|f|$ and $|g|$ and

$$\int (af+bg)\,d\mu = a\int f\,d\mu + b\int g\,d\mu$$

$$\left|\int f\,d\mu\right| \le \int |f|\,d\mu \le \sup|f|\,\mu(\Omega)$$

From now on, let μ be a finite measure.

Bounded functions. Let f be measurable with $|f|<M$.

Fact There are simple f_j s.t. $\sup|f_j-f|\longrightarrow 0$.

Proof. Assume $|f|<M$. Write

$$[-M,M) = \bigcup_{j=1}^{n} I_j, \quad I_j = a_j + [-\tfrac{M}{n}, \tfrac{M}{n})$$

where the a_j are s.t. $[-M,M) = \bigcup_{j=1} I_j$. Then set

$$f_n = \sum_{j=1}^{n} a_j\, 1_{A_j} \text{ where } A_j = f^{-1}(I_j).$$

Clearly, $|f_n(\omega)-f(\omega)| \le \frac{M}{n} \longrightarrow 0$.

Fact. If f_j are simple and $\sup|f_j-f|\longrightarrow 0$ then $\int f_j\,d\mu$ is a Cauchy sequence. The limit

$$\int f\,d\mu \overset{\text{def}}{=} \lim_{j\to\infty}\int f_j\,d\mu$$

is independent of the approximating sequence.

Proof. By the properties of simple integrals,

$$\left|\int f_j \, d\mu - \int f_k \, d\mu\right| \le \int |f_j - f_k| \, d\mu \le \sup |f_j - f_k| \, \mu(\Omega).$$

Since $\sup|f_j - f| \to 0$, (f_j) is Cauchy, i.e. $\sup |f_j - f_k| \to 0$. Similarly, if $\sup |f_j - f| \to 0$ and $\sup |\hat{f}_j - f| \to 0$, then $|f_j - \tilde{f}_j| \to 0$ uniformly.

Fact. The integral for bounded functions has the following properties: If f, g are bounded, $a, b \in \mathbb{R}$,

$$\int (af + bg) \, d\mu = a \int f \, d\mu + b \int g \, d\mu.$$

$$\left| \int f \, d\mu \right| \le \int |f| \, d\mu \le \sup |f| \, \mu(\Omega).$$

In fact,

$$\left|\int f \, d\mu\right| \le \mu(\{\omega : |f(\omega)| > 0\}) \sup |f|$$

Proof. Take limits of the corresponding properties for simple functions. For the last property, note

$$\tilde{\mu}(A) = \mu(A \cap B) \quad \text{where } B = \{\omega : |f(\omega)| > 0\}$$

defines a measure and

$$\int f \, d\mu = \int f \, d\hat{\mu}, \quad \int |f| \, d\mu = \int |f| \, d\tilde{\mu}, \quad \tilde{\mu}(\Omega) = \mu(B)$$

The last property thus follows from the second.

<u>Aside</u>: Convergence of measurable functions

<u>Defn.</u> Let (f_n) be a sequence of measurable fnctns.

- $f_n \to f$ uniformly if $\sup |f_n - f| \to 0$
- $f_n \to f$ pointwise or everywhere if $|f_n(\omega) - f(\omega)| \to 0$ for every $\omega \in \Omega$.

If there is a measure P on (Ω, Σ) also define

- $f_n \to f$ almost everywhere (or almost surely), written $f_n \to f$ a.e. or a.s. if there is $N \in \Sigma$ s.t.
$$P(N) = 0, \quad |f_n(\omega) - f(\omega)| \to 0 \text{ for } \omega \in N^c.$$
- $f_n \to f$ in measure (or in probability) if
$$P(\{\omega \in |f_n(\omega) - f(\omega)| \geq \varepsilon) \to 0 \quad \text{for every } \varepsilon > 0.$$

<u>Fact.</u> $f_n \to f$ uniformly $\Rightarrow$ $f_n \to f$ pointwise
$\Rightarrow$ $f_n \to f$ a.e. $\Rightarrow$ $f_n \to f$ in measure

<u>Proof.</u> Only the last implication is nontrivial. But note

$$f_n \to f \text{ a.e. } \Rightarrow \bigcap_n \bigcup_{m \geq n} \{\omega : |f_m(\omega) - f(\omega)| \geq \varepsilon\} \subset N$$
$$\Rightarrow 0 = \lim_{n \to \infty} P\Big[\bigcup_{m \geq n} |f_m(\omega) - f(\omega)| \geq \varepsilon\Big]$$
<u>count. add.</u>
$$\geq \lim_{n \to \infty} P\big[|f_n(\omega) - f(\omega)| \geq \varepsilon\big].$$

Notation. Given A a measurable set, define

$$\int_A f\,d\mu = \int 1_A f\,d\mu.$$

and note $\int f\,d\mu = \int_A f\,d\mu + \int_{A^c} f\,d\mu$.

Bounded convergence theorem. Let f_n be measurable functions with $|f_n| \le M$ for some constant M. Then

$$f_n \to f \text{ in measure} \Rightarrow \lim_{n\to\infty} \int f_n\,d\mu = \int f\,d\mu.$$

Proof. Replacing f_n by $f_n - f$ it suffices to prove that if $f_n \to 0$ in measure then $\int |f_n|\,d\mu \to 0$.

$$\int |f_n|\,d\mu \le \underbrace{\int_{|f_n|\le\varepsilon} |f_n|\,d\mu}_{\le \varepsilon} + \underbrace{\int_{|f_n|>\varepsilon} |f_n|\,d\mu}_{\le M \underbrace{\mu(\{|f_n|>\varepsilon\})}_{\to 0 \text{ as } n\to\infty}}$$

$\Rightarrow \limsup_{n\to\infty} \int |f_n|\,d\mu \le \varepsilon \quad \forall \varepsilon > 0$

$\Rightarrow \lim_{n\to\infty} \int |f_n|\,d\mu = \limsup_{n\to\infty} \int |f_n|\,d\mu = 0.$

Example. $\Omega = [0,1]$ with Lebesgue measure. Then

- $f_n(x) = x^n$, $|f_n(x)| \le 1$, $f_n \to 0$ a.e. $\Rightarrow \int f_n\,dx = \frac{1}{n+1} \to 0$.
- $f_n(x) = nx^n$, $f_n \to 0$ a.e. but $\int f_n\,dx = 1 \not\to 0$.

Nonnegative functions.

Defn. f is a nonnegative measurable function if $f : \Omega \to [0, +\infty]$ is measurable with respect to the Borel σ-field on $[0, +\infty]$.

Defn. Let f be a nonnegative measurable function. Then define

$$\int f \, d\mu = \sup \left\{ \int g \, d\mu : g \text{ bounded}, \ 0 \le g \le f \right\}.$$

Fatou's Lemma. Let f_n be nonneg. meas. functions. If $f_n \to f$ in measure, then

$$\int f \, d\mu \le \liminf_{n\to\infty} \int f_n \, d\mu .$$

In general,

$$\int \liminf_{n\to\infty} f_n \, d\mu \le \liminf_{n\to\infty} \int f_n \, d\mu .$$

Proof. Let g be bounded and $0 \le g \le f$. Then $g \wedge f_n$ is bounded and

$$g \wedge f_n \to g \wedge f = g \text{ in measure.}$$

By the BCT,

$$\int g \, d\mu = \lim_{n\to\infty} \int g \wedge f_n \, d\mu \le \liminf_{n\to\infty} \int f_n \, d\mu .$$

Thus $\int f \, d\mu \le \liminf \int f_n \, d\mu$.

For the general case, replace f_n by $\tilde{f}_n = \inf_{k \geq n} f_k$. Then $\tilde{f}_n \to \tilde{f} = \liminf f_n$ and $\tilde{f}_n \leq f_n$.

<u>Monotone Convergence Theorem.</u> Let f_n be nonneg. meas. and $f_n \uparrow f$ pointwise. Then

$$\int f_n \, d\mu \uparrow \int f \, d\mu.$$

<u>Proof</u>. Clearly $\int f_n \, d\mu \leq \int f \, d\mu$, so $\lim_{n\to\infty} \int f_n \, d\mu \leq \int f \, d\mu$. The other direction follows from Fatou:

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

<u>Fact</u>. For all nonneg. meas. f, g and all $a, b \geq 0$:

$$\int (af + bg) \, d\mu = a \int f \, d\mu + b \int g \, d\mu$$

$$\int f \, d\mu \leq \int g \, d\mu \quad \text{if } f \leq g$$

$$\int f \, d\mu = 0 \quad \text{iff} \quad \mu(\{f \geq 0\}) = 0.$$

<u>Proof</u>. Let $f_n = n \wedge f$, $g_n = n \wedge g$. Then f_n, g_n are bounded and $f_n \uparrow f$, $g_n \uparrow g$. The first claim follows from MCT and the corresponding which we already know:

$$\int (af_n + bg_n) \, d\mu = a \int f_n \, d\mu + b \int g_n \, d\mu.$$

The second claim is obvious from the definition.
The third claim follows since

$$f=0 \text{ a.e.} \Leftrightarrow f_n=0 \text{ a.e. } \forall n.$$

and the corresponding claim for bounded functions.

Arbitrary functions.

Defn. $f:\Omega\to[0,\infty]$ measurable is integrable if

$$\int f\,d\mu < \infty.$$

$f:\Omega\to\mathbb{R}$ measurable is integrable if $|f|$ is. Set

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

where $f^+ = f\vee 0$, $f^- = (-f)\vee 0$.

Fact. If f,g are integrable, then $af+bg$, $a,b\in\mathbb{R}$ are integrable and

$$\int (af+bg)\,d\mu = a\int f\,d\mu + b\int g\,d\mu$$

$$\left|\int f\,d\mu\right| \le \int |f|\,d\mu$$

Moreover, for any measurable f,

$$\int |f|\,d\mu = \int f\,d\mu = 0 \quad \text{if} \quad \mu(\{f\neq 0\})=0$$

Proof. Only the first claim is slightly nontrivial.

Note that $f+g=(f+g)_+-(f+g)_-=f_+-f_-+g_+-g_-$.

$\Rightarrow (f+g)_+ + f_- + g_- = (f+g)_- + f_+ + g_+$

$$\Rightarrow \int (f+g)_+ \, d\mu + \int f_- \, d\mu + \int g_- \, d\mu = \int (f+g)_- \, d\mu + \int f_+ \, d\mu + \int g_+ \, d\mu$$

$$\Rightarrow \int (f+g) \, d\mu = \int f \, d\mu + \int g \, d\mu.$$

Similarly $\int af \, d\mu = \int af_+ \, d\mu - \int af_- \, d\mu = a \int f \, d\mu$.

<u>Dominated Convergence Theorem</u>. Let f_n be meas., and assume $f_n \to f$ in measure. Let g be integrable and $|f_n| \leq g$ for all n. Then

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

Proof. $g+f_n$ and $g-f_n$ are nonnegative and $g \pm f_n \to g \pm f$ in measure.

By Fatou, $\liminf_{n\to\infty} \int (g \pm f_n) \, d\mu \geq \int (g \pm f) \, d\mu$

$$\Rightarrow \liminf_{n\to\infty} \pm \int f_n \, d\mu \geq \pm \int f \, d\mu$$

$$\Rightarrow \liminf_{n\to\infty} \int f_n \, d\mu \geq \int f \, d\mu \geq \limsup_{n\to\infty} \int f_n \, d\mu$$

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

1.6. Transformations

Defn. Given two measurable spaces $(\Omega_1, \Sigma_1), (\Omega_2, \Sigma_2)$, $T: \Omega_1 \to \Omega_2$ is measurable if $T^{-1}(A) \in \Sigma_1$ for every $A \in \Sigma_2$. If μ is a measure on (Ω_1, Σ_1) the induced measure on (Ω_2, Σ_2), also called pullback or image measure, $T_* \mu = \mu \circ T^{-1}$ is

$$T_*\mu(A) = \mu(T^{-1}(A)) \quad \text{for every } A \in \Sigma_2.$$

Thm. If $f: \Omega_2 \to \mathbb{R}$ is measurable, $g = f \circ T: \Omega_1 \to \mathbb{R}$ is measurable and g is integrable w.r.t. μ iff f is integrable w.r.t. $T_*\mu$ and

$$\int_{\Omega_2} f \, dT_*\mu = \int_{\Omega_1} g \, d\mu.$$

Proof. Recall the simplifying assumption that μ is fin.

1. If $f = 1_A$, $A \in \Sigma_2$, this is by definition of $T_*\mu$.
2. If f is a simple function, this follows from 1. and linearity of the integral.
3. If f is bounded, it extends by uniform limits.
4. If f is nonnegative, it follows by monotone limits.
5. For general f, decompose into f^+ and f^-.

1.7. Product measures

<u>Defn.</u> Given measurable spaces (Ω_1, Σ_1), (Ω_2, Σ_2), the product space (Ω, Σ) is given by $\Omega = \Omega_1 \times \Omega_2$ and Σ the σ-field generated by the rectangles

$$A_1 \times A_2, \qquad A_1 \in \Sigma_1, \ A_2 \in \Sigma_2.$$

<u>Fact.</u> The finite disjoint unions of the rectangles $A_1 \times A_2$ form a field $\mathcal{F}$ and so $\Sigma = \sigma(\mathcal{F})$.

Given two finite measures μ_1 and μ_2 on (Ω_1, Σ_1), (Ω_2, Σ_2), define a finitely additive measure on $\mathcal{F}$ by

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$$

<u>Fact.</u> μ is well-defined, i.e., if

$$E = \bigcup_i (A_1^i \times A_2^i) = \bigcup_j (B_1^j \times B_2^j)$$

then $\sum_i \mu_1(A_1^i)\mu_2(A_2^i) = \sum_j \mu_1(B_1^j)\mu_2(B_2^j)$.

<u>Prop.</u> μ is countably additive on $\mathcal{F}$.

In particular, μ extends to a measure on $\Sigma = \sigma(\mathcal{F})$. Also write $\mu = \mu_1 \otimes \mu_2$.

Proof. For $E \in \mathcal{F}$ and $\omega_2 \in \Omega_2$, define

$$E_{\omega_2} = \{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in E\}.$$

Then $\omega_2 \mapsto \mu_1(E_{\omega_2})$ is a simple function, so measurable, and

$$\mu(E) = \int_{\Omega_2} \mu_1(E_{\omega_2})\, d\mu_2$$

Let $E_n \in \mathcal{F}$, $E_n \downarrow \emptyset$. Need to show $\mu(E_n) \downarrow 0$.

Since $E_n \in \mathcal{F}$, $E_n \downarrow \emptyset$, $E_{n,\omega_2} \downarrow \emptyset$ for all $\omega_2 \in \Omega_2$.

Since μ_1 is countably additive,

$$\mu_1(E_{\omega_2}) \downarrow 0 \qquad \forall \omega_2 \in \Omega_2.$$

Since μ_1 is a finite measure, $\mu_1(E_{\omega_2})$ is bounded. By the BCT therefore

$$\mu(E_n) = \int \mu_1(E_{n,\omega_2})\, d\mu_2 \to 0.$$

Thus μ is countably additive.

Aside: <u>Monotone Class Theorem</u>. Let (Ω, Σ) be a measurable space and $\mathcal{F}$ a field s.t. $\Sigma = \sigma(A)$. Let $\mathcal{V} \subset \{f : \Omega \to \mathbb{R} \text{ bounded}\}$ s.t.

(i) $1 \in \mathcal{V}$, $1_A \in \mathcal{V}$ $\forall A \in \mathcal{F}$.
(ii) $f_n \in \mathcal{V}$, $f_n \geq 0$, f is bounded, $f_n \uparrow f \Rightarrow f \in \mathcal{V}$.

Then $\mathcal{V}$ contains all bounded measurable functions

<u>Proof</u>. Note that $1_A \in \mathcal{V}$ for all $A \in \Sigma$ by the monotone class lemma for sets.
Since $\mathcal{V}$ is a vector space, it contains all simple functions. Therefore, given $f \geq 0$ bounded,

$$f_n = 2^{-n} \lfloor 2^n f \rfloor \in \mathcal{V}.$$

Since $f_n \uparrow f$, hence $f \in \mathcal{V}$. For $f : \Sigma \to \mathbb{R}$ bounded, decompose as $f = f^+ - f^-$ with $f^{\pm} \geq 0$ bounded.

<u>Prop.</u> Let f be a measurable function on (Ω, Σ). Then $\omega_1 \mapsto f(\omega_1, \omega_2)$ is measurable for each $\omega_2 \in \Omega_2$.

<u>Proof</u>. Let $\mathcal{V}$ be the set of bounded f for which the claim holds. The assumptions of the MCT apply, so $\mathcal{V}$ contains all bounded measurable f. Finally, approximate f as limit $n \wedge f$ and take $n \to \infty$.

<u>Prop</u>. Let f be a bounded or nonnegative measurable function on (Ω, Σ). Then

$$f_1(\omega_1) = \int f(\omega_1, \omega_2)\, d\mu_2$$

is again a bounded resp. nonnegative measurable function on (Ω_1, Σ_1).

<u>Proof</u>. Apply MCT.

<u>Fubini's Theorem</u>. (a) Let f be nonneg. meas. on (Ω, Σ)

$$\Rightarrow \int f\, d\mu = \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2)\, d\mu_2 \right) d\mu_1$$

(b) Let f be meas. on (Ω, Σ) and μ-integrable. Let

$$A_1 = \left\{ \omega_1 \in \Omega_1 : \int |f(\omega_1, \omega_2)|\, d\mu_2 < \infty \right\}.$$

For $\omega_1 \in A_1$, define $f_1 : \Omega_1 \to \mathbb{R}$ by

$$f_1(\omega_1) = \int f(\omega_1, \omega_2)\, d\mu_2 \quad \text{if } \omega_1 \in A_1,$$
$$= 0 \quad \text{if } \omega_2 \notin A_1.$$

Then $\mu_1(A_1^c) = 0$, f_1 is μ_1-integrable, $\mu_1(f_1) = \mu(f)$.

Proof. (a) The usual pattern (variations possible):

- For $f = 1_A$, $A = A_1 \times A_2$ this is the definition of μ.
- By linearity it extends to $f = 1_A$, $A \in \mathcal{F}$.
- By the monotone class theorem (and using the monotone convergence theorem to verify its assumption), the assertion follows for $f \geq 0$ bounded meas.
- For general $f \geq 0$ meas. it follows from monotone convergence.

(b) Let f be μ-integrable. Then by (a),

$$\int |f| \, d\mu = \int_{\Omega_1} \int_{\Omega_2} |f(\omega_1, \omega_2)| \, d\mu_2 \, d\mu_1 < \infty$$

so that $\mu(A_1^c) = 0$. Let

$$f_1^{(\pm)}(\omega_1) = \int f^{\pm}(\omega_1, \omega_2) \, d\mu_2$$

$$\Rightarrow f_1 = \left(f_1^{(+)} - f_1^{(-)}\right) 1_{A_1}$$

By (a), therefore

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu = \int f_1^{(+)} \, d\mu_1 - \int f_1^{(-)} \, d\mu_1$$
$$= \int f_1 \, d\mu_1$$

as needed.

1.8. Distribution and expectation

<u>Defn</u>. Let (Ω, Σ, P) be a probability space and $X : \Omega \to \mathbb{R}$ a random variable.

$$X_* P = \alpha$$

is a probability measure on $\mathcal{B}(\mathbb{R})$ called the <u>distribution of X</u> and its distribution function

$$F(x) = \alpha((-\infty, x]) = P[X \leq x]$$

is called the <u>distribution function of X</u>.

<u>Defn</u>. The expectation and variance of X are

$$E[X] = \int X(\omega)\, dP = \int x \, d\alpha$$

$$Var[X] = E[X^2] - E[X]^2 = E[(X - E(X))^2]$$

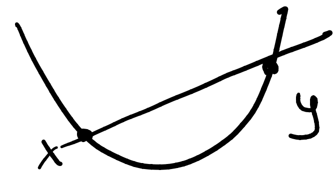
provided $\int |x| \, d\alpha < \infty$ resp. $\int x^2 \, d\alpha < \infty$. Similarly, the p-th moment of X is

$$E[X^p]$$

provided $E[|X|^p] < \infty$ if p is not even.

Aside: Let $\phi : I \subset \mathbb{R} \to \mathbb{R}$ be convex:

$$\phi(tx+(1-t)y) \le t\phi(x) + (1-t)\phi(y).$$

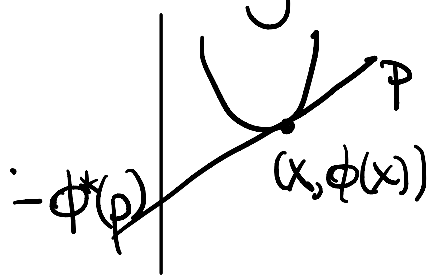


where I is an interval.

The Legendre transform of ϕ is defined by

$$\phi^*(p) = \sup_x \left(px - \phi(x) \right)$$

for $p \in I^*$ where $I^* = \{ \phi^*(p) < \infty \}$.



Then for ϕ convex:

$$\phi(x) = \sup_p \left(px - \phi^*(p) \right) = \phi^{**}(x)$$

Indeed, for all ϕ, the defn. implies $\phi(x) \ge \phi^{**}(x)$:

$$\phi^*(p) \ge px - \phi(x) \ \forall x, p \Rightarrow \phi(x) \ge \phi^{**}(x)$$

For convex ϕ, for every x, there is p s.t.

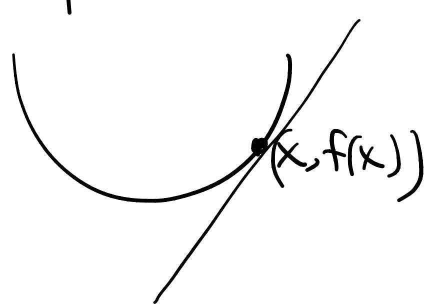
$$\phi(y) \ge \phi(x) + p(y-x) \quad \forall y$$

$$\Leftrightarrow \phi(y) - py \ge \phi(x) - px \quad \forall y$$

$$\Leftrightarrow py - \phi(y) \le px - \phi(x) \quad \forall y$$

$$\Leftrightarrow \phi^*(p) \le px - \phi(x)$$

$$\Leftrightarrow \phi(x) \le px - \phi^*(p) \le \phi^{**}(x)$$



<u>Thm</u> (Jensen's inequality). Let $\phi : \mathbb{R} \to \mathbb{R}$ be convex and f and $\phi \circ f$ be both integrable. Then if α is a prob. measure on $\mathbb{R}$:

$$\int \phi \circ f \, d\alpha \geq \phi\left(\int f \, d\alpha \right).$$

In other words, for any random variable X,

$$\mathbb{E}(\phi(X)) \geq \phi(\mathbb{E}(X)).$$

<u>Proof</u>.

$$p f(x) - \phi^*(p) \leq \phi(f(x))$$

$$\Rightarrow p \int f \, d\alpha - \phi^*(p) \leq \int \phi \circ f \, d\alpha$$

$$\Rightarrow \phi\left(\int f \, d\alpha\right) = \sup_p \left(p \int f \, d\alpha - \phi^*(p) \right) \leq \int \phi \circ f \, d\alpha.$$

<u>Cor</u>. For any $p > 1$, $\mathbb{E}(|X|) \leq \mathbb{E}(|X|^p)^{1/p}$.

<u>Proof</u>. $\phi(x) = x^p$ is convex on $[0, \infty)$.

Defn. For a vector of r.v. $X = (X_1, \dots, X_n)$, $\alpha = X_* P$ is the joint distribution of $(X_1, \dots, X_n)$. It is a probability measure on $\mathcal{B}(\mathbb{R}^n)$.

Let $\pi_i : \mathbb{R}^n \to \mathbb{R}$ be the coordinate maps, $\pi_i(x_1, \dots, x_n) = x_i$. Then $\alpha_i = (\pi_i)_* \alpha$ are called the marginals of α.

2. Weak convergence

2.1. Characteristic functions

Defn. If α is a probability measure on $\mathcal{B}(\mathbb{R})$,

$$\phi(t) = \hat{\alpha}(t) = \int e^{ixt} \, d\alpha$$

is called characteristic function (or Fourier transform) of the measure α.

Thm. ϕ is uniformly continuous and positive definite, i.e.,

$$\sum_{i,j=1}^{n} \phi(t_i - t_j)\, \xi_i \bar{\xi}_j \geq 0 \quad \forall\ \xi \in \mathbb{C}^n,\ t \in \mathbb{R}^n.$$

Proof.

$$\sum \phi(t_i - t_j)\, \xi_i \bar{\xi}_j = \int \underbrace{\sum_{i,j} e^{i(t_i - t_j)x}\, \xi_i \bar{\xi}_j}_{|\sum_i e^{it_i x} \xi_i|^2} \, d\alpha \geq 0 \quad \Rightarrow \phi \text{ is p.d.}$$

$$|\phi(t) - \phi(s)| \leq \int \underbrace{|e^{itx} - e^{isx}|}_{= |e^{i(t-s)x} - 1| \to 0} \, d\alpha \xrightarrow{t-s \to 0} 0 \text{ by BCT} \quad \Rightarrow \phi \text{ is unif. cont.}$$

How to recover the distribution from ϕ?

Assume $\alpha = f(x)\,dx$, $F(x) = \int_{-\infty}^{x} f(t)\,dt$, so $F'(x) = f(x)$.

Then ϕ is the Fourier transform of f and the inversion formula (for <u>nice</u> f) gives

$$F'(x) = \frac{1}{2\pi}\int e^{-itx}\phi(t)\,dt$$

$$\Rightarrow F(b) - F(a) = \frac{1}{2\pi}\int \phi(t) \underbrace{\int_a^b e^{-itx}\,dx}_{\frac{e^{-itb}-e^{-ita}}{-it}}\,dt$$

$$= \lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \phi(t)\,\frac{e^{-itb}-e^{-ita}}{-it}\,dt$$

<u>Thm</u>. If a, b are points of continuity for F then

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

<u>Proof</u>. Let

$$u(T,x) = \int_0^T \frac{\sin(tx)}{t}\,dt \qquad \operatorname{sign} x = \begin{cases} +1 & x>0 \\ -1 & x<0 \\ 0 & x=0 \end{cases}$$

Then $|u(T,x)| \le C$ and $\lim_{T\to\infty} u(T,x) = \operatorname{sign}(x)\,\frac{\pi}{2}$.

$$\Rightarrow \frac{1}{2\pi}\int_{-T}^{T}\left(\frac{e^{-itb}-e^{-ita}}{-it}\int e^{itx}\,d\alpha\right)dt$$

$$=\frac{1}{2\pi}\int d\alpha\int_{-T}^{T}\frac{e^{-it(b-x)}-e^{-it(a-x)}}{-it}\,dt$$

$$=\frac{1}{2\pi}\int d\alpha\int_{-T}^{T}\frac{\sin(t(x-a))-\sin(t(x-b))}{t}\,dt$$

$$=\frac{1}{\pi}\int d\alpha\left(u(T,x-a)-u(T,x-b)\right)$$

By the BCT, using $|u(T,x)|\leq C$,

$$\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}(\cdots)=\frac{1}{2}\int d\alpha\Big(\underbrace{\operatorname{sign}(x-a)-\operatorname{sign}(x-b)}_{\begin{cases}0 & \text{if } x<a\\ +1 & \text{if } x=a\\ 2 & \text{if } a<x<b\\ +1 & \text{if } x=b\\ 0 & \text{if } x>b\end{cases}}\Big)$$

$$=2\,1_{a<x\leq b}+\left(1_{x=a}-1_{x=b}\right)$$

$$=F(b)-F(a)+\underbrace{\frac{1}{2}\left(F(a)-F(a_-)-F(b)+F(b_-)\right)}_{=0 \text{ at cont. points}}$$

Prop. If F is increasing the set of points of discontinuity is countable.

If F_α and F_β are distribution functions of α and β and $F_\alpha = F_\beta$ on all continuity points, then $F_\alpha = F_\beta$, so $\alpha = \beta$.

Hence, ϕ determines the distribution uniquely.

Proof For any increasing F, the set $\{x : F(x_+) \neq F(x_-)\}$ is countable. Indeed, the intervals $(F(x_-), F(x_+))$, $x \in \mathbb{R}$ are disjoint (or empty)

$\Rightarrow 1 = F(\infty) - F(-\infty) = \sum_x \big(F(x_+) - F(x_-)\big).$

$\Rightarrow \{x : F(x_+) \neq F(x_-)\}$ is countable.

Thus for any x there are $x_i > x$ which are continuity points for F_α and F_β s.t. $x_i \downarrow x$. Thus

$$F_\alpha(x) \underset{\substack{\uparrow \\ \text{right-cont.}}}{=} \lim_{i\to\infty} F_\alpha(x_i) \underset{\substack{\uparrow \\ F_\alpha = F_\beta \text{ on} \\ \text{cont. points}}}{=} \lim_{i\to\infty} F_\beta(x_i) \underset{\substack{\uparrow \\ \text{right-cont.}}}{=} F_\beta(x)$$

Hence $F_\alpha = F_\beta$, so $\alpha = \beta$.

2.2. Weak convergence

Example. The total variation distance between two measures μ and ν on (Ω, Σ) is

$$d_{TV}(\mu,\nu) = \sup_{A\in\Sigma} |\mu(A) - \nu(A)|.$$

Note that if $\mu = \delta_x$ and $\nu = \delta_y$ on $\mathcal{B}(\mathbb{R})$ then

$$d_{TV}(\mu,\nu) = 1 \quad \text{whenever } x \neq y.$$

If Ω is a metric space such as $\mathbb{R}$ we would like a distance s.t. $d(\delta_x, \delta_y) \to 0$ as $x \to y$.

Defn. A sequence α_n of prob. distr. on $\mathbb{R}$ converges weakly to α (written $\alpha_n \Rightarrow \alpha$ or $\alpha_n \xrightarrow{w} \alpha$) if

$$\lim_{n\to\infty} \alpha_n(I) = \alpha(I) \quad \forall \text{ interval } I = [a,b]$$
$$\text{s.t. } \alpha(\{a\}) = \alpha(\{\beta\}) = 0$$

Equivalently,

$$\lim_{n\to\infty} F_n(x) = F(x) \quad \text{for every } x \text{ that is a continuity point for } F.$$

This is also written $F_n \Rightarrow F$.

__Lévy–Cramér Continuity Theorem.__ TFAE:

(i) $\alpha_n \Rightarrow \alpha$ (or $F_n \Rightarrow F$)

(ii) $\lim_{n\to\infty} \int f \, d\alpha_n = \int f \, d\alpha \quad \forall f: \mathbb{R} \to \mathbb{R}$ bounded continuous

(iii) $\lim_{n\to\infty} \hat{\alpha}_n(t) = \hat{\alpha}(t) \quad \forall t \in \mathbb{R}$

Moreover, if $\phi(t) = \lim_{n\to\infty} \hat{\alpha}_n(t)$ exists for all $t \in \mathbb{R}$ and ϕ is continuous at 0 then $\phi = \hat{\alpha}$ for some probability measure α.

__Proof__ of (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii). Let $\varepsilon > 0$.

Step 1. There are cont. points $a<b$ of F s.t.

$$F(a) \le \varepsilon, \quad 1 - F(b) \le \varepsilon$$

$$\Rightarrow F_n(a) \le 2\varepsilon, \quad 1 - F_n(b) \le 2\varepsilon \quad \text{for } n \ge n_0(\varepsilon).$$

Step 2. Using that f is uniformly continuous on $[a,b]$, for any $\delta > 0$, there is $\varepsilon > 0$ s.t. $|f(x) - f(y)| < \delta$ if $|x-y| < \varepsilon$. Let a_j be cont. pts. of F, $a_{j+1} - a_j < \varepsilon$ s.t.

$$(a,b] = \bigcup_{j=1}^{N} I_j, \quad I_j = (a_j, a_{j+1}].$$

Step 3. Let $h = \sum 1_{I_j} f(a_j)$. Then $\sup_{[a,b]} |h - f| < \delta$ and

$$\left|\int f\,d\alpha - \sum_{j=1}^{N} f(a_j)[F(a_{j+1})-F(a_j)]\right| \le \delta + 2M\varepsilon$$
$$\left|\int f\,d\alpha_n - \sum_{j=1}^{N} f(a_j)[F_n(a_{j+1})-F_n(a_j)]\right| \le \delta + 4M\varepsilon$$
where $M = \sup|f|$. Since $F_n(a_j) \to F(a_j)$ $\forall j$,
$$\limsup_{n\to\infty} \left|\int f\,d\alpha - \int f\,d\alpha_n\right| \le 2\delta + 6M\varepsilon.$$
$$\Rightarrow \lim_{n\to\infty} \left|\int f\,d\alpha - \int f\,d\alpha_n\right| = 0.$$

This proves (i) $\Rightarrow$ (ii).

(ii) $\Rightarrow$ (iii) is trivial since (iii) is (ii) with the special choice $f(x) = e^{itx}$.

<u>Proof</u> of (iii) $\Rightarrow$ (i).

Step 1. Since $F_n \in [0,1]$, for every r, there is a subsequence s.t. $F_n(r)$ converges. Extract a diag. subsequence $\Lambda \subset \mathbb{N}$ s.t.
$$F_n(r) \longrightarrow \tilde{F}(r) \quad \forall r \in \mathbb{Q},\ n \in \Lambda,\ n \to \infty.$$
Step 2. Let $F(x) = \inf_{\substack{r>x \\ r\in\mathbb{Q}}} \tilde{F}(r)$. Notice F is right-cont. and increasing.

Step 3. For every cont. point x of F,

$$\lim_{n\to\infty} F_n(x) = F(x).$$

Indeed, let $r > x$, $r \in \mathbb{Q}$. Then $F_n(x) \le F_n(r) \longrightarrow \tilde{F}(r)$.

$\Rightarrow \limsup_{n\to\infty} F_n(x) \le \tilde{F}(r)$.

$\Rightarrow \limsup_{n\to\infty} F_n(x) \le F(x)$.

To show convergence, let $y < x$ and $r \in \mathbb{Q}$, $y < r < x$.

$\Rightarrow \liminf F_n(x) \ge \liminf F_n(r) = \tilde{F}(r) \ge F(y)$.

$\Rightarrow \liminf F_n(x) \ge \inf_{y<x} F(y) = F(x-) = F(x)$

since x is a cont. point of F.

Step 4. $\phi(0)=1$ and ϕ is cont. at 0

$\Rightarrow F(-\infty)=0$, $F(+\infty)=1$.

$\Rightarrow F(x) = \alpha((-\infty, x])$ for some prob. meas. α.

Indeed, we will show

$$1 - F_n\left(\tfrac{2}{T}\right) + F_n\left(-\tfrac{2}{T}\right) \le 2\left[1 - \frac{1}{2T}\int_{-T}^{T} \phi_n(t)\, dt\right]$$

For T s.t. $\frac{2}{T}$ are cont. points for F, take limit:

$$1-F\left(\frac{2}{T}\right)+F\left(-\frac{2}{T}\right) \leq 2\left[1-\frac{1}{2T}\int_{-T}^{T}\phi(t)\,dt\right]$$

Let $T\to 0$ along cont. points,

$$1-F(0)+F(-\infty) \leq 2[1-\phi(0)] = 0.$$

Conclusion. Thus $F_n \Rightarrow F$ and $\hat{\alpha}_n = \phi$ as $n\to\infty$ along a subsequence and ϕ is ch. function of F by (i) $\Rightarrow$ (iii). This works for any subseq., hence the limit is unique.

<u>Lemma</u>. For any prob. dist. α,

$$1-F_\alpha\left(\frac{2}{T}\right)+F_\alpha\left(-\frac{2}{T}\right) \leq 2\left[1-\frac{1}{2T}\int_{-T}^{T}\hat{\alpha}(t)\,dt\right]$$

<u>Proof</u>. $$\frac{1}{2T}\int_{-T}^{T}\hat{\alpha}(t)\,dt = \int \underbrace{\left[\frac{1}{2T}\int_{-T}^{T} e^{itx}\,dt\right]}_{\frac{\sin(Tx)}{Tx}} d\alpha$$

$$\leq \int_{|x|<\ell}\left|\frac{\sin Tx}{Tx}\right| d\alpha + \int_{|x|\geq \ell}\left|\frac{\sin Tx}{Tx}\right| d\alpha$$

$$\leq \alpha(|x|<\ell) + \frac{1}{T\ell}\alpha(|x|\geq \ell)$$

$$\Rightarrow 1-\frac{1}{2T}\int_{-T}^{T}\hat{\alpha}(t)\,dt \geq \left(1-\frac{1}{T\ell}\right)\alpha(|x|\geq\ell) \underset{\ell=\frac{2}{T}}{=} \frac{1}{2}\underbrace{\alpha\left(|x|\geq\frac{2}{T}\right)}_{\text{LHS}}.$$

Defn. A sequence of r.v. X_n converges in distribution or in law to a r.v. X if the distributions α_n of X_n converge weakly to the distribution α of X.

Thm. Let $\mathcal{A}$ be a family of prob. measures on $\mathbb{R}$. Suppose $\mathcal{A}$ is uniformly tight:

$$\lim_{\ell\to\infty} \sup_{\alpha\in\mathcal{A}} \alpha(|x|\geq \ell) = 0.$$

Then for any sequence $\alpha_n \in \mathcal{A}$ there is a subseq. that converges weakly to a limiting prob. measure.

Rk. Uniform tightness is equivalent to

$$\lim_{h\to 0} \sup_{\alpha\in\mathcal{A}} \sup_{t\leq|h|} |1-\phi_\alpha(t)| = 0$$

Proof. Same as last theorem.

Rk. The above can be generalized to prob. measures on a metric space.

- $\alpha_n \Rightarrow \alpha$ then by defn if $\int f\,d\alpha_n \to \int f\,d\alpha$ $\forall f$ bd. cont.
- $\mathcal{A}$ is tight if $\forall \varepsilon>0$ $\exists K$ compact st. $\sup_{\alpha\in A} \alpha(K^c) < \varepsilon$.

Thm. Let $\alpha_n \Rightarrow \alpha$. Then

$\forall C \subset \mathbb{R}$ closed: $\limsup_{n\to\infty} \alpha_n(C) \le \alpha(C)$

$\forall G \subset \mathbb{R}$ open: $\liminf_{n\to\infty} \alpha_n(G) \ge \alpha(G)$.

Moreover, if $\alpha(\partial A) = 0$ then $\lim_{n\to\infty} \alpha_n(A) = \alpha(A)$.

Proof. Let

$$f(x) = \frac{1}{1+d(x,C)}, \quad d(x,C) = \inf_{y\in C} |x-y|.$$

Since C is closed, $f(x) = 1 \Leftrightarrow x \in C$ and $f(x) < 1$ for $x \notin C$. Therefore

$$f(x)^k \downarrow \mathbb{1}_C(x) \quad \text{as } k \to \infty.$$

Since f^k is continuous,

$$\lim_{n\to\infty} \int f(x)^k \, d\alpha_n = \int f(x)^k \, d\alpha$$

$$\Rightarrow \limsup_{n\to\infty} \alpha_n(C) \le \lim_{n\to\infty} \int f(x)^k \, d\alpha_n = \int f(x)^k \, d\alpha$$

$$\Rightarrow \limsup_{n\to\infty} \alpha_n(C) \le \alpha(C)$$

Since G is open G^c is closed and the second statement follows.

The third statement is immediate since $\partial A = \bar{A} \setminus \mathring{A}$.

2.3. Bochner's Theorem

Bochner's Theorem. If $\phi: \mathbb{R} \to \mathbb{C}$ is pos. def, cont. at 0, $\phi(0)=1$, then $\phi = \hat{\alpha}$ for a probability measure α.

Fact. Let $\phi: \mathbb{R} \to \mathbb{C}$ be pos. def. Then

(a) $\phi(t)e^{ita}$ is pos. def. for any $a \in \mathbb{R}$.

(b) $\phi(-t) = \overline{\phi(t)}$ and $|\phi(t)| \le \phi(0)$ for all t.

(c) $|\phi(t)-\phi(s)|^2 \le 4\phi(0)|1-\phi(t-s)|$ for all t, s.

Proof. (a) $\sum_{i,j} \phi(t_i - t_j) e^{i(t_i-t_j)a} \xi_i \overline{\xi_j}$

$$= \sum_{i,j} \phi(t_i - t_j) \eta_i \overline{\eta_j} \ge 0 \quad , \quad \eta_i = e^{it_i a}\xi_i$$

(b) $$\begin{bmatrix} \phi(0) & \phi(t) \\ \phi(-t) & \phi(0) \end{bmatrix} = \begin{bmatrix} \phi(0) & \phi(t) \\ \phi(-t) & \phi(0) \end{bmatrix}^* \Rightarrow \phi(t) = \overline{\phi(-t)}$$

$$\det \begin{bmatrix} \phi(0) & \phi(t) \\ \phi(-t) & \phi(0) \end{bmatrix} \ge 0 \Rightarrow \phi(0)^2 - \phi(t)\phi(-t) \ge 0$$
$$\Rightarrow |\phi(t)| \le \phi(0)$$

(c)
$$\det \begin{bmatrix} \phi(0) & \phi(t-s) & \phi(t) \\ \overline{\phi(t-s)} & \phi(0) & \phi(s) \\ \overline{\phi(t)} & \overline{\phi(s)} & \phi(0) \end{bmatrix} \ge 0$$

$$\Rightarrow \quad 0 \le 1 + \phi(t-s)\phi(s)\overline{\phi(t)} + \phi(t)\overline{\phi(t-s)}\,\overline{\phi(s)}$$
$$- |\phi(t-s)|^2 - |\phi(t)|^2 - |\phi(s)|^2$$

$$\Rightarrow \quad 0 \le 1 - |\phi(s)-\phi(t)|^2 - |\phi(t-s)|^2$$
$$- (1-\phi(t-s))\phi(s)\overline{\phi(t)} - (1-\overline{\phi(t-s)})\phi(t)\overline{\phi(s)}.$$

$$\le 1 - |\phi(s)-\phi(t)|^2 - |\phi(t-s)|^2 + 2|1-\phi(t-s)|$$

$$\Rightarrow \quad |\phi(s)-\phi(t)|^2 \le \underbrace{1-|\phi(t-s)|^2}_{(1-|\phi(t-s)|)(1+|\phi(t-s)|)} + 2|1-\phi(t-s)|$$

$$\underbrace{(1-|\phi(t-s)|)}_{\le |1-\phi(t-s)|}\underbrace{(1+|\phi(t-s)|)}_{\le 2}$$

$$\le 4|1-\phi(t-s)|.$$

<u>Fact.</u> If ϕ_α are pos. def. and ν a measure on α, then $\int \phi_\alpha \, d\nu$ also is pos. def.

<u>Proof</u> of Bochner's Theorem. Let ϕ be pos. def., continuous, and <u>in addition integrable</u>. Let

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

Step 1. $f(x) \ge 0$.

$$f(x) = \lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \left(1-\frac{|t|}{T}\right) e^{-itx}\phi(t)\,dt \qquad \text{(DCT)}$$

$$= \lim_{T\to\infty} \frac{1}{2\pi T}\int_0^T\int_0^T e^{-i(t-s)x}\phi(t-s)\,dt\,ds$$

Indeed, the change of variables $\binom{t}{s} \mapsto \binom{u}{v} = \binom{t-s}{t+s}$ has Jacobian 2 and thus

$$\int_0^T \int_0^T g(t-s)\,dt\,ds = \frac{1}{2}\int_{-T}^{T} g(u) \left(\int_0^{2T} \underbrace{1_{\frac{u+v}{2}\in[0,T]} 1_{\frac{v-u}{2}\in[0,T]}}_{\substack{\begin{cases} u+v\in[0,2T] \Rightarrow v\in[-u,2T-u] \\ v-u\in[0,2T] \Rightarrow v\in[u,2T+u] \end{cases} \\ \Rightarrow v\in[|u|,2T-|u|]}} dv \right) du$$

$$= \frac{1}{2}\int_{-T}^{T} g(u)\,(2T-2|u|)\,du$$

$$= T\int_{-T}^{T} g(u)\left(1-\frac{|u|}{T}\right)du$$

Thus

$$f(x) = \lim_{T\to\infty} \frac{1}{2\pi T}\int_0^T\int_0^T e^{-itx} e^{+ixs}\phi(t-s)\,dt\,ds$$

$$= \lim_{T\to\infty} \frac{1}{2\pi T} \lim_{N\to\infty} \left(\frac{T}{N}\right)^2 \sum_{i=1}^{N}\sum_{j=1}^{N} e^{-it_ix} e^{+it_jx}\phi(t_i-t_j) \geq 0$$

where $t_i = \frac{T}{N} i$

Riemann sum approximation

pos. def.

Step 2. $\phi(t) = \frac{1}{2\pi}\int e^{itx} f(x)\,dx$

Let $f_\sigma(x) = f(x)\, e^{-\frac{1}{2}\sigma^2 x^2}$. Then

$$\int e^{itx} f_\sigma(x)\,dx = \int e^{itx} e^{-\frac{1}{2}\sigma^2x^2} \frac{1}{2\pi}\int e^{-isx}\phi(s)\,ds\,dx$$

$$= \underline{\frac{1}{2\pi}} \int \phi(s) \underbrace{\left[\int e^{i(t-s)x} e^{-\frac{1}{2}\sigma^2x^2}\,dx\right]}_{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-s)^2}{2\sigma^2}} = p_{\sigma^2}(t-s)} ds$$

$$\overset{t=0}{\Longrightarrow} \int f_\sigma(x)\,dx = \int \phi(s)\, p_{\sigma^2}(s)\,ds \le \phi(0)\int p_{\sigma^2}(s)\,ds$$

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

By Fatou's Lemma,

$$\int f(x)\,dx = \int \underline{\lim}_{\sigma\to 0} f_\sigma(x)\,dx \le \underline{\lim}_{\sigma\to 0} \int f_\sigma(x)\,dx \le 1$$

Thus $|f_\sigma(x)\, e^{itx}| \le f(x)$ is integrable. By DCT,

$$\int e^{itx} f(x)\,dx = \lim_{\sigma\to 0}\int \phi(s)\, p_{\sigma^2}(t-s)\,ds$$

$$= \lim_{\sigma\to 0}\int \phi(s+t)\, p_{\sigma^2}(s)\,ds$$

$$= \lim_{\sigma\to 0}\int \phi(t+\sigma s)\, p_1(s)\,ds = \phi(t)$$

Step 3. Now assume ϕ is pos. def. and continuous, but not necessarily integrable. Then

$$\phi_\sigma(t) = \phi(t) e^{-\frac{\sigma^2 t^2}{2}} = \int \underbrace{\phi(t)\, e^{ity}}_{\text{pos. def.}} p_{\sigma^2}(y)\, dy$$

is also pos. def. and integrable.

By Steps 1 and 2, $\phi_\sigma = \hat{\alpha}_\sigma$ for some prob. measure α_σ

Since $\phi_\sigma(t) \to \phi(t)$ for all t, the Cont. Theorem implies that $\phi = \hat{\alpha}$ for some prob. measure α.

3. Independent sums

3.1. Kolmogorov's Consistency Theorem

Defn. A family of prob. measures P_n on $\mathbb{R}^n$ is consistent if

$$\pi_* P_{n+1} = P_n \quad \text{where } \pi(x_1, \dots, x_{n+1}) = (x_1, \dots, x_n).$$

Let $\Omega = \mathbb{R}^{\mathbb{N}} = \{(x_1, x_2, \dots) : x_i \in \mathbb{R}\}$ and $\Sigma = \sigma(\mathcal{F})$ where $\mathcal{F}$ is the field generated by cylinder sets

$$B = \{\omega : (x_1, \dots, x_n) \in A\}, \quad A \in \mathcal{B}(\mathbb{R}^n).$$

Kolmogorov's Consistency Thm. Given a consistent family of distributions P_n on $\mathbb{R}^n$ there exists a unique prob. measure P on (Ω, Σ) s.t.

$$\pi_* P = P_n \quad \text{where } \pi(x_1, \dots) = (x_1, \dots, x_n).$$

Proof. By consistency, one can define P on $\mathcal{F}$ by

$$P(B) = P_n(A).$$

To prove countable additivity, let $B_j \in \mathcal{F}$ s.t. $B_j \downarrow \emptyset$. To show $P(B_j) \downarrow 0$ assume $P(B_j) \geq \delta$ for all j, some $\delta > 0$.

$\Rightarrow B_j = \pi^{-1}(\tilde{B}_j)$ for some $\tilde{B}_j \in \mathcal{B}(\mathbb{R}^{n_j})$.

There is a compact set $\tilde{K}_j \subset \hat{B}_j$ s.t. (inner reg.).

$$P_{n_j}(\hat{B}_j \setminus \tilde{K}_j) \le \delta 2^{-j-1}.$$

Let $K_j = \pi_j^{-1}(\tilde{K}_j)$

$$D_j = \bigcap_{k=1}^{j} K_j = \pi^{-1}(\tilde{D}_j) \quad \text{for some } \tilde{D}_j \subset \tilde{K}_j \subset \mathbb{R}^{n_j}$$

Then $D_j \subset B_j$, so $D_j \downarrow \emptyset$ and $D_j \neq \emptyset$ since

$$\begin{aligned} P(D_j) &= P\Big(\bigcap_{k=1}^{j} K_k\Big) \\ &= P\Big(\bigcap_{k=1}^{j} B_k \setminus (B_k \setminus K_k)\Big) \\ &= P\Big(B_j \setminus \bigcup_{k=1}^{j} (B_k \setminus K_k)\Big) \\ &\ge \delta - \sum_{k=1}^{j} \underbrace{P(B_k \setminus K_k)}_{P_{n_k}(\hat{B}_k \setminus \tilde{K}_k) \le \delta 2^{-k-1}} \ge \frac{\delta}{2} > 0 \end{aligned}$$

Note if $\omega^{(j)} = (x_1^j, x_2^j, \ldots) \in D_j$ then $(x_1^j, \ldots, x_{n_k}^j) \in \tilde{D}_k$ for $k \le j$.

For each i there is a subsequence Λ_i s.t.

$$x_i^j \to x_i \quad (j \to \infty,\ j \in \Lambda_i).$$

Take diagonal subsequence Λ s.t.

$$x_i^j \to x_i \quad \forall i \quad (j \to \infty,\ j \in \Lambda).$$

Then $\omega = (x_1, x_2, \ldots) \in D_k$ for all k. Thus $D_k \not\downarrow \emptyset$.

The proof used the inner regularity of Borel measures.

Thm Let α be a prob. measure on $\mathbb{R}^n$. Then for any $B \in \mathcal{B}(\mathbb{R})$ there is $K \subset B$ compact s.t.

$$\alpha(B \setminus K) < \varepsilon.$$

Proof. Exercise.

3.2. Independence and convolution

<u>Defn</u>. Events $(A_\alpha)_\alpha$ are independent if for any finite subcollection $\alpha_1, \dots, \alpha_n$:

$$P\Big[\bigcap_{j=1}^{n} A_{\alpha_j}\Big] = \prod_j P[A_{\alpha_j}]$$

Random variables $(X_\alpha)_\alpha$ are independent if the events $\{X_\alpha \in A\}_{\alpha, A}$ where $A \in \mathcal{B}(\mathbb{R})$ are independent, i.e., for any finite subcollection $\alpha_1, \dots, \alpha_n$:

$$P\Big[\bigcap_j \{X_{\alpha_j} \in A_j\}\Big] = \prod_j P[X_{\alpha_j} \in A_j] \quad \forall A_1, \dots, A_n \in \mathcal{B}(\mathbb{R}).$$

<u>Prop</u>. Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be fields s.t.

$$P(A_1 \cap A_2) = P(A_1)P(A_2) \quad \forall A_i \in \mathcal{F}_i$$

Then $\sigma(\mathcal{F}_1)$ and $\sigma(\mathcal{F}_2)$ are independent.

<u>Proof</u>. Given $A_1 \in \mathcal{F}_1$, define measures

$$\mu(A) = P(A \cap A_1), \quad \nu(A) = P(A)P(A_1).$$

Then $\mu(A) = \nu(A)$ $\forall A \in \mathcal{F}_2$. By the uniqueness of the extension this then holds for all $A \in \sigma(\mathcal{F}_2)$.

Now repeat with $A \in \sigma(\mathcal{F}_2)$ given.

Fact. Random variables X and Y on (Ω, Σ, P) are independent iff the joint distribution on $\mathbb{R}^2$ is the product measure $\alpha \otimes \beta$ where α, β are the distributions of X and Y.

Fact. If X and Y are independent then the distr. of $X+Y$ is $\alpha * \beta = f_*(\alpha \otimes \beta)$, $f(x,y) = x+y$, and

$$(\alpha * \beta)(A) = \int \alpha(A-y)\, d\beta = \int \beta(A-x)\, d\alpha$$

Fact. $\widehat{\alpha * \beta}(t) = \hat{\alpha}(t)\, \hat{\beta}(t)$.

Proof. $\widehat{\alpha * \beta}(t) = \int e^{ixt}\, d(\alpha * \beta)$

$$= \int e^{i(x+y)t}\, d\alpha\, d\beta$$

$$= \left(\int e^{ixt}\, d\alpha\right)\left(e^{iyt}\, d\beta\right) = \hat{\alpha}(t)\, \hat{\beta}(t).$$

Fact. If $X_1, \dots, X_n$ are independent random variables,

$$\mathrm{Var}[X_1 + \dots + X_n] = \mathrm{Var}[X_1] + \dots + \mathrm{Var}[X_n].$$

Proof. $\mathrm{Var}[X_1 + \dots + X_n] = E[(X_1 + \dots + X_n - E[X_1] - \dots E[X_n])^2]$

$$= \sum_i E[(X_i - E[X_i])^2] + \sum_{i \neq j} \underbrace{E[(X_i - E[X_i])(X_j - E[X_j])]}_{=0 \text{ by indep.}}$$

$$= \sum_i \mathrm{Var}[X_i].$$

3.3. Weak law of large numbers

Thm. Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed (i.i.d) random variables with $\mathbb{E}[|X_i|] < \infty$ and $\mathbb{E}[X_i] = m$. Then $\frac{1}{n} S_n = \frac{1}{n}(X_1 + \cdots + X_n)$ converges to m in probability:

$$P\left[\left|\frac{S_n}{n} - m\right| \geq \delta\right] \to 0 \quad \forall \delta > 0.$$

Markov Inequality. $P[|X| > \delta] \leq \frac{E[|X|]}{\delta}$

Chebyshev Inequality. $P[|X - E[X]| > \delta] \leq \frac{Var[X]}{\delta^2}$

Proof. $1_{|X| \geq \delta} \leq \frac{1}{\delta}|X|$ and $1_{|X - E[X]| > \delta} \leq \frac{1}{\delta^2}(X - E[X])^2$

Lemma. Weak LLN holds assuming $Var[X_i] < \infty$.

Proof.
$$\begin{aligned} P\left[\left|\frac{1}{n} S_n - m\right| > \delta\right] &\leq \frac{1}{\delta^2} Var\left(\frac{S_n}{n}\right) = \frac{1}{\delta^2 n^2} \sum_i Var(X_i) \\ &= \frac{1}{\delta^2 n} Var(X_1) \\ &\to 0 \quad (n \to \infty) \end{aligned}$$

Proof 1 of weak LLN. Let
$$X_i^c = X_i 1_{|X_i| \leq c}$$
$$m^c = E[X_i^c].$$

$$\Rightarrow \frac{1}{n}\sum_i X_i = \frac{1}{n}\sum X_i^c + \frac{1}{n}\sum (X_i - X_i^c).$$

$$\Rightarrow E\left[\left|\frac{1}{n}\sum X_i - m\right|\right] = E\left[\left|\frac{1}{n}\sum X_i^c - m^c\right|\right]$$

$$+ E\left[\left|\frac{1}{n}\sum (X_i - X_i^c) - (m - m^c)\right|\right]$$

$$\leq \underbrace{\mathrm{Var}\left(\frac{1}{n}\sum X_i^c\right)}_{\leq \frac{c^2}{n} \to 0}^{1/2} + 2E[|X_i - X_i^c|]$$

$$\Rightarrow \limsup_{n\to\infty} E\left[\left|\frac{1}{n}\sum X_i - m\right|\right] \leq 2E[|X_i - X_i^c|].$$

Since $\lim_{c\to\infty} E[|X_i - X_i^c|] = 0$, therefore

$$\lim_{n\to\infty} E\left[\left|\frac{1}{n}\sum X_i - m\right|\right] = 0$$

$\Rightarrow \frac{1}{n}X_i \to m$ in probability by Markov's ineq.

<u>Proof 2 of weak LLN</u>. Let ϕ be the characteristic function of X_i. Then $\frac{1}{n}S_n$ has chacteristic fnctn.

$$\psi_n(t) = \phi\left(\frac{t}{n}\right)^n.$$

Since $\mathbb{E}[|X|]<\infty$, ϕ is differentiable at 0,

$$\phi'(0) = im, \quad m = \mathbb{E}[X_i].$$

$$\Rightarrow \phi\left(\frac{t}{n}\right) = 1 + \frac{imt}{n} + o\left(\frac{1}{n}\right).$$

$$\Rightarrow \lim_{n\to\infty} \psi_n(t) = \lim_{n\to\infty} \left(1 + \frac{imt}{n} + o\left(\frac{1}{n}\right)\right)^n = e^{imt} = \hat{\delta}_m(t).$$

$$\Rightarrow \text{Law}\left(\frac{1}{n}S_n\right) \Rightarrow \delta_m \text{ weakly}.$$

<u>Exercise.</u> Let $X_1, X_2, \ldots$ be a sequence of random variables and X a random variable (all on the same probability space).

(a) $X_i \to X$ in probability $\Rightarrow$ $X_i \to X$ in law

(b) $X_i \to 0$ in law $\Rightarrow$ $X_i \to 0$ in probability.

3.4. Central Limit Theorem

<u>Thm</u>. Let $X_1, X_2, \dots$ be i.i.d, $E[X_i]=0$, $E[X_i^2] = \sigma^2 \in (0,\infty)$.
Then the distribution of $\frac{1}{\sqrt{n}} S_n = \frac{1}{\sqrt{n}}(X_1 + X_2 + \cdots)$ converges weakly to the normal dist. with dens.

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

<u>Proof</u>. Let ϕ be the characteristic function of X_i.
Then the characteristic function ψ_n of $\frac{1}{\sqrt{n}} S_n$ is

$$\psi_n(t) = \phi(t/\sqrt{n})^n.$$

Since
$$\phi(t) = \phi(0) + \phi'(0)t + \frac{1}{2}\phi''(0)t^2 + o(t^2)$$
$$= 1 + \frac{1}{2}\sigma^2 t^2 + o(t^2),$$

$$\psi_n(t) = \left(1 + \frac{1}{2}\frac{\sigma^2 t^2}{n} + o\left(\frac{t^2}{n}\right)\right)^n = e^{n\left(\frac{1}{2}\frac{\sigma^2 t^2}{n} + o\left(\frac{t^2}{n}\right)\right)}$$

$$\Rightarrow \lim_{n\to\infty} \psi_n(t) = e^{\frac{1}{2}\sigma^2 t^2} = \hat{p}(t)$$

$$\Rightarrow \text{Law}\left(\frac{1}{\sqrt{n}} S_n\right) \Rightarrow N(0,\sigma^2).$$

3.5 Borel-Cantelli Lemma and Kolmogorov's 0-1 law

Borel-Cantelli Lemma For any events $A_1, A_2, \dots$

$$\sum_n P(A_n) < \infty \Rightarrow P\Big[\underbrace{\omega : \lim_{n\to\infty} 1_{A_n}(\omega) = 0}_{A_n \text{ happens finitely often}}\Big] = 1.$$

Moreover, if the A_i are independent, $\Leftrightarrow$.

Proof. $\sum_n P(A_n) < \infty \Rightarrow S(\omega) = \sum_n 1_{A_n}(\omega) < \infty$ a.e.

since $E[S] < \infty$.

$\Rightarrow \lim_{n\to\infty} 1_{A_n} = 0$ a.e.

Suppose the A_i are indep. and $\sum P(A_n) = \infty$.

$$\Rightarrow P\Big[\bigcup_{n\geq m} A_n\Big] = 1 - P\Big[\bigcap_{n\geq m} A_n^c\Big]$$

$$= 1 - \prod_{n\geq m} \big(\underbrace{1 - P[A_n]}_{\leq e^{-P[A_n]}}\big) \quad \text{(indep)}$$

$$\geq 1 - e^{-\sum_{n\geq m} P[A_n]} = 1$$

$\Rightarrow P[A_n \text{ happens finitely often}] = 0.$

__Defn__ Given random variables $X_1, X_2, \dots$, let

$$\sigma(X_1, X_2, \dots) = \sigma\Big(\bigcup_i X_i^{-1}(\mathcal{B}(\mathbb{R}))\Big)$$

be the smallest σ-field with respect to which the X_i are all measurable.

Define $\mathcal{F}^n = \sigma(X_n, X_{n+1}, \dots)$, $\quad \mathcal{F}_n = \sigma(X_1, \dots, X_n)$

$$\mathcal{F}^\infty = \bigcap_n \mathcal{F}^n, \qquad \mathcal{F}_\infty = \sigma\Big(\bigcup \mathcal{F}_n\Big)$$

$\mathcal{F}^\infty$ is called the __tail σ-field__.

__Kolmogorov's 0-1 law__ Suppose $X_1, X_2, \dots$ are independent random variables. Then any $A \in \mathcal{F}^\infty$ has $P(A) = 0$ or $=1$. Any $\mathcal{F}^\infty$-measurable random variable is almost surely constant.

__Proof__ For any $A \in \mathcal{F}_n$, $B \in \mathcal{F}^{n+1}$ given by

$$A = \{X_1 \in A_1, \dots, X_n \in A_n\}, \quad A_i \in \mathcal{B}(\mathbb{R})$$

$$B = \{X_{n+1} \in B_{n+1}, \dots, X_{n+1} \in B_{n+k}\}, \quad B_i \in \mathcal{B}(\mathbb{R})$$

$P(A \cap B) = P(A)P(B)$ by independence of $X_1, X_2, \dots$.

The above A form a field that generates $\mathcal{F}_n$ and the above B a field that generates $\mathcal{F}^{n+1}$. Thus by uniqueness of extension $\mathcal{F}_n$ and $\mathcal{F}^{n+1}$ are independent.

$\Rightarrow$ $\mathcal{F}_n$ is indep. of $\mathcal{F}^{\infty}$

$\Rightarrow$ $\mathcal{F}_{\infty}$ is indep. of $\mathcal{F}^{\infty}$.

Since $\mathcal{F}^{\infty} \subset \mathcal{F}_{\infty}$, for any $A \in \mathcal{F}^{\infty}$,

$$P(A) = P(A \cap A) = P(A)^2$$

$\Rightarrow$ $P(A) = 0$ or $P(A) = 1$.

3.6. Strong Law of Large Numbers

Strong Law of Large Numbers. Let $X_1, X_2, \ldots$ be i.i.d., $E[X_i]=0$. Then

$$\lim_{n\to\infty} \frac{S_n}{n} = \lim_{n\to\infty} \frac{X_1+\cdots+X_n}{n} = 0 \qquad \text{a.s.}$$

Proof. if $E[X_i^4]=M<\infty$.

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

$$= E\left[\sum_{i=1}^n X_i^4 + 6\sum_{1\le i<j\le n} X_i^2 X_j^2\right]$$

since $E[X_i X_j^3] = E[X_i X_j X_k^2] = E[X_i X_j X_k X_l] = 0$

$$\le nM + 6\binom{n}{2}M = (n+3n(n-1))M \le 3n^2 M$$

since $E[X_i^2 X_j^2] = E[X_i^2]E[X_j^2]$

$$\le E[X_i^4]^{1/2} E[X_j^4]^{1/2} = M$$

$$\Rightarrow P\Big[\underbrace{\frac{S_n}{n} > \delta}_{A_n}\Big] \le \frac{3n^2 M}{n^4 \delta^4}$$

$\Rightarrow \frac{S_n}{n} > \delta$ happens finitely often a.s.

<u>Proof</u> of general case. Let $Y_n = X_n 1_{|X_n| \le n}$.

Step 1. It suffices to prove $\frac{1}{n}\sum Y_n \to 0$.

$$\sum_{n=1}^{\infty} P[|X_n| > n] \le \int_0^{\infty} P[|X_1| > t] = E[|X_1|] < \infty$$

$$\Rightarrow \quad P[|X_n| \ne |Y_n| \text{ i.o.}] = 0$$

$$\Rightarrow \left|\sum X_n - \sum Y_n\right| < \infty \quad \text{a.s.}$$

$$\Rightarrow \left|\frac{1}{n}\sum Y_n - \frac{1}{n}\sum X_n\right| \to 0$$

Step 2. $\sum_n \frac{1}{n^2}\mathrm{Var}(Y_n) \le 4E[|X_1|] < \infty$.

$$\mathrm{Var}(Y_n) \le E[Y_n^2] = \int_0^{\infty} 2y\, P[|Y_n| > y]\, dy$$

$$\le \int_0^{n} 2y\, P[|X_1| > y]\, dy.$$

$$\Rightarrow \sum \frac{1}{n^2} E[Y_n^2] \le \sum \frac{1}{n^2}\int_0^{\infty} 1_{y<n}\, 2y\, P[|X_1| > y]\, dy$$

$$= \int_0^{\infty} \underbrace{\left(\sum_{n>y} \frac{1}{n^2}\right) 2y}_{\le 4}\, P[|X_1| > y]\, dy$$

$$\le 4E[|X_1|]$$

Indeed, if $y \geq 1$

$$\sum_{n>y} \frac{1}{n^2} \leq \int_{\lceil y \rceil}^{\infty} \frac{1}{x^2}\, dx = \frac{1}{\lceil y \rceil} \leq \frac{1}{\lfloor y \rfloor}$$

$$\Rightarrow 2y \sum_{n>y} \frac{1}{n^2} \leq 2 \frac{y}{\lfloor y \rfloor} \leq 4$$

and if $y \in (0,1)$ then also

$$2y \sum_{n>y} \frac{1}{n^2} \leq 2 \sum_{n \geq 1} \frac{1}{n^2} \leq 2\left(1 + \int_1^{\infty} \frac{1}{x^2}\right) \leq 4.$$

Step 3. $\frac{1}{n(k)} \underbrace{(Y_1 + \cdots + Y_{n(k)})}_{T_{n(k)}} \to 0$ if $n(k) = \lfloor \alpha^k \rfloor$, $\alpha > 1$

$$\sum_k P(|T_{n(k)} - E T_{n(k)}| > \varepsilon n(k)) \leq \frac{1}{\varepsilon^2} \sum_k \frac{1}{n(k)^2} \operatorname{Var}(T_{n(k)})$$

$$\leq \frac{1}{\varepsilon^2} \sum_k \frac{1}{n(k)^2} \sum_{m=1}^{n(k)} \operatorname{Var}(Y_m)$$

$$= \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \operatorname{Var}(Y_n) \underbrace{\sum_{k:\, n(k) \geq n} n(k)^{-2}}$$

$$\leq \sum_{k:\, n(k) \geq n} (2\alpha^k)^2 \leq 4 \sum_{k:\, \alpha^k \geq n} \alpha^{2k} \leq 4(1-\alpha^2)^{-1} n^{-2}$$

$$\leq \frac{4}{\varepsilon^2 (1-\alpha^2)} \sum_n n^{-2} \operatorname{Var}(Y_n) \leq \frac{16}{\varepsilon^2 (1-\alpha^2)} E|X_1|$$

By Borel-Cantelli, $\frac{1}{n(k)} T_{n(k)} - \underbrace{\frac{1}{n(k)} E[T_{n(k)}]}_{\to E[X_1] \text{ by DCT}} \to 0$ a.s.

$\Rightarrow \frac{1}{n(k)} T_{n(k)} \to 0$

Step 4. $\frac{1}{n}(Y_1 + \cdots + Y_n) \to 0$

WLOG $Y_i \geq 0$. Then

$$\frac{T_{n(k)}}{n(k+1)} \leq \frac{T_m}{m} \leq \frac{T_{n(k+1)}}{n(k)} \quad \text{if } n(k) \leq m < n(k+1)$$

Since $n(k+1)/n(k) \to \alpha$ $(k \to \infty)$ LHS and RHS converge to 0 as $k \to \infty$ a.s.

$\Rightarrow T_m/m \to 0$ a.s.

3.7. Infinitely divisible distributions

Defn. Let α be a prob. measure on $\mathbb{R}$ and

$$e_a(\alpha) = e^{-a} \sum_{j=0}^{\infty} \frac{a^j}{j!} \alpha^j \quad \text{where } \alpha^j = \underbrace{\alpha * \cdots * \alpha}_{j \text{ times}}.$$

The measure $e_a(\alpha)$ is called compound Poisson distribution associated with α.

If $\alpha = \delta_1$ then $e_a(\delta_1)$ is the std. Poisson distribution.

Fact $\widehat{e_a(\alpha)}(t) = e^{-a} \sum_{j=0}^{\infty} \frac{a^j}{j!} \hat{\alpha}(t)^j = e^{a(\hat{\alpha}(t)-1)}$

$$e_{a+b}(\alpha) = e_a(\alpha) * e_b(\alpha)$$

$$e_a(\alpha) = \underbrace{e_{\frac{a}{n}}(\alpha) * \cdots * e_{\frac{a}{n}}(\alpha)}_{n\text{-times}}$$

Defn. A prob. distr. β that can be written as

$$\beta = \underbrace{\beta_n * \cdots * \beta_n}_{n}$$

for some β_n and any n is infinitely divisible.

Exercise. If α and β are infinitely divisible then $\alpha * \beta$ is infinitely divisible.

Exercise. The normal distri. is infinitely divisible.

Consider r.v. $\{X_{n,j} : 1 \leq j \leq k_n\}$ with distr. $\alpha_{n,j}$.

Assume $\lim\limits_{n\to\infty} \sup\limits_{1\leq j\leq k_n} P[X_{n,j} > \delta] = 0 \quad \forall \delta > 0$.

Let μ_n be the distr. of $\sum_{j=1}^{k_n} X_{n,j}$.

Accompanying Laws Thm. Assume $E[X_{n,j}] = 0$ for all n,j. Then

$$\underset{\alpha_{n,1} * \cdots * \alpha_{n,k_n}}{\underset{\parallel}{\mu_n}} \Rightarrow \mu \iff \underset{\beta_{n,1} * \cdots * \beta_{n,k_n}}{\underset{\parallel}{\lambda_n}} \Rightarrow \mu$$

where $\beta_{n,j} = e_1(\alpha_{n,j})$ is the associated compound Poisson distribution.

We will not prove this Thm.

The characteristic function of λ_n is

$$\hat{\lambda}_n(t) = \prod_{j=1}^{k_n} e^{\hat{\beta}_{n,j}(t)-1}$$

$$= \exp\left[\sum_{j=1}^{k_n}\int(e^{itx}-1)\,d\beta_{n,j}\right]$$

$$= \exp\left[\int(e^{itx}-1)\,dM_n\right]$$

$$= \exp\left[\int\left(e^{itx}-1-it\theta(x)\right)dM_n + itb_n\right]$$

where $a_n = \int\theta\,dM_n$

and θ is any bd. function. Assume

$$\theta(x) = x + O(x^3).$$

<u>Defn</u>. M is an admissible Lévy measure if

$$\int\frac{x^2}{1+x^2}\,dM < \infty.$$

<u>Thm</u>. For any admissible Lévy measure M, $\sigma^2 > 0$, $a \in \mathbb{R}$, there is an infinitely divisible measure μ s.t.

$$\hat{\mu}(t) = \exp\left[\int\left(e^{itx}-1-\theta(x)\right)dM + ita - \frac{\sigma^2 t^2}{2}\right]$$

Denote $\mu = e(M, \sigma^2, a)$.

Thm. $\mu_n = e(M_n, \sigma_n^2, a_n) \Rightarrow \mu$ iff $\mu = e(M, \sigma^2, a)$ and the following conditions hold:

- $\lim_{n\to\infty} \int f \, dM_n = \int f \, dM$ $\forall f$ cont. bd. $f(x)=0$, $|x|<\varepsilon$
- $\lim_{n\to\infty} \left[\sigma_n^2 + \int_{-\ell}^{\ell} x^2 \, dM_n\right] = \left[\sigma^2 + \int_{-\ell}^{\ell} x^2 \, dM\right]$ $\forall \ell$ s.t. $M(\{-\ell\}\cup\{\ell\})=0$.
- $\lim_{n\to\infty} a_n = a$.

Cor. (Lévy-Khintchine repr.) Any infinitely divisible measure μ is of the form $\mu = e(M, \sigma^2, a)$.

4. Conditioning

4.1. Radon-Nikodym Theorem

<u>Defn</u> A <u>signed measure</u> on (Ω,Σ) is a set function $\lambda:\Sigma\to\mathbb{R}$ that is countably additive:

$$\lambda\left(\bigcup A_n\right)=\sum \lambda(A_n) \quad \text{for disjoint } A_n$$

$$\overset{\text{Exerc.}}{\Longleftrightarrow} \quad \lambda(A_n)\longrightarrow\lambda(A) \quad \text{for } A_n\uparrow A \text{ or } A_n\downarrow A.$$

<u>Example</u>. Let μ_1 and μ_2 be finite positive measures. Then $\lambda=\mu_1-\mu_2$ is a signed measure.

<u>Hahn-Jordan Decomposition</u> A signed measure λ on (Ω,Σ) can be decomposed as

$$\lambda=\mu_+-\mu_-$$

where $\mu_\pm$ are positive measures that are orthogonal, i.e., there are disjoint $\Omega_\pm\in\Sigma$ s.t. $\mu_+(\Omega_-)=\mu_-(\Omega_+)=0$.

<u>Lemma</u>. If λ is a signed measure on (Ω,Σ) then

$$\sup_{A\in\Sigma}|\lambda(A)|<\infty.$$

<u>Proof</u>. Facts:

$x=a+b \Rightarrow |x|\geq|a|-|b|$

(i) $\lambda(\Omega)\in\mathbb{R} \Rightarrow \big||\lambda(A)|-|\lambda(A^c)|\big|\leq|\lambda(\Omega)|<\infty \;\forall A\in\Sigma$

(ii) $\sup_{B\subset A} |\lambda(B)| < \infty$ & $\sup_{B\subset A^c} |\lambda(B)| < \infty$

$$\Rightarrow \sup_B |\lambda(B)| \leq \sup_B \left(|\lambda(B\cap A)| + |\lambda(B \cap A^c)| \right)$$

Suppose there is A s.t. $\sup_{B\subset A} |\lambda(B)| = \infty$. Then for any $N>0$ there is $A_1 \subset A$ s.t.

$$|\lambda(A_1)| \geq N \quad \text{and} \quad \sup_{B\subset A_1} |\lambda(B)| = \infty.$$

Indeed, we can pick $E \subset A$ s.t. $|\lambda(E)| \geq 2N$ and then by (i) $\big||\lambda(A\setminus E)| - |\lambda(E)|\big| \leq |\lambda(A)| < |\lambda(\Omega)| < \infty$ and $|\lambda(A\setminus E)| \geq 2N - |\lambda(\Omega)| \geq N$ if N is sufficiently large.

Thus both E and $A\setminus E$ satisfy the first condition. By (ii) at least one of them has to satisfy the second condition

Iterating this, there are A_j s.t. $A_{j+1} \subset A_j$ and

$$|\lambda(A_j)| \geq j \quad \text{and} \quad \sup_{B\subset A_j} |\lambda(B)| = \infty.$$

Thus A_j is decreasing but $|\lambda(A_j)| \to \infty$, a contrad. to count. additivity which gives $\lambda(A_j) \to \lambda(A) \in \mathbb{R}$ if $A_j \downarrow A$.

Lemma. Suppose $\lambda(A) > 0$. Then there is $\bar{A} \subset A$ that is *totally positive*, i.e.,

$$\inf_{B \subset \bar{A}} \lambda(B) \geq 0$$

such that $\lambda(\bar{A}) \geq \lambda(A)$.

Proof. Let $m = \inf_{B \subset A} \lambda(B)$. Then

- $m \leq \lambda(\emptyset) = 0$
- $m > -\infty$ by the previous lemma.

WLOG we can assume $m \neq 0$, otherwise $\bar{A} = A$ works.

Choose $B_1 \subset A$ s.t. $\lambda(B_1) \leq \frac{m}{2} < 0$. Then

$$A_1 = A \setminus B_1 \text{ satisfies } \lambda(A_1) \geq \lambda(A)$$
$$\inf_{B \subset A_1} \lambda(B) \geq \frac{m}{2}$$

Iteratively find A_j s.t. $A_{j+1} \subset A_j$ and

$$\lambda(A_j) \geq \lambda(A) \quad \text{and} \quad \inf_{B \subset A_j} \lambda(B) \geq \frac{m}{2^j}$$

Let $\bar{A} = \bigcap A_j$. Then $\lambda(\bar{A}) \geq \lambda(A)$ and $\inf_{B \subset \bar{A}} \lambda(B) \geq 0$.

<u>Proof</u> (Hahn–Jordan Decomp.)

Let $m_+ = \sup_A \lambda(A) < \infty$.

WLOG $m_+ > 0$, otherwise $\lambda(A) \leq 0 \ \forall A$ and $\Omega_- = \Omega$, $\Omega_+ = \emptyset$ works.

There are A_j s.t. $\lambda(A_j) \geq m_+ - \frac{1}{j}$ and (last lemma) $\bar{A}_j$ totally positive s.t. $\lambda(\bar{A}_j) \geq m_+ - \frac{1}{j}$.

$\Rightarrow \Omega_+ = \bigcup_j \bar{A}_j$ is totally positive

$\lambda(\Omega_+) = m_+$.

Claim: $\Omega_- = \Omega \setminus \Omega_+$ is totally negative

Otherwise there is $B \subset \Omega_-$ s.t. $\lambda(B) > 0$.

$\Rightarrow \lambda(\Omega_+ \cup B) = \lambda(\Omega_+) + \lambda(B) > m_+$ — contradiction.

Define $\mu_\pm = \lambda|_{\Omega_\pm}$.

<u>Defn.</u> If λ is a signed measure with Hahn–Jordan decomposition $\lambda = \mu_+ - \mu_-$ then the measure

$$|\lambda| = \mu_+ + \mu_-$$

is the <u>total variation</u> measure of λ.

Example. Let μ be a positive measure and $f: \Omega \to \mathbb{R}$ be μ-integrable. Then

$$\lambda(A) = \int_A f \, d\mu$$

defines a signed measure, $\mu_{\pm}(A) = \int_A f^{\pm} \, d\mu$, $\Omega_{\pm} = \{ \pm f \geq 0 \}$, and $|\lambda|(A) = \int_A |f| d\mu$.

Defn. A signed measure λ is absolutely continuous with respect to a positive measure μ, written $\lambda \ll \mu$, if

$$\mu(A) = 0 \Rightarrow \lambda(A) = 0 \quad \forall A \in \Sigma.$$

Radon-Nikodym Thm. If $\lambda \ll \mu$ on (Ω, Σ) then there is a μ-integrable (Ω, Σ)-measurable function f s.t.

$$\lambda(A) = \int_A f \, d\mu \quad \forall A \in \Sigma.$$

f is uniquely defined almost everywhere.

Defn. $f = \frac{d\lambda}{d\mu}$ is the Radon-Nikodym derivative of λ with respect to μ.

__Proof.__ Let

$$\lambda_a = \lambda - a\mu, \quad \Omega_+(a) = \text{totally pos. subset of } \lambda_a.$$
(defined up to λ-measure 0 set)

Note if $\frac{d\lambda}{d\mu} = f$ then $\{f > a\} = \Omega_+(a)$ up to meas. 0.

Up to sets of measure 0, $\Omega_+(a) \subset \Omega_+(b)$ if $b < a$.

WLOG, by removing null sets, $\Omega_+(a)$ is decreasing along $a \in \mathbb{Q}$.

Thus set

$$f(\omega) = \sup\{a \in \mathbb{Q} : \omega \in \Omega_+(a)\}.$$

Claim: f is measurable

$$\{\omega : f(\omega) > x\} = \{\omega : \omega \in \Omega_+(y) \text{ for some } y > x, y \in \mathbb{Q}\}$$
$$= \bigcup_{\substack{y > x \\ y \in \mathbb{Q}}} \Omega_+(y) \quad \Rightarrow f \text{ is measurable.}$$

Claim: $\lambda(|f| = \infty) = \mu(|f| = \infty) = 0$

$$A \subset \bigcap_a \Omega_+(a) \Rightarrow \lambda(A) - a\mu(A) \geq 0 \quad \forall a \in \mathbb{Q}$$
$$\Rightarrow \mu(A) = 0$$
$$\Rightarrow \lambda(A) = 0 \text{ by absolute cont.}$$

$$\Rightarrow \lambda(f = +\infty) = \mu(f = +\infty) = 0$$

$$A \cap \Omega_+(a) = \emptyset \ \forall a \Rightarrow \lambda(A) - a\mu(A) \leq 0 \ \forall a \in \mathbb{Q}$$
$$\Rightarrow \mu(A) = 0$$
$$\Rightarrow \lambda(A) = 0$$

$$\Rightarrow \lambda(f = -\infty) = \mu(f = -\infty) = 0$$

Claim: $\int f \, d\mu < \infty$ and $\lambda(A) = \int_A f \, d\mu$.

Let $E_{a,b} = \{\omega : a \leq f(\omega) \leq b\}$

$$\subseteq \Omega_+(a') \cap \Omega_+(b')^c \ \forall a' < a, b' > b.$$

$$\Rightarrow a\mu(A) \leq \lambda(A) \leq b\mu(A) \quad \forall A \subset E_{a,b}$$

$$\Rightarrow \lambda(A) - (b-a)\mu(A) \leq a\mu(A)$$
$$\lambda(A) + (b-a)\mu(A) \geq b\mu(A)$$

Let $E_n = E_{nh,(n+1)h}$ for some fixed $h > 0$, $n \in \mathbb{Z}$.

$$\Rightarrow \lambda(A \cap E_n) - h\mu(A \cap E_n) \leq nh\mu(A \cap E_n)$$
$$\leq \int_{A \cap E_n} f \, d\mu$$
$$\leq (n+1)h\,\mu(A \cap E_n)$$
$$\leq \lambda(A \cap E_n) + h(A \cap E_n)$$

$$\Rightarrow \lambda(A) - h\mu(A) \leq \int_A f \, d\mu \leq \lambda(A) + h\mu(A).$$

Exercise: Prove that f is integrable and

$$\lambda(A) = \int_A f \, d\mu.$$

Rk. If f and $\hat{f}$ are as in the theorem then $f = \tilde{f}$ a.e.

Rk. If $\lambda(A) \geq 0$ $\forall A \in \Sigma$ then $f \geq 0$ a.e.

4.2. Conditional expectation

Defn. Let (Ω, Σ, P) be a probability space and $B \in \Sigma$ satisfy $P(B) > 0$. Then the conditional probability of $A \in \Sigma$ given B is defined by

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

The conditional expectation is defined by

$$E[X \mid A] = \frac{\int_A X(\omega)\, dP}{P(A)}.$$

Let ξ be a r.v. taking discrete values a_j on A_j. with $\Omega = \bigcup A_i$ disjoint.

$$\Rightarrow P(B) = \sum_i P(A_j) P(B \mid A_j) = \sum_j P(B \mid \overbrace{\xi = a_j}^{A_j}) P(\xi = a_j)$$

How to extend this to $P(\xi = a) = 0$?

In the case ξ takes discrete values, consider

$$\Sigma' = \{\{\xi \in E\} : E \subset \{a_j\}\} \quad \text{(the } \sigma\text{-field gen. by } \xi\text{)},$$
$$= \sigma(A_j)$$

$$\Rightarrow \begin{aligned} P(B \mid \Sigma') &: \omega \in \Omega \mapsto P(B \mid A_j) \text{ if } \omega \in A_j \\ E[X \mid \Sigma'] &: \omega \in \Omega \mapsto E[X \mid A_j] \text{ if } \omega \in A_j \end{aligned}$$

are Σ'-measurable random variables.

Let $\Sigma' \subset \Sigma$ be a sub-σ-field.

If $\lambda \ll \mu$ then the RN thm can be applied on (Ω, Σ) and (Ω, Σ') giving two densities

$$f = \frac{d\lambda}{d\mu}\Big|_{\Sigma} \qquad f' = \frac{d\lambda}{d\mu}\Big|_{\Sigma'}$$

Note f is Σ-meas. and f' is Σ'-meas. ($\Rightarrow$ Σ-meas). but f is not Σ'-meas. in general.

<u>Defn.</u> Let X be a integrable r.v. on (Ω, Σ, P) and $\Sigma' \subset \Sigma$ a sub-σ-field. Define

$$E[X|\Sigma'] = \frac{d\lambda}{dP}\Big|_{\Sigma'} \quad \text{with} \quad \lambda(A) = \int_A X(\omega)\, dP$$

Thus $E[X|\Sigma']$ is a Σ'-measurable random variable.

<u>Fact.</u> (i) $E[1|\Sigma'] = 1$ a.e. $\left[\lambda = P \Rightarrow \frac{d\lambda}{dP}\Big|_{\Sigma'} = 1\right]$

(ii) $E[E[X|\Sigma']] = E[X]$ $\left[\int \frac{d\lambda}{dP}\Big|_{\Sigma'} dP = \int \frac{d\lambda}{dP}\Big|_{\Sigma'} dP\Big|_{\Sigma'} = \int d\lambda\right]$

(iii) $X \geq 0 \Rightarrow E[X|\Sigma'] \geq 0$ a.e.

(iv) $E[a_1 X_1 + b_1 X_2 | \Sigma'] = a_1 E[X_1|\Sigma'] + a_2 E[X_2|\Sigma']$ a.e. $\forall a_1, a_2$ const

Prop $E[\,|E[X|\Sigma']|\,] \leq E[|X|]$

Proof. Let $d\lambda = X\,dP$. Then

$$\int |X|\,dP = \int X_+\,dP + \int X_-\,dP = \sup_{A\in\Sigma} \lambda(A) - \inf_{A\in\Sigma} \lambda(A)$$

$$\vee|$$

$$\int |E[X|\Sigma']|\,dP = \quad \cdots \quad = \sup_{A\in\Sigma'} \lambda(A) - \inf_{A\in\Sigma'} \lambda(A).$$

Prop If Z is a bounded Σ'-meas. random var.,

$$E[XZ|\Sigma'] = Z\,E[X|\Sigma'] \quad \text{a.e.}$$

Proof. Let $Z = 1_E$, $E \in \Sigma'$. Then

$$\lambda^{XZ}(A) = \int_A 1_E\,X\,dP = \int_{A\cap E} X\,dP = \lambda^X(A\cap E).$$

$$\Rightarrow \left.\frac{d\lambda^{XZ}}{dP}\right|_{\Sigma'} = \left.\frac{d\lambda^X}{dP}\right|_{\Sigma'} 1_E$$

Extend from indicator to simple to bounded.

Tower property. If $\Sigma'' \subset \Sigma' \subset \Sigma$ then

$$E[X|\Sigma''] = E[\,E[X|\Sigma']\,|\Sigma'']$$

<u>Proof.</u> $E[X|\Sigma'] = \frac{d\lambda}{dP}\Big|_{\Sigma'}$

$$\Rightarrow E[E[X|\Sigma']|\Sigma''] = \frac{d\mu}{dP}\Big|_{\Sigma''}, \quad \mu(A) = \int_A \frac{d\lambda}{dP}\Big|_{\Sigma'} dP$$

$$= \int \frac{d\lambda}{dP} dP = \int X \, dP$$

$$= \frac{d\lambda}{dP}\Big|_{\Sigma''}$$

<u>Conditional Jensen ineq.</u> Let $\phi : \mathbb{R} \to \mathbb{R}$ be convex. Then

$$E[\phi(X)|\Sigma'] \geq \phi(E[X|\Sigma']) \quad \text{a.e.}$$

$$E[\phi(X)] \geq E[\phi(E[X|\Sigma'])]$$

<u>Proof.</u> Since ϕ is convex,

$$\phi = \sup_{y \in \mathbb{R}} [yx - \phi^*(y)] = \sup_{y \in \mathbb{R}} [yx - \phi^*(y)]$$

$$\Rightarrow E[\phi(X)|\Sigma] \geq E\Big[\sup_{y \in \mathbb{R}} [yX - \phi^*(y)] \,\Big|\, \Sigma'\Big]$$

$$\geq \sup_{y \in \mathbb{R}} E[yX - \phi^*(y) | \Sigma'] \quad \text{a.e.}$$

$$= \phi(E[X|\Sigma'])$$

Let $L^2(\Omega, \Sigma, \mu) = \{f : \Omega \to \mathbb{R},\ \Sigma\text{-meas.} : \int |f|^2 \, d\mu < \infty\}/\sim$
where $f \sim g$ if $f = g$ a.e.

$L^2(\Omega, \Sigma, \mu)$ is a Hilbert space with inner product

$$(X, Y) = \int XY \, d\mu$$

<u>Fact</u>. The map $X \mapsto E[X|\Sigma']$ is the orthogonal projection from $L^2(\Omega, \Sigma, \mu)$ to $L^2(\Omega, \Sigma', \mu)$.

<u>Prop</u>. The conditional expectation $Y = E[X|\Sigma']$ is the unique (up to a.e.) Σ'-measurable r.v. s.t.

$$\int_A X \, dP = \int_A Y \, dP \quad \forall A \in \Sigma'$$

<u>Proof</u>. If Y and Y' satisfy the conditions,

$$\int_A (Y - Y') \, dP = 0 \quad \forall A \in \Sigma'$$

Let $A = \{Y - Y' \geq \varepsilon > 0\}$. Then

$$\varepsilon P(A) \leq \int_A (Y - Y') \, dP = 0 \Rightarrow P(A) = 0.$$

$\Rightarrow P(Y - Y' > 0) = 0$, likewise $P(Y' - Y > 0) = 0$.

4.3. Conditional probability

Define $P(\omega, A|\Sigma') = E[\mathbb{1}_A|\Sigma'](\omega)$.

Thm.

(i) $P(\omega, \Omega|\Sigma') = 1$, $P(\omega, \phi|\Sigma') = 0$ a.e.-ω

(ii) $0 \leq P(\omega, A|\Sigma') \leq 1$ a.e.-ω $\forall A \in \Sigma$

(iii) For any count. disjoint $A_i \in \Sigma$,

$$P(\omega, \cup A_i|\Sigma') = \sum_i P(\omega, A_i|\Sigma') \quad \text{a.e.-}\omega$$

(iv) $P(\omega, A|\Sigma') = \mathbb{1}_A(\omega)$ a.e.-ω $\forall A \in \Sigma'$

Proof. Exercise from properties of $E[\cdot|\Sigma']$

Rk. Each case involves possibly different null sets (depending on the A_i).

Can one construct a version of the cond. prob. involving only one null set for all properties?

This is called regular conditional probability when possible.

Thm. Let P be a prob. measure on $([0,1], \mathcal{B}([0,1]))$. Let $\Sigma' \subset \mathcal{B}([0,1])$ be a sub-σ-field. Then there exist prob. distr. Q_x, $x \in [0,1]$ s.t.

$$x \mapsto Q_x(A) \text{ is } \Sigma'\text{-meas. } \forall A \in \mathcal{B}([0,1])$$

$$\int f(y)\, Q_x(dy) = E^P[f \mid \Sigma'] \quad P\text{-a.e.}$$

If Σ' is count. generated then

$$Q_x(A) = 1_A(x) \quad \forall A \in \mathcal{B}([0,1]).$$

5. Markov Chains

5.1. Definition. Let $(\mathcal{X}, \mathcal{F})$ be a state space.

X_0 is distr. according to some prob. μ_0 on $(\mathcal{X}, \mathcal{F})$.

Given $X_0 = x_0, \dots, X_{k-1} = x_{k-1}$, we would like to define the distribution of X_k by

$$P(X_k \in A) = \pi_k(x_0, \dots, x_{k-1}; A).$$

Assume:

- for each $(x_0, \dots, x_{k-1})$, π_k is a prob. meas. on $(\mathcal{X}, \mathcal{F})$.
- for each $A \in \mathcal{F}$, $(x_0, \dots, x_{k-1}) \mapsto \pi_k(x_0, \dots, x_{k-1}; A)$ is measurable w.r.t. $(\mathcal{X}^n, \mathcal{F}^n)$ (product space).

Then if μ_{k-1} is the distribution of $(X_0, \dots, X_{k-1})$ on $(\mathcal{X}^k, \mathcal{F}^k)$ define μ_k on $(\mathcal{X}^{k+1}, \mathcal{F}^{k+1}) = (\mathcal{X}^k, \mathcal{F}^k) \times (\mathcal{X}, \mathcal{F})$ by

$$\mu_k(A \times A_k) = \int \pi_k(\cdot, A_k)\, d\mu_{k-1}.$$

By Kolmogorov's consistency thm, one can define a measure P on $\mathcal{X}^{\mathbb{N}}$ in this way (if the conditions are satisfied, which is the case if $\mathcal{X}$ is a complete separable metric space).

Defn (X_k) is a Markov process if the transition probabilities satisfy

$$\pi_k(x_0, \dots, x_{k-1}; \cdot) = \pi_{k-1,k}(x_{k-1}; \cdot) \qquad \forall k$$

It is a time-homogeneous Markov process if

$$\pi_{k-1,k}(x_{k-1}; \cdot) = \pi(x_{k-1}; \cdot) \quad \forall k.$$

For a Markov process, define

$$\pi_{k,k+\ell+1}(x, A) = \int \pi_{k,k+\ell}(x, dy)\, \pi_{k+\ell,k+\ell+1}(y, A)$$
$$= \int \pi_{k,k+1}(x, dy_{k+1}) \cdots \pi_{k+\ell,k+\ell+1}(y_{k+\ell}, A).$$

Chapman-Kolmogorov eqns.

$$\pi_{k,n}(x, A) = \int \pi_{k,m}(x, dy)\, \pi_{m,n}(y, A) \quad \forall\, k < m < n$$

In the time-homogeneous case, define

$$\pi^{(k+1)}(x, A) = \int \pi^{(k)}(x, dy)\, \pi(y, A), \quad \pi^{(1)}(x, A) = \pi(x, A).$$

Chapman-Kolmogorov eqns.

$$\pi^{(k+\ell)}(x, A) = \int \pi^{(k)}(x, dy)\, \pi^{(\ell)}(y, A)$$

Prop. Let $\Sigma_m = \sigma(X_0, \ldots, X_m)$. Then

$$P[X_n \in A \mid \Sigma_m] = \pi_{m,n}(X_m, A) \quad \text{a.e. if } m < n$$

Proof. Need to show that

$$P[\{X_n \in A\} \cap B] = \int_B \pi_{m,n}(X_m, A)\, dP \quad \forall B \in \Sigma_m,\ A \in \mathcal{F}.$$

But indeed,

$$\begin{aligned}
&P[\{X_n \in A\} \cap B] \\
&= \int_{\{X_n \in A\} \cap B} dP = \int_B d\mu_0(x_0)\, \pi_{0,1}(x_0, dx_1) \cdots \pi_{m-1,m}(x_{m-1}, dx_m) \\
&\qquad\qquad \times \underbrace{\int \pi_{m,m+1}(x_m, dx_{m+1}) \cdots \pi_{n-1,n}(x_{n-1}, A)}_{\pi_{m,n}(x_m, A)} \\
&= \int_B \pi_{m,n}(X_m, A)\, dP.
\end{aligned}$$

Thm. Let P be a probability measure on the product space $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ with product σ-field. TFAE:

(i) $E[g(Z) \mid X,Y] = E[g(Z) \mid Y] \quad \forall g : \mathcal{Z} \to \mathbb{R}$ bd.

(ii) $E[f(X) \mid Y,Z] = E[f(X) \mid Y] \quad \forall f : \mathcal{X} \to \mathbb{R}$ bd.

(iii) $E[f(X)g(Z) \mid Y] = E[f(X) \mid Y]\, E[g(Z) \mid Y]$

$\forall f : \mathcal{X} \to \mathbb{R},\ g : \mathcal{Z} \to \mathbb{R}$ bd.

Proof. Assume (i). Then (iii) holds:

$$E[f(X)g(Z) \mid Y] = E[\underbrace{E[f(X)g(Z) \mid X,Y]}_{f(X)\,E[g(Z) \mid X,Y]} \mid Y]$$

$$\overset{(i)}{=} f(X)\, E[g(Z) \mid Y]$$

$$= E[f(X) \mid Y]\, E[g(Z) \mid Y]$$

Assume (iii). Then

$$E[f(X)\, b(Y)\, E[g(Z) \mid X,Y]]$$

$$= E[f(X)\, b(Y)\, g(Z)] = E[b(Y)\, E[f(X) g(Z) \mid Y]]$$

$$\overset{(iii)}{=} E[b(Y)\, E[f(X) \mid Y]\, E[g(Z) \mid Y]]$$

$$= E[b(Y) f(X)\, E[g(Z) \mid Y]]$$

$$\Leftrightarrow E[f(X)\, b(Y)\, (E[g(Z)\mid X,Y] - E[g(Z)\mid Y])] = 0$$

Since f and b are arbitrary, also

$$E[h(X,Y)\, (\underbrace{E[g(Z)\mid X,Y] - E[g(Z)\mid Y]}))] = 0$$

Taking $h(X,Y) =$ ", this implies

$$E[g(Z)\mid X,Y] = E[g(Z)\mid Y] \quad \text{a.e.}$$

<u>Interpretation</u>. X = past, Y = present, Z = future

<u>Defn</u>. A prob. measure μ on $(\mathcal{X}, \mathcal{F})$ is an <u>invariant measure</u> for the Markov chain if

$$\mu(A) = \int \pi(y, A)\, \mu(dy) \qquad \forall A \in \mathcal{F}.$$

This is written also as $\mu = \mu\pi$.

<u>Rk</u>. If $X_0 \sim \mu$ and μ is invariant, then P has marginal μ at every time, and $(X_0, X_1, \dots)$ can be extended to a stationary process $(X_n)_{n\in\mathbb{Z}}$.

5.2. Stopping times and strong Markov property

Defn. A random variable $\tau : \Omega \to \mathbb{N}_0 \cup \{+\infty\}$ is a stopping time if

$$\{\omega : \tau(\omega) \le n\} \text{ is } \Sigma_n\text{-measurable } \forall n$$

$$\Leftrightarrow \{\omega : \tau(\omega) = n\} \text{ is } \Sigma_n\text{-measurable } \forall n.$$

Example.

- $\tau = \min\{n : X_n \in A\}$ is a stopping time.
- $\tau = \min\{n : X_{n+1} \in A\}$ is not a stopping time.

Fact. Let τ be a stopping time. Then

$$\Sigma_\tau = \{A : A \in \Sigma_\infty \text{ and } A \cap \{\tau \le n\} \in \Sigma_n \text{ for all } n\}$$

is a σ-field and τ is Σ_τ-measurable.

Strong Markov Property. Let τ be a stopping time. Then a.e. on $\{\tau < \infty\}$:

$$P[X_{\tau+1} \in A_1, \dots, X_{\tau+n} \in A_n \mid \Sigma_\tau]$$

$$= \int_{A_1} \cdots \int_{A_n} \pi(X_\tau, dx_1) \cdots \pi(x_{n-1}, dx_n)$$

<u>Proof</u>. Let $A \in \Sigma_\tau$ and $A \subset \{\tau < \infty\}$. Then

$$P[A \cap \{X_{\tau+1} \in A_1\} \cap \cdots \cap \{X_{\tau+n} \in A_n\}]$$

$$= \sum_k P[A \cap \{\tau = k\} \cap \{X_{k+1} \in A_1\} \cap \cdots \cap \{X_{k+n} \in A_n\}]$$

$$= \sum_k \int_{A \cap \{\tau=k\}} dP \left(\int_{A_1} \cdots \int_{A_n} \pi(X_k, dx_{k+1}) \cdots \pi(x_{k+n-1}, dx_{k+n}) \right)$$

$$= \sum_k \int_{A \cap \{\tau=k\}} dP \left(\int_{A_1} \cdots \int_{A_n} \pi(X_\tau, dx_{k+1}) \cdots \pi(x_{k+n-1}, dx_{k+n}) \right)$$

$$= \int_A dP \int_{A_1} \cdots \int_{A_n} \pi(X_\tau, dx_1) \cdots \pi(x_{n-1}, dx_n).$$

5.3. Countable state space

We assume the state space $\mathcal{X}$ is $\{1, 2, \dots\}$ and consider time-homogeneous Markov chains.

Then $\Pi(x,y) = \Pi(x, \{y\})$, $x, y \in \mathcal{X}$, is a matrix s.t.

$$\Pi(x,y) \geq 0, \qquad \sum_y \Pi(x,y) = 1 \quad \forall x.$$

Such matrices are called <u>stochastic matrices</u>.

$\Pi^{(n)}(x,y)$ = entries of the n-th matrix power of Π.

$\Pi^{(0)}(x,y) = \delta_{x,y}$ = entries of identity matrix by conv

<u>Defn</u>. The state $x \in \mathcal{X}$ <u>communicates</u> with $y \in \mathcal{X}$ if

$$\Pi^{(n)}(x,y) > 0 \qquad \text{for some } n.$$

The Markov chain is <u>irreducible</u> if all states communicate with each other.

Let τ_x be the first passage time to x:

$$\tau_x = \inf\{n \geq 1 : X_n = x\}$$

Let $f_n(x) = P_x[\tau_x = n]$.

$\quad X_0 = x$

$$\Rightarrow f_n(x) = P_x[X_j \neq x \text{ for } 1 \le j \le n-1,\ X_n = x]$$
$$= \sum_{\substack{y_1 \neq x \\ y_2 \neq x \\ \vdots \\ y_{n-1} \neq x}} \pi(x, y_1)\,\pi(y_1, y_2) \cdots \pi(y_{n-1}, x)$$

and

$$P_x(\tau_x < \infty) = \sum_n f_n(x) \le 1.$$

<u>Defn</u>. The state x is <u>recurrent</u> if $P_x(\tau_x < \infty) = 1$, <u>transient</u> otherwise.

Let $V_x = \#$ visits to $x = \sum_{n=0}^{\infty} 1_{X_n = x}$

<u>Fact</u>. x is recurrent $\Leftrightarrow P_x[V_x = \infty] = 1 \Leftrightarrow E_x[V_x] = \infty$
x is transient $\Leftrightarrow P_x[V_x = \infty] = 0 \Leftrightarrow E_x[V_x] < \infty$

<u>Proof</u>. By the strong Markov property,

$$P_x[V_x > r] = P_x[\tau_x < \infty]^r$$

$$P_x[V_x = \infty] = \lim_{r \to \infty} P_x[V_x > r] = \lim_{r \to \infty} P_x[\tau_x < \infty]^r = \begin{cases} 1 & \text{rec.} \\ 0 & \text{trans.} \end{cases}$$

$$E_x[V_x] = \sum_{r \ge 1} P[V_x > r] = \frac{1}{1 - P_x[\tau_x < \infty]} \begin{cases} = \infty & \text{rec} \\ < \infty & \text{trans.} \end{cases}$$

Define the Green's function

$$G(x,y) = E_x[V_y] = \sum_n \pi^{(n)}(x,y)$$

<u>Prop</u>. For an irreducible chain, either all states are transient or all are recurrent:

$$G(x,x) = \infty \text{ for some } x \Rightarrow G(x,x) = \infty \text{ for all } x.$$

<u>Proof</u>. By irreducibility there exist n, m s.t.

$$\pi^{(n)}(x,y) > 0, \quad \pi^{(m)}(y,x) > 0.$$

$$\Rightarrow \pi^{(r+n+m)}(x,x) \geq \pi^{(n)}(x,y)\pi^{(r)}(y,y)\pi^{(m)}(y,x).$$

$$\Rightarrow \underbrace{\sum_r \pi^{(r)}(y,y)}_{G(y,y)} \leq \frac{1}{\pi^{(n)}(x,y)\,\pi^{(m)}(y,x)} \underbrace{\sum_r \pi^{(r)}(x,x)}_{G(x,x)}$$

Therefore $G(y,y) = \infty \Rightarrow G(x,x) = \infty$ and vice versa.

<u>Lemma</u>. $G(x,x) = \dfrac{1}{1 - P_x(\tau_x < \infty)}$

$$G(x,y) = P_x(\tau_y < \infty)\, G(y,y)$$

and if the chain is irreducible it is recurrent iff

$$G(x,y) = \infty \ \& \ P_x(\tau_y < \infty) = 1 \qquad \text{for all } x,y.$$

Proof. $G(x,x) = 1/(1-P_x(\tau_x<\infty))$ was shown in last pf.

$$E_x[V_y] = E_x[\underbrace{E[V_y \mid \Sigma_y]}_{E_y[V_y] \text{ by strong Markov property}} 1_{\tau_y<\infty}]$$

$$= P_x(\tau_y<\infty)\, E_y[V_y]$$

If the chain is irreducible then for any y there is n s.t.

$$P_x(\tau_y<\infty) \geq \pi^{(n)}(x,y) > 0.$$

Moreover, if the chain is recurrent,

$$1 = P_x[V_x=\infty] = P_x\left[\sum_k 1_{X_k=x} = \infty\right]$$

$$\leq P_x\left[\sum_{k\geq n+1} 1_{X_k=x} > 0\right]$$

(Markov prop.) $$= \sum_z \underbrace{P_x[X_n=z]}_{\pi^{(n)}(x,z)} \underbrace{P_z\left[\sum_{k\geq 1} 1_{X_k=x} > 0\right]}_{P_z[\tau_x<\infty]}$$

Since $\sum \pi^{(n)}(x,z) = 1$ and $\pi^{(n)}(x,y) > 0$,

$$P_y[\tau_x<\infty] = 1.$$

__Fact.__ If $P_x[\tau_y < \infty] > 0$ then $P_x[T_y < T_x] > 0$.

__Proof.__ Suppose $P_x[T_y < T_x] = 0$. Then $P_x[T_y < \infty] = 0$ by the strong Markov property.

Let $m(x) = E_x[T_x]$ be the __expected return time__.

__Defn.__ The recurrent state x is

__positive recurrent__ if $m(x) < \infty$

__null recurrent__ if $m(x) = \infty$.

__Prop.__ For an irreducible recurrent chain, all states are of the same recurrence type.

__Proof.__

$$E_x[\tau_x] \geq E_x[\tau_x 1_{T_y < \tau_x}]$$

$$= E_x[E_y[\tau_x] 1_{T_y < \tau_x}] = E_y[\tau_x] \underbrace{P(\tau_y < \tau_x)}_{p > 0}$$

$$\Rightarrow E_y[\tau_x] \leq \frac{1}{p} E_x[\tau_x]$$

In the other direction,

$$E_x[\tau_y] \le E_x[\tau_x 1_{\tau_y \le \tau_x}] + \underbrace{E_x[\tau_y 1_{\tau_x < \tau_y}]}_{E_x[(\tau_x + E_x[\tau_y]) 1_{\tau_x < \tau_y}]}$$

$$= E_x[\tau_x] + E_x[\tau_y](1-p)$$

$$\Rightarrow E_x[\tau_y] \le \frac{1}{p} E_x[\tau_x]$$

Thus

$$E_y[\tau_y] \le E_y[\tau_y 1_{\tau_y > \tau_x}] + E_y[\tau_x 1_{\tau_y \le \tau_x}]$$

$$= E_y[(\tau_x + E_x[\tau_y]) 1_{\tau_y > \tau_x}] + E_y[\tau_x 1_{\tau_y \le \tau_x}]$$

$$\le E_y[\tau_x] + E_x[\tau_y] \le \frac{2}{p} E_x[\tau_x]$$

Thus if x is pos. rec. so is y.

<u>Defn</u>. A state x is aperiodic if $\pi^{(n)}(x,x) > 0$ for all $n \ge n_0(x)$. A Markov chain is aperiodic if all states are aperiodic.

<u>Exercise</u>. For an irreducible Markov chain, if there is an aperiodic state, all states are aperiodic.

Thm. Consider an aperiodic positive recurrent Markov chain. Then, for all x, y,

$$\lim_{n\to\infty} \pi^{(n)}(x,y) = q(y) = \frac{1}{m(y)}.$$

and q is an invariant distribution:

$$\sum q(x) = 1, \quad q\pi = q \quad \Big(\text{i.e. } \sum_x q(x)\pi(x,y) = q(y)\Big) \quad \forall y$$

Proof. Fix x and set

$$f_n = P_x[\tau_x = n], \quad p_n = \pi^{(n)}(x,x), \quad m = m(x).$$

Then:

$$(*)\begin{cases} f_n \geq 0 \text{ for all } n, \quad f_n > 0 \\ \sum_n f_n = 1, \quad \sum_n n f_n = m \\ p_n = \sum_{j=1}^{n} f_j p_{n-j}, \quad p_0 = 1 \end{cases}$$

for $n \in S$ where S is such that every $n \geq n_0$ can be written $n = j_1 + \cdots + j_k$, $j_i \in S$

Lemma. If $(*)$ holds then

$$\lim_{n\to\infty} p_n = \frac{1}{m}$$

The lemma implies $\lim_{n\to\infty} \pi^{(n)}(x,x) = \frac{1}{m(x)}$.

For general x, y,

$$\pi^{(n)}(x,y) = \sum_k P_x[\tau_y = k]\, \pi^{(n-k)}(y,y)$$

and $\sum_k P_x[\tau_y = k] = P_x[\tau_y < \infty] = 1$ by recurrence. Thus

$$\lim_{n\to\infty} \pi^{(n)}(x,y) \overset{(DCT)}{=} \sum_k P_x[\tau_y = k] \frac{1}{m(y)} = \frac{1}{m(y)} \quad \forall\, x,y.$$

Claim: $\sum_y q(y) = 1$.

Indeed, by Fatou,

$$\sum_y q(y) = \sum_y \lim_{n\to\infty} \pi^{(n)}(x,y) \leq \liminf_{n\to\infty} \sum_y \pi^{(n)}(x,y) = 1.$$

On the other hand,

$$\pi^{(n+1)}(x,y) = \sum_z \pi^{(n)}(x,z)\, \pi(z,y)$$

$$\Rightarrow\ q(y) \geq \sum_z q(z)\, \pi(z,y) \qquad \text{by Fatou}$$

$$\Rightarrow\ \sum_y q(y) \geq \sum_{z,y} q(z)\, \pi(z,y) = \sum_z q(z)$$

$$\Rightarrow\ q(y) = \sum_z q(z)\, \pi(z,y)$$

$$\Rightarrow\ q(y) = \sum_z q(z)\, \pi^{(n)}(z,y) \overset{(n\to\infty)}{=} \sum_z q(z) \sum_y q(y) = 1.$$

Proof of lemma. Clearly $p_n \le 1$ for all n.

Choose a subseq. s.t. $p_{n_k} \to q_0 = \limsup_n p_n$.

and $p_{n_k+j} \to q_j$ for all $j \in \mathbb{Z}$.

$\Rightarrow$ $q_n = \sum_{j=1}^{\infty} f_j q_{n-j}$ for all $n \in \mathbb{Z}$. $(*_n)$

$\Rightarrow$ $q_0 = \sum_{j=1}^{\infty} f_j q_{-j}$ $(*_0)$

Claim: $q_j = q_0$ for all j

Note that $q_j \le q_0 = \limsup p_n$ for all j.

Since $f_j > 0$ for all $j \in S$, $q_{-j} = q_0$ for $j \in S$ by $(*_0)$.

Then $q_{-j} = q_0$ for $j = j_1 + j_2$, $j_i \in S$ by $(*_j)$.

Then $q_{-j} = q_0$ for $j = j_1 + \cdots + j_k$, $j_i \in S$ again by $(*_j)$.

By assumption on S, $q_{-j} = q_0$ for $j \ge n_0$.

By $(*_{-n_0+1})$,

$$q_{-n_0+1} = \sum_{j=1}^{\infty} f_j q_{-n_0+1-j} = q_0 \sum_{j=1}^{\infty} f_j = q_0$$

By induction $q_n = q_0$ for all $n \in \mathbb{Z}$.

Claim: Let $T_j = \sum_{i=j}^{\infty} f_i$. Then $T_1 = 1$, $\sum_{j=0}^{\infty} T_j = m$ and

$$\sum_{j=1}^{N} T_j p_{N-j+1} = 1 - p_N$$

Indeed, $p_n = \sum_{j=1}^{n} f_j p_{n-j}$ implies

$$\sum_{n=1}^{N} p_n = \sum_{n=1}^{N} \sum_{j=1}^{n} f_j p_{n-j} = \sum_{n=1}^{N} \sum_{j=0}^{n-1} f_{n-j} p_j$$

$$= \sum_{j=0}^{N-1} p_j \sum_{n=j+1}^{N} f_{n-j}$$

$$= \sum_{j=0}^{N-1} p_j \sum_{i=1}^{N-j} f_i = \sum_{j=0}^{N-1} p_j \left(1 - T_{N-j+1}\right)$$

$$\Rightarrow 1 - p_N = \sum_{j=1}^{N} T_j p_{N-j+1}$$

Conclusion: Recall that $p_{n_k+j} \to q_0$, $p_{n_k+j} \leq q_0$.

If $\sum_{j=0}^{\infty} T_j = m < \infty$ then by DCT,

$$1 - q_0 = q_0 \sum_{j=1}^{\infty} T_j \Rightarrow m = \sum_{j=0}^{\infty} T_j = \frac{1}{q_0}$$

$$\Rightarrow \limsup p_n = \frac{1}{m}$$

An analogous argument shows

$$\liminf p_n = \frac{1}{m}.$$

If $\sum T_j = \infty$ then by Fatou

$$1 - q_0 = \liminf_{k\to\infty} (1 - p_{n_k}) = \liminf_{k\to\infty} \sum T_j p_{n_k - j + 1}$$

$$\geq q_0 \sum T_j \Rightarrow q_0 = 0.$$

$\Rightarrow \lim p_n = 0.$

<u>Example</u>. Simple random walk on $\mathbb{Z}^d$

$d \leq 2$, symmetric: recurrent (null recurrent)

$d > 2$ or asymmetric: transient

5.4. Example: Polya urn

Defn A bounded function $\mathcal{U}: \mathcal{X} \to \mathbb{R}$ is harmonic with respect to the transition probabilities of a Markov chain if

$$\mathcal{U}(x) = \int \pi(x, dy)\, \mathcal{U}(y) \qquad \forall x \in \mathcal{X}.$$

Let $\mathcal{U}$ be bounded harmonic. Then

$$E[\mathcal{U}(X_{n+1}) - \mathcal{U}(X_n) \mid \mathcal{F}_n] = \int \pi(X_n, dy)(\mathcal{U}(y) - \mathcal{U}(X_n)) = 0.$$

Thus $\eta_n = \mathcal{U}(X_n) - \mathcal{U}(X_{n-1})$ satisfies

$$E[\eta_n] = 0, \quad E[\eta_n \eta_m] = 0 \quad \forall n \neq m.$$

Thus

$$\mathcal{U}(X_n) = \mathcal{U}(X_0) + \eta_1 + \cdots + \eta_n$$

$$\Rightarrow E_x[\mathcal{U}(X_n)^2] = \mathcal{U}(x)^2 + E_x[\eta_1^2] + \cdots + E_x[\eta_n^2] \leq \sup \mathcal{U}$$

$\Rightarrow \mathcal{U}(X_n)$ is a Cauchy sequence in L^2

$\Rightarrow \xi = \lim_{n\to\infty} \mathcal{U}(X_n)$ exists, $E_x[\xi] = \mathcal{U}(x)$

Rk. $\mathcal{U}(X_n)$ is a martingale and the limit exists a.s.

Polya urn: $\mathcal{X} = \mathbb{N} \times \mathbb{N}$, $\pi((p,q),(p+1,q)) = \frac{p}{p+q}$

$\pi((p,q),(p,q+1)) = \frac{q}{p+q}$.

Thm. $\xi_n = \frac{P_n}{P_n + Q_n} \longrightarrow \xi$ in $L^2(P)$ and a.s.

and, if $(P_0, Q_0) = (p,q)$, ξ has distribution

$$\frac{1}{\beta(p,q)} x^{p-1}(1-x)^{q-1} \, dx$$

Lemma. For any $x \in (0,1)$,

$$F_x(p,q) = \frac{1}{\beta(p,q)} x^{p-1}(1-x)^{q-1}, \qquad \beta(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$$

is bounded harmonic and so is $\frac{p}{p+q}$.

Proof. $\frac{p}{p+q} F(p+1,q) + \frac{q}{p+q} F(p,q+1) = F(p,q)$

Lemma For any $f: [0,1] \to \mathbb{R}$ continuous,

$$F_f(p,q) = \int f(x) \, F_x(p,q) \, dx$$

is bounded harmonic and

$$\lim_{\substack{p,q \to \infty \\ \frac{p+q}{p} \to x}} F_f(p,q) = f(x).$$

Sketch.

$$\lim_{r\to\infty} F_f(xr, (1-x)r) = \lim_{r\to\infty} \int_0^1 f(y) \underbrace{\frac{1}{B(xr,(1-x)r)}} y^{xr-1}(1-y)^{(1-x)r-1}\, dy$$

$$n! = \sqrt{2\pi n}\, e^{-n} n^n \qquad \sim \frac{1}{\sqrt{2\pi r x(1-x)}\, x^{xr}(1-x)^{(1-x)r}}$$

$$\approx \lim_{r\to\infty} \int_0^1 f(y) \left(\frac{y}{x}\right)^{xr} \left(\frac{1-y}{1-x}\right)^{(1-x)r} dy$$

$$\approx \lim_{r\to\infty} \int_0^{1/x} f(xs)\; s^{xr} \left(\frac{1-sx}{1-x}\right)^{(1-x)r} ds$$

$\to 0$ unless $s \geq 1$ $\qquad$ $\to 0$ unless $s \leq 1$

$$\to f(x)$$

Proof of Thm. Since $U(p,q) = \frac{p}{p+q}$ is bd. harmonic,

$$\xi = \lim_{n\to\infty} U(P_n, Q_n) = \frac{P_n}{P_n + Q_n} \text{ exists in } L^2.$$

Since F_f is bd. harmonic, $F_f(p,q) \to f(x)$ if $\frac{p}{p+q} \to x$, $p \to \infty$,

$$\lim_{n\to\infty} F_f(P_n, Q_n) \text{ exists in } L^2$$

$$\lim_{n\to\infty} F_f(P_n, Q_n) = f(\xi)$$

On the other hand

$$\mathbb{E}_{p,q}[f(\xi)] = \lim_{n\to\infty} \mathbb{E}_{p,q}[F_f(P_n, Q_n)]$$
$$= F_f(p,q)$$
$$= \int f(x) \frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1}\,dx$$

$\Rightarrow \xi \sim \mathrm{Beta}(p,q)$.

6. Martingales

6.1. Aside: L^p spaces

Let $\|f\|_{L^p(\mu)} = \|f\|_p = \left(\int |f|^p \, d\mu\right)^{1/p}$.

Hölder's inequality. Let $p, q \in [1, \infty]$ be s.t. $\frac{1}{p} + \frac{1}{q} = 1$.
Then

$$\int |fg| \, d\mu \leq \|f\|_{L^p(\mu)} \|g\|_{L^q(\mu)}.$$

Proof. WLOG $\|f\|_{L^p(\mu)} \in (0, \infty)$ and then $\|f\|_{L^p(\mu)} = 1$.
Then define a prob. meas. $P(A) = \int_A |f|^p \, d\mu$.
Thus $E[X] = \int X |f|^p \, d\mu$.
By Jensen, $E[|X|] \leq E[|X|^q]^{1/q}$.

$$\Rightarrow \int |fg| \, d\mu = \int \frac{|g|}{|f|^{p-1}} |f|^p 1_{|f|>0} \, d\mu$$

$$= E\left[\frac{|g|}{|f|^{p-1}} 1_{|f|>0}\right]$$

$$\leq E\left[\frac{|g|^q}{|f|^{(p-1)q}} 1_{|f|>0}\right]^{1/q}$$

$$= \left(\int |g|^q \cancel{\frac{|f|^p}{|f|^{(p-1)q}}}^{=1} 1_{|f|>0} \, d\mu\right)^{1/q} \leq \|g\|_{L^q(\mu)}.$$

Minkowski Inequality.

$$\|f+g\|_{L^p} \le \|f\|_{L^p} + \|g\|_{L^p}$$

Proof. Assume $p>1$ and $\|f\|_{L^p} + \|g\|_{L^p} < \infty$. Then

$$|f+g|^p \le (|f|+|g|)^p \le 2^p(|f|^p+|g|^p)$$

$\Rightarrow \|f+g\|_{L^p} < \infty$.

$$\int |f+g|^p \, d\mu \le \int |f|\,|f+g|^{p-1} \, d\mu + \int |g|\,|f+g|^{p-1} \, d\mu$$

$$\le \left(\|f\|_{L^p} + \|g\|_{L^p}\right) \underbrace{\| |f+g|^{p-1} \|_{L^q}}_{\|f+g\|_{L^p}^{p/q}}$$

$$\Rightarrow \|f+g\|_{L^p} = \|f+g\|_{L^p}^{p(1-\frac{1}{q})} \le \|f\|_{L^p} + \|g\|_{L^q}.$$

Completeness. Let $p \in [1,\infty]$. For any $(f_n) \subset L^p$ s.t.

$$\|f_n - f_m\|_{L^p} \to 0 \quad (n,m \to \infty)$$

there is $f \in L^p$ s.t.

$$\|f_n - f\|_{L^p} \to 0.$$

<u>Proof</u> (for $p<\infty$). Find a subsequence (n_k) s.t.

$$S = \sum_{k=1}^{\infty} \|f_{n_{k+1}} - f_{n_k}\|_{L^p} < \infty$$

$$\Rightarrow \Big\| \sum_{k=1}^{K} |f_{n_{k+1}} - f_{n_k}| \Big\|_{L^p} \underset{\substack{\uparrow \\ \text{Minkowski}}}{\leq} S$$

$$\Rightarrow \Big\| \sum_{k=1}^{\infty} |f_{n_{k+1}} - f_{n_k}| \Big\|_{L^p} \underset{\substack{\uparrow \\ \text{monotone conv.}}}{\leq} S$$

$$\Rightarrow \sum_{k=1}^{\infty} |f_{n_{k+1}} - f_{n_k}| < \infty \text{ a.e.}$$

$\Rightarrow f_{n_k}(\omega)$ converges for $\omega \in \Omega \setminus N$ with $\mu(N) = 0$

Define $f(\omega) = \begin{cases} \lim f_{n_k}(\omega) & \text{for } \omega \in \Omega \setminus N \\ 0 & \text{for } \omega \in N. \end{cases}$

Since (f_n) is Cauchy in L^p,

$$\int |f_n - f_m|^p \, d\mu \leq \varepsilon \quad \text{for } m \geq n \geq n(\varepsilon).$$

$$\Rightarrow \int |f_n - f_{n_k}|^p \, d\mu \leq \varepsilon \quad \text{for } n \geq n(\varepsilon),\ k \geq k(\varepsilon).$$

$$\Rightarrow \int |f_n - f|^p \, d\mu = \int \liminf |f_n - f_{n_k}|^p \, d\mu \underset{\substack{\uparrow \\ \text{Fatou}}}{\leq} \underbrace{\liminf \int |f_n - f_{n_k}| \, d\mu}_{\leq \varepsilon}$$

$$\Rightarrow f \in L^p,\ \|f - f_n\|_{L^p} \to 0.$$

<u>Cor.</u>

$$L^p(\Omega, \mathcal{F}, \mu) = \{ f : \Omega \to \mathbb{R} \text{ meas.}, \|f\|_{L^p} < \infty \} /_{\sim}$$

where $f \sim g$ if $f = g$ a.e. is a Banach space.

<u>Duality</u>. Let $q \in [1, \infty]$. Then the map for every $g \in L^q$,

$$\Lambda_g : L^p \longrightarrow \mathbb{R}, \quad \Lambda_g(f) = \int fg \, d\mu$$

is bounded with $\|\Lambda_g\| = \sup_{\|f\|_{L^p} \leq 1} |\Lambda_g(f)| = \|g\|_{L^q}$ and if $p < \infty$ then for any bounded linear Λ on L^p there is $g \in L^q$ s.t. $\Lambda = \Lambda_g$.

From now, assume μ is a finite measure.

<u>Defn</u>. $(g_j) \subset L^1$ is <u>uniformly integrable</u> if it is bounded and

$$\sup_j \int_{|g_j| > t} |g_j| \, d\mu \longrightarrow 0 \quad (t \to \infty).$$

<u>Weak compactness</u>. Let $(g_j) \subset L^q$ be bounded, $q > 1$. Then there is a subsequence s.t.

$$(g_j, f) \longrightarrow (g, f) \ \forall f \in L^p.$$

The same holds if $(g_j) \subset L^1$ is uniformly integrable.
($q > 1$: Banach–Alaoglu, $q = 1$: Dunford–Pettis)

<u>Vitali's convergence theorem</u> (finite measure version).
Let (f_j) and f be measurable. TFAE:

(i) $f_j \in L^1$ for all j, $f \in L^1$ and $f_j \to f$ in L^1

(ii) (f_j) is uniformly integrable and $f_j \to f$ in measure.

6.2. Martingales and Doob's inequality

Defn. Let $(\Omega, \mathcal{F}, P)$ be a probability space. A martingale of length n is a sequence of random variables $X_1, \dots, X_n$ and sub-σ-fields $\mathcal{F}_i$ s.t.

- $E[|X_i|] < \infty$, X_i is $\mathcal{F}_i$-measurable
- $\mathcal{F}_{i+1} \supseteq \mathcal{F}_i$ for every i
- $X_i = E[X_{i+1} | \mathcal{F}_i]$ a.e. for every $i = 1, \dots, n-1$.

An infinite martingale sequence is defined in the same way.

Rk. Let $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $X_0 = E[X_i]$ (which is indep. of i). Then $X_0, X_1, \dots$ is a martingale as well.

Rk. If (X_i) is a martingale, $Y_i = X_i - X_{i-1}$ satisfies

$$E[Y_{i+1} | \mathcal{F}_i] = 0 \quad \text{a.e.}$$

Such a sequence is called martingale difference sequence.

Rk. Let $(\mathcal{F}_i)$ be an increasing sequence of σ-fields and X a random variable s.t. $E[|X|]<\infty$. Then

$$X_i = E[X \mid \mathcal{F}_i]$$

is a martingale (called Doob martingale ass. with X).

Defn. $(X_i, \mathcal{F}_i)$ is a

- submartingale if $X_i \le E[X_{i+1} \mid \mathcal{F}_i]$ a.e. $\forall i$
- supermartingale if $X_i \ge E[X_{i+1} \mid \mathcal{F}_i]$ a.e. $\forall i$

Lemma. Let $(X_i, \mathcal{F}_i)$ be a martingale and φ a convex function s.t. $\varphi(X_i)$ is integrable for every i. Then $(\varphi(X_i), \mathcal{F}_i)$ is submartingale.
In particular, $(|X_i|^p, \mathcal{F}_i)$ is a submartingale if $p \ge 1$.

Pf. Jensen.

Doob's Inequality. Let $X_1, \dots, X_n$ be a martingale. Then

$$P\left[\sup_{1\le j\le n} |X_j| \ge \ell\right] \le \frac{1}{\ell} E\left[|X_n| \, 1_{\sup_{1\le j\le n}|X_j| \ge \ell}\right] \le \frac{1}{\ell} E[|X_n|]$$

&

$$E\left[\sup_{1\le j\le n} |X_j|^p\right] \le \left(\frac{p}{p-1}\right)^p E[|X_n|^p] \quad \text{for any } p>1.$$

Proof. Let $S = \sup_{1 \le j \le n} |X_j|$. Then

$$\{S \ge \ell\} = \bigcup_j E_j, \quad E_j = \{|X_1| < \ell, \dots, |X_{j-1}| < \ell, |X_j| \ge \ell\}.$$

$$\Rightarrow P(E_j) \le \frac{1}{\ell} E[|X_j| \, 1_{E_j}] \underset{\uparrow}{\le} \frac{1}{\ell} E[|X_n| \, 1_{E_j}]$$

$|X_j|$ is a submartingale

$$\Rightarrow P[S \ge \ell] \le \frac{1}{\ell} \sum_j E[|X_n| \, 1_{E_j}] = \frac{1}{\ell} E[|X_n| \, 1_{S \ge \ell}]$$

The second claim is a consequence of the next general lemma.

Lemma. Let $p > 1$. Then for r.v. $X \ge 0$, $Y \ge 0$,

$$P[Y \ge \ell] \le \frac{1}{\ell} \int_{|Y| > \ell} X \, dP$$

$$\Rightarrow E[Y^p] \le \left(\frac{p}{p-1}\right)^p E[X^p]$$

Proof. Let $T(y) = P[Y \ge y]$. Then

$$E[Y^p] = p \int_0^\infty y^{p-1} T(y) \, dy$$

$$\le p \int_0^\infty \frac{dy}{y} y^{p-1} \int_{Y \ge y} X \, dP \quad \text{(assumption)}$$

$$\Rightarrow \quad E[Y^p] \le p \int X \left[\int_0^Y y^{p-2}\, dy \right] dP \qquad \text{(Fubini)}$$

$$= \frac{p}{p-1} \int X\, Y^{p-1}\, dP$$

$$\le \frac{p}{p-1} E[X^p]^{1/p} \underbrace{E[Y^{q(p-1)}]^{1/q}}_{E[Y^p]^{1-\frac{1}{p}}} \qquad \text{(Hölder)}$$

$$\Rightarrow \quad E[Y^p] \le \left(\frac{p}{p-1}\right)^p E[X^p] \qquad \text{if } E[Y^p] < \infty$$

In general, apply to $Y \wedge N$ to get

$$E[(Y \wedge N)^p] \le \left(\frac{p}{p-1}\right)^p E[X]^p$$

$$\overset{N\to\infty}{\Rightarrow} \quad E[Y^p] \le \left(\frac{p}{p-1}\right) E[X^p].$$

6.3. Martingale Convergence Theorems

Let $X \in L^p$, $p \geq 1$ and set $X_n = E[X \mid \mathcal{F}_n]$.

Then $|X_n|^p$ is a submartingale and $E[|X_n|^p] \leq E[|X|^p]$.

<u>Thm</u>. $\lim_{n\to\infty} E[|X_n - X|^p] \to 0$

<u>Proof</u>. Assume X is bounded. Then $X \in L^2$ and we have already seen $X_n \to X$ in L^2. Indeed,

$$E[X_n^2] = E[X_0^2] + E[\eta_1^2] + \cdots + E[\eta_n^2], \quad \eta_n = X_n - X_{n-1}.$$

$$\Rightarrow \sum_n E[\eta_n^2] < E[|X|^2] < \infty$$

$$\Rightarrow E[(X_n - X_m)^2] \leq \sum_{k=m+1}^{n} E[\eta_k^2] \to 0 \quad (n, m \to \infty)$$

Thus $(X_n)_n$ is Cauchy in L^2 and therefore $Y = \lim_{n\to\infty} X_n$ exists (by completeness of L^2).

Claim: $X = Y$ a.s.

$$\int_A Y \, dP \underset{\substack{\uparrow \\ X_n \to Y \text{ in } L^2 \\ \Rightarrow X_n \to Y \text{ in } L^1}}{=} \lim_{n\to\infty} \int_A X_n \, dP \underset{\substack{\uparrow \\ \text{defn. of } X_n}}{=} \int_A X \, dP \quad \forall A \in \bigcup_m \mathcal{F}_m$$

$\Rightarrow$ Same for $A \in \mathcal{F} = \sigma(\bigcup \mathcal{F}_m) \Rightarrow X = Y$ a.s.

Since X is bd, also $X_n \to X$ in L^p $\forall p \in [1, \infty)$.
For any $X \in L^p$, there is $X' \in L^\infty$ s.t. $\|X - X'\|_{L^p} < \varepsilon$.
Then $X_n' = E[X' | \mathcal{F}_n]$ satisfies $\|X_n' - X_n\|_{L^p} < \varepsilon$ $\forall n$.
Since $X_n' \to X'$ in L^p,

$$\limsup_{n\to\infty} \|X_n - X\|_{L^p} < 2\varepsilon \quad \forall \varepsilon > 0$$

$$\Rightarrow \lim_{n\to\infty} \|X_n - X\|_{L^p} = 0.$$

<u>Thm</u>. Let (X_n) be a martingale that is bd in L^p: $(p>1)$.

$$\sup_n E[|X_n|^p] < \infty$$

Then there is $X \in L^p$ s.t. $X_n = E[X | \mathcal{F}_n]$.

<u>Proof</u>. Since (X_n) is bd in L^p, $p>1$, which is the dual space of L^q, $\frac{1}{p}+\frac{1}{q}=1$, the Banach-Alaoglu Thm implies that there is a subsequence s.t. $X_{n_j} \to X$ weakly in L^p, i.e.,

$$E[X_{n_j} Y] \longrightarrow E[XY] \quad \forall Y \in L^q$$

$$\Rightarrow \int_A X_{n_j}\, dP \longrightarrow \int_A X\, dP \quad \forall A$$

For $A \in \mathcal{F}_m$, since (X_n) is a martingale:

$$\lim_{j\to\infty} \int_A X_{n_j}\, dP = \int_A X_m\, dP$$

Thus

$$\int_A X_m\, dP = \int_A X\, dP \quad \forall A \in \mathcal{F}_m.$$

$$\Rightarrow X_m = E[X \mid \mathcal{F}_m].$$

<u>Thm</u>. Let $X \in L^p$, $p \geq 1$. Then $X_n = E[X \mid \mathcal{F}_n] \to X$ a.s.

<u>Proof</u>. Assume $p=1$. Let

$$M = \{X \in L^1 : E[X \mid \mathcal{F}_n] \to X \text{ a.s.}\}.$$

Note M is a linear space, so if it is dense and closed in L^1 then $M = L^1$ proving the result.

Let $M_n = \{X \in L^1 : X \text{ is } \mathcal{F}_n\text{-measurable}\}$.

Then M_n is a closed subspace of L^1 and $\bigcup_n M_n$ is dense in L^1.

Note that $M_n \subset M$, so $\bigcup_n M_n \subset M$ and M is dense in L^1.

It thus suffices to prove M is closed.

Let $Y_j \in M$ and $Y_j \to X$ in L^1. Need to show $X \in M$, i.e., if $X_n = E[X | F_n]$ then $X_n \to X$ a.s.

By Doob's inequality and Jensen,

$$P[\sup_{1 \le n \le N} |X_n| \ge e] \le \frac{1}{e} E[|X_N|] \le \frac{1}{e} E[|X|].$$

Claim: $\limsup X_n = \liminf X_n$ a.s. (Since $X_n \to X$ in L^1 this implies $X_n \to X$ a.s.)

Let $Y_{n,j} = E[Y_j | \mathcal{F}_n]$. Write

$$X = Y_j - (X - Y_j)$$
$$X_n = Y_{n,j} - (X_n - Y_{n,j})$$

$$\Rightarrow \limsup_n X_n - \liminf_n X_n \le \limsup_n Y_{n,j} - \liminf_n Y_{n,j}$$
$$+ \limsup_n (X_n - Y_{n,j}) - \liminf_n (X_n - Y_{n,j})$$
$$\overset{Y_j \in M}{=} \limsup_n (X_n - Y_{n,j}) - \liminf_n (X_n - Y_{n,j})$$
$$\le 2 \sup |X_n - Y_{n,j}|$$

$$\Rightarrow P[\limsup X_n - \liminf X_n \ge \varepsilon] \le P[\sup_n |X_n - Y_{n,j}| \ge \tfrac{\varepsilon}{2}]$$
$$\overset{\text{Doob}}{\le} \frac{2}{\varepsilon} E[|X - Y_j|] \longrightarrow 0 .$$

Thm. If X_n is a uniformly integrable martingale, i.e.

$$\sup_n E[|X_n| \mathbb{1}_{|X_n|>t}] \to 0 \quad \text{as } t \to \infty,$$

then there is X s.t. $X_n = E[X | \mathcal{F}_n]$.

Proof. Uniform integrability implies weak compactness in L^1. The rest is the same.

6.4. Doob decomposition

<u>Doob decomposition theorem</u> Let (X_n) be a sub-martingale. Then $X_n = M_n + A_n$ where

(i) M_n is a martingale

(ii) $A_{n+1} \geq A_n$ a.e. $\forall n$, $A_1 = 0$

(iii) A_n is $\mathcal{F}_{n-1}$ measurable $\forall n \geq 2$

The decomposition is unique.

<u>Proof.</u> The condition implies

$$A_n - A_{n-1} = X_n - X_{n-1} + M_{n-1} - M_n$$

$$\Rightarrow A_n - A_{n-1} = E[X_n - X_{n-1} \mid \mathcal{F}_{n-1}] \geq 0$$

$$\Leftrightarrow A_n = A_{n-1} + E[X_n - X_{n-1} \mid \mathcal{F}_{n-1}] \geq A_{n-1}$$

Thus the decomposition is unique and $A_n \geq A_{n-1}$ if (X_n) is a submartingale.

<u>Rk.</u> Such a decomposition without monotonicity holds for any adapted process (X_n).

Fact. Let (X_n) be a nonnegative martingale. Then (X_n) is bounded in L^1:

$$E[|X_n|] = E[X_n] = E[X_1]$$

Thm. Let (X_n) be an L^1 bounded martingale. Then there are nonnegative martingales (Y_n) and (Z_n) s.t. $X_n = Y_n - Z_n$.

Proof. For $n \geq j$ define

$$Y_{j,n} = E[|X_n| \mid \mathcal{F}_j].$$

Since $(|X_n|)$ is a submartingale,

$$Y_{j,n+1} - Y_{j,n} = E[|X_{n+1}| - |X_n| \mid \mathcal{F}_j]$$
$$= E[E[|X_{n+1}| - |X_n| \mid \mathcal{F}_n] \mid \mathcal{F}_j] \geq 0 \text{ a.s.}$$

Since $Y_{j,n} \geq 0$ and $E[Y_{j,n}] = E[|X_n|] \leq E[|X_1|]$, there is $Y_j \in L^1$ s.t. $Y_{n,j} \xrightarrow{n\to\infty} Y_j$ in L^1 (by mon. conv.).

Limits of martingales are martingales and $Y_{j,n}$ is a martingale for $j \leq n$. Thus Y_j is a mart.

$$Y_j + X_j = \lim_{n\to\infty} E[|X_n| + X_n \mid \mathcal{F}_j] \geq 0$$

Thus $X_j = (X_j + Y_j) - Y_j$ with $X_j + Y_j$ and Y_j nonneg.

6.5. Optional stopping

Recall the defn. of stopping time w.r.t. $(\mathcal{F}_n)$.

<u>Example</u>. Let τ be a stopping time and f an increasing function s.t. $f(n) \geq n$. Then $f(\tau)$ is again a stopping time.

<u>Example</u> Let τ_1 and τ_2 be stopping times. Then $\tau_1 \vee \tau_2 = \max\{\tau_1, \tau_2\}$ and $\tau_1 \wedge \tau_2 = \min\{\tau_1, \tau_2\}$ are stopping times. Thus $\tau_n = \tau \wedge n$ is a bounded stopping time such that $\tau_n \to \tau$.

<u>Optional stopping theorem</u>. Let (X_n) be a martingale. Let $0 \leq \tau_1 \leq \tau_2 \leq N$ be two <u>bounded</u> stopping times. Then
$$E[X_{\tau_2} \mid \mathcal{F}_{\tau_1}] = X_{\tau_1} \quad \text{a.s.}$$

<u>Proof</u>. Since $\mathcal{F}_{\tau_1} \subset \mathcal{F}_{\tau_2} \subset \mathcal{F}_N$, it suffices to show
$$E[X_k \mid \mathcal{F}_\tau] = X_\tau$$
if τ is a stopping time s.t. $\tau \leq k$. Indeed then
$$\underbrace{E[X_k \mid \mathcal{F}_{\tau_1}]}_{X_{\tau_1}} = E[\underbrace{E[X_k \mid \mathcal{F}_{\tau_2}]}_{X_{\tau_2}} \mid \mathcal{F}_{\tau_1}].$$

To show the claim, let $A \in \mathcal{F}_\tau$. Define

$$E_j = \{\tau = j\} \quad \Rightarrow \quad \Omega = \bigcup_{j=1}^{k} E_j .$$

Since $A \cap E_j \in \mathcal{F}_j$ the martingale property shows

$$\int_{A \cap E_j} X_k \, dP = \int_{A \cap E_j} X_j \, dP = \int_{A \cap E_j} X_\tau \, dP$$

$$\Rightarrow \int_A X_k \, dP = \int_A X_\tau \, dP \qquad \forall A \in \mathcal{F}_\tau .$$

<u>Cor.</u> The optional stopping theorem extends to sub- and supermartingales (by Doob decomposition).

<u>Cor.</u> If (X_n) is a martingale and τ a <u>bounded</u> stopping time, then

$$E[X_\tau] = E[X_0]$$

<u>Ex.</u> Boundedness is important: if ξ_i are i.i.d., uniform on ± 1, then $X_n = \xi_1 + \dots + \xi_n$, $X_0 = 0$, is a martingale, $\tau = \inf\{n \geq 0 : X_n = 1\}$ is stopping time, $P[\tau < \infty] = 1$, but $E[X_\tau] = 1 \neq 0 = E[X_0]$.

Rk. If τ is an unbounded stopping time, can apply OST to $\tau \wedge k$.

$$\Rightarrow \quad E[X_{\tau\wedge k}] = E[X_0].$$

Since $\tau\wedge k \to \tau$ a.s., $X_{\tau\wedge k} \to X_\tau$ a.s., and to pass to the limit it suffices to show that $(X_{\tau\wedge k})_k$ is uniformly integrable. This is the case if, e.g.,

$$S = \sup_{0\le n\le \tau} |X_n| \in L^1$$

Indeed, then $|X_{\tau\wedge k}| \le S$. $\quad\|$

Defn. Given a random process $(X_n)_n$ and a stopping time τ the stopped process X^τ is defined by $X^\tau_n = X_{n\wedge\tau}$.

Thm. Given an adapted integrable process (X_n), i.e., $X_n \in \mathcal{F}_n$ and $E[|X_n|] < \infty$ for all n, TFAE:

(a) X is a martingale

(b) for all bounded stopping times τ and all stopping times σ,

$$E[X_\tau \mid \mathcal{F}_\sigma] = X_{\tau\wedge\sigma} \quad \text{a.s.}$$

(c) for all stopping times τ, X^τ is a martingale

(d) for all bounded stopping times $\sigma \le \tau$:

$$E[X_\tau] = E[X_\sigma]$$

<u>Proof</u>. (a) $\Rightarrow$ (b) is the OST applied with $\tau_1 = \tau$ and $\tau_2 = \tau \wedge \sigma$. Both are bounded, so

$$E[X_\tau \mid \mathcal{F}_\sigma] = E[X_\tau \mid \mathcal{F}_{\tau\wedge\sigma}] = X_{\tau\wedge\sigma}.$$

(b) with $\sigma = n \ge m$ and τ replaced by $\tau \wedge m$ implies (c):

$$E[X^\tau_m \mid \mathcal{F}_n] = E[X_{\tau\wedge m} \mid \mathcal{F}_n] = X_{\tau\wedge m}$$

(d) also follows from (b) or OST.

(d) $\Rightarrow$ (a) Let $A \in \mathcal{F}_m$ and $n \ge m$.

Then $\tau = m \mathbb{1}_A + n \mathbb{1}_{A^c}$ is a stopping time, $\tau \le n$.

$$\Rightarrow E[X_n \mathbb{1}_A] - E[X_m \mathbb{1}_A] = E[X_n] - E[X_\tau] \underset{(d)}{=} 0$$

$$\Rightarrow E[X_n \mid \mathcal{F}_m] = X_m \text{ a.s.}$$

__Cor.__ Let X be a __uniformly integrable__ martingale, and let τ, σ be any stopping times (not bounded). Then

$$E[X_\tau] = E[X_0]$$

and

$$E[X_\tau | \mathcal{F}_\sigma] = X_{\tau \wedge \sigma}.$$

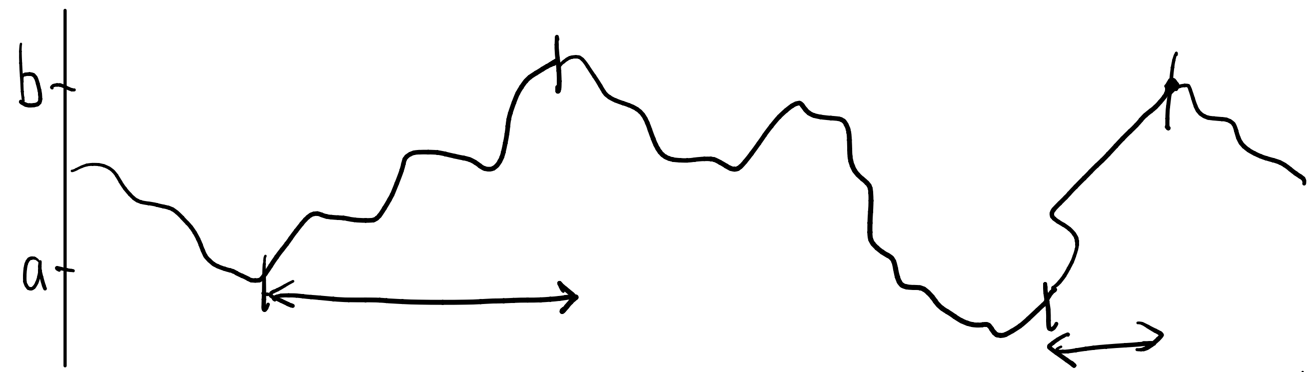
__Proof.__ Similar to earlier remark.

6.6. Upcrossing inequality

Let (X_n) be a random process, $a<b$.

An upcrossing of $[a,b]$ is an interval $\{j, j+1, \dots, k\}$ s.t. $X_j < a$ and $X_k > b$.

Let $U_n(a,b)$ be the number of completed upcrossing up to time n.



Upcrossing inequality. Let $X_1, \dots, X_n$ be a martingale. Then

$$\mathbb{E}[U_n(a,b)] \le \frac{1}{b-a} \mathbb{E}[(a-X_n)_+] \le \frac{1}{b-a}\left(|a| + \mathbb{E}[|X_n|]\right).$$

Proof. Define
$$\tau_1 = n \wedge \inf\{k : X_k \le a\}$$
$$\tau_2 = n \wedge \inf\{k \ge \tau_1 : X_k \ge b\}$$
$$\tau_3 = n \wedge \inf\{k \ge \tau_2 : X_k \le a\}$$
$$\vdots$$

Since $\tau_k \ge \tau_{k-1}+1$ and $\inf \emptyset = \infty$, $\tau_n = n$.

Define

$$D = \sum_{j=1}^{n} (X_{\tau_{2j}} - X_{\tau_{2j-1}}).$$

Each if $\tau_{2j} < n$, the interval $[\tau_{2j-1}, \tau_{2j}]$ corresponds to an upcrossing, $X_{\tau_{2j}} - X_{\tau_{2j-1}} \geq b-a$, and there are $U_n(a,b)$ upcrossings.

$$\Rightarrow D \geq (b-a)\, U(a,b) + R_n$$

$$R_n \geq \begin{cases} 0 & \text{if } \tau_{2\ell} < n = \tau_{2\ell+1} \quad \text{(incomplete downcr.)} \\ (X_n - a) & \text{if } \tau_{2\ell-1} < n = \tau_{2\ell} \quad \text{(incomplete upcross)} \end{cases}$$

By OST, $E[D] = 0$. Thus

$$E[U_n(a,b)] \leq \frac{1}{b-a} E[-R_n] \leq \frac{1}{b-a} E[(a - X_n)_+].$$

<u>Cor.</u> Let (X_n) be a martingale bounded in L^1. Then X_n has a limit a.s. and the limit is in L^1.

<u>Proof.</u> The number of upcrossings of $[a,b]$ is finite a.s. By Doob's ineq., $|X_n|$ is bounded almost surely:

$$P[\sup_n |X_n| > \ell] \leq \frac{1}{\ell} \sup_n E[X_n]$$

$$\overset{\ell \to \infty}{\Rightarrow} P[\sup_n |X_n| = \infty] = 0.$$

Let
$$\Omega_0 = \{\sup_n |X_n| < \infty\} \cap \left(\bigcap_{\substack{a<b \\ a,b\in\mathbb{Q}}} \{U(a,b) < \infty\}\right)$$
$\Rightarrow P(\Omega_0) = 1$

and X_n converges on Ω_0 (otherwise, $\limsup_n X_n > \liminf_n X_n$ and there would be an interval with infinitely many upcrossings).

Let $X_\infty = \lim_{n\to\infty} X_n \mathbb{1}_{\Omega_0}$. Then Fatou gives
$$E[|X_\infty|] = E[\liminf_n |X_n| \mathbb{1}_{\Omega_0}] \le \liminf_n E[|X_n|].$$
so X_∞ is integrable.

<u>Defn</u>. $(\hat{\mathcal{F}}_n)_{n\ge0}$ is a <u>backward filtration</u> if
$$\mathcal{F} \supseteq \hat{\mathcal{F}}_n \supseteq \hat{\mathcal{F}}_{n+1}, \qquad \hat{\mathcal{F}}_\infty = \bigcap_{n\ge0} \hat{\mathcal{F}}_n$$
Alternatively we can consider $(\mathcal{F}_n)_{n\le0}$, $\mathcal{F}_n = \hat{\mathcal{F}}_{-n}$.

Then $(\hat{X}_n)_{n\ge0}$ or $(X_n)_{n\le0}$ given by $X_n = \hat{X}_{-n}$ is a backward martingale if (X_n) satisfies the usual martingale defn: $X_n \in L^1$ for all n and
$$X_n = E[X_{n+1} | \mathcal{F}_n]$$

Backward martingale convergence thm. Let $(X_n)_{n\le 0}$ be a backwards martingale. Then X_n converges a.s. and in L^1 as $n\to-\infty$.

Rk. L^1 convergence is automatic.

Proof. Let $U_n(a,b)$ denote the number of upcrossings of $[a,b]$ by $X_{-n},\dots,X_0$. Then

$$E[U_n(a,b)] \le \frac{1}{b-a} E[(a-X_0)^+].$$

$\Rightarrow E[U(a,b)] < \infty$ by monotone convergence.

The rest of the proof is as in the almost sure martingale convergence theorem.

Note $X_n = E[X_0 \mid \mathcal{F}_n]$ so (X_n) is uniformly int. Thus $X_n \to X_{-\infty}$ as $n\to-\infty$ also in L^1.

6.7. Martingales and Markov chains

Thm. Let $(X_n)_{n\geq 0}$ be an adapted process. Then (X_n) is a Markov chain with transition prob. π iff

$$M_n^f = f(X_n) - f(X_0) - \sum_{k=0}^{n-1}\Big(\underbrace{\pi f(X_k)}_{\int \pi(X_k, dy) f(y)} - f(X_k)\Big)$$

is a martingale for all bounded cont. f.

Cor. If f is bd. harmonic ($\pi f = f$) then $f(X_n)$ is a martingale as we have seen before.

Proof. (X_n) MC $\Rightarrow$ $(M_n^f)_n$ martingale: similar to argument we have seen if f is harmonic.

In the other direction,

$$E[f(X_n) \mid \mathcal{F}_{n-1}] = \underbrace{E[M_n^f \mid \mathcal{F}_{n-1}]}_{=} + f(X_0) + \sum_{k=0}^{n-1}(\pi f(X_k) - f(X_k))$$

$$M_{n-1}^f = f(X_{n-1}) - f(X_0) - \sum_{k=0}^{n-2}(\cdots)$$

$$= \cancel{f(X_{n-1})} + \pi f(X_{n-1}) - \cancel{f(X_{n-1})}$$

$$= \pi f(X_{n-1}).$$

__Defn.__ Given a measurable set A in the state space, the __hitting time__ of A is

$$\tau_A = \inf\{n \geq 0 : X_n \in A\}.$$

The function

$$u_A(x) = P_x[\tau_A < \infty]$$

is the __exit probability__ from A^c.

__Fact.__ For $x \in A$, $u_A(x) = 1$, and for $x \notin A$,

$$u_A(x) = \pi(x, A) + \int_{A^c} \pi(x, dy)\, u_A(y)$$

$$= \int_X \pi(x, dy)\, u_A(y)$$

Thus u_A solves the Dirichlet problem

$$(*) \quad \begin{cases} (\pi - I)v = 0 & \text{on } A^c \\ v = 1 & \text{on } A \end{cases}$$

__Thm.__ $u_A(x) = P_x[\tau_A < \infty]$ is the smallest solution to $(*)$

Proof. Let V be a nonneg. soln to (*).
Define $W = \min\{V, 1\}$. Then $0 \le W \le 1$ and $W(x) = 1$ for $x \in A$. For $x \notin A$,

$$\pi W(x) = \int \pi(x,dy) W(y) \le \int \pi(x,dy) V(y) = V(x)$$

Since also $\pi W(x) \le 1$, hence for $x \notin A$,

$$\pi W(x) \le W(x) = \min\{V(x), 1\}$$

This also holds for $x \in A$. Thus $\pi W(x) \le W(x)$ for all $x \in X$.

Thus $(W(X_n))_n$ is a supermartingale. Therefore, for any bounded stopping time τ,

$$E_x[W(X_\tau)] \le E_x[W(X_0)] = W(x).$$

We would like to take $\tau = \tau_A$ but τ_A is not bounded. Thus take $\tau = \tau_A \wedge N$. Then

$$E_x[W(X_{\tau_A \wedge N})] \le E_x[W(X_0)] = W(x).$$

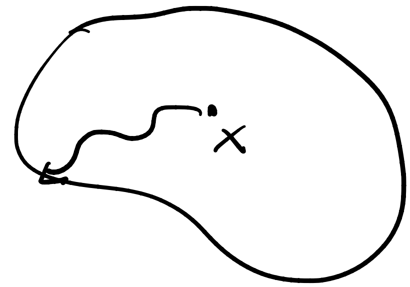
On $\{\tau_A < \infty\}$, $\tau \wedge N \to \tau$ and $W(X_{\tau \wedge N}) \to W(X_\tau) = 1$ as $N \to \infty$.

$$\Rightarrow W(x) \ge \limsup_{N\to\infty} E[W(X_{\tau_A \wedge N}) 1_{\tau_A < \infty}] = P[\tau_A < \infty].$$

$$\Rightarrow V(x) \ge W(x) \ge U_A(x).$$

<u>Prop</u>. For x s.t. $U_A(x)=1$, and any bounded soln to

$$(**)\begin{cases}(\Pi-1)V=0 & \text{on } A^c\\ V=f & \text{on } A\end{cases}$$



it follows that

$$V(x)=E_x[V(\tau_A)].$$

<u>Proof</u>. Let V be a bd. soln to $(**)$ and $h=(\Pi-1)V$

$\Rightarrow V(X_n)-V(X_0)-\sum_{j=1}^{n} h(X_{j-1})$ is a martingale.

Since $h(x)=0$ for $x\notin A$, $h(X_{j-1})=0$ for $j\leq\tau_A$,

$$V(x)\overset{OST}{=}E_x[V(X_{\tau_A\wedge N})].$$

If x is s.t. $U_A(x)=P_x[\tau_A<\infty]=1$, then can take $N\to\infty$ to get

$$V(x)\overset{DCT}{=}E_x[V(X_{\tau_A})].$$

7. Stationary processes

7.1. Ergodic theorems

Defn. A sequence of random variables $(\xi_n)_{n\in\mathbb{Z}}$ is a stationary stochastic process if the joint distr. of $(\xi_{n_1},\dots,\xi_{n_k})$ and $(\xi_{n+n_1},\dots,\xi_{n+n_k})$ are the same for all $k\geq 1$ and $n, n_1,\dots,n_k\in\mathbb{Z}$.

Examples. I.i.d. sequences, Markov chains starting from invariant measure extended to negative n.

In general, the ξ_n can take values in a measurable space $(X,\mathcal{B})$. We assume $(X,\mathcal{B})$ is such that Kolmogorov's consistency theorem applies. Then there is a measure P on $\Omega = X^{\mathbb{Z}}$ with the same finite-dimensional distributions. Define

$$T:\Omega\to\Omega,\quad (T\omega)_n = \omega_{n+1}.$$

Stationarity now means $T_*P = P$.

Defn. Given a prob. space $(\Omega,\mathcal{F},P)$, $T:\Omega\to\Omega$ is a measure preserving transformation if $T_*P=P$. Given $T:\Omega\to\Omega$, a measure is T-invariant if $T_*P=P$.

From now assume $(\Omega, \mathcal{F}, P)$ is a probability space, $T : \Omega \to \Omega$ is an invertible P-preserving map.
For any measurable map $\xi : (\Omega, \mathcal{F}) \to (X, \mathcal{B})$,

$$\xi_n(\omega) = \xi(T^n \omega)$$

then defines a stationary process.

For $f : \Omega \to \mathbb{R}$ define

$$Uf(\omega) = f(T\omega).$$

<u>Fact</u>. U is an isometry on L^p, i.e.

$$\|Uf\|_{L^p} = \left(\int |f(T\omega)|^p \, dP\right)^{1/p} = \left(\int |f(\omega)|^p \, dP\right)^{1/p} = \|f\|_{L^p},$$

and

$$\int Uf \, dP = \int f \, dP.$$

Moreover, U is unitary on L^2 with inverse

$$U^{-1} f(\omega) = \int f(T^{-1}\omega) \, dP.$$

<u>Mean ergodic theorem</u>. Let H be a Hilbert space and U unitary on H. Then for any $f \in H$,

$$\frac{1}{n}\left(f + Uf + \cdots + U^{n-1} f\right) \longrightarrow \pi(f) \quad \text{in } H$$

where $\pi : H \to H_0 = \{f \in H : Uf = f\}$ is the orth. proj.

<u>Proof</u>. H_0 is a closed subspace. Define

$$H_0^\perp = \{g : (g,f) = 0 \ \ \forall f : U^* f = f\} = \ker((I-U)^*)^\perp$$
$$= \overline{\operatorname{im}(I-U)}$$

$$Uf = f \Leftrightarrow U^{-1}f = U^* f = f$$

where we used that $(\ker T^*)^\perp = \overline{\operatorname{im} T}$.

Indeed, if $f \in \ker T^*$ then $(f, Tg) = (T^* f, g) = 0$.
$\Rightarrow \operatorname{im} T \subset (\ker T^*)^\perp \Rightarrow \overline{\operatorname{im} T} \subset (\ker T^*)^\perp$

If $f \in (\operatorname{im} T)^\perp \Rightarrow (f, Tg) = 0 \ \forall g \Rightarrow (T^* f, g) = 0 \ \forall g$
$\Rightarrow (\operatorname{im} T)^\perp \subset \ker T^*$

Since $H = (\overline{\operatorname{im} T}) \oplus (\operatorname{im} T)^\perp = (\ker T^*) \oplus (\ker T^*)^\perp$ one has $\overline{\operatorname{im} T} = (\ker T^*)^\perp$.

Let $A_n f = \frac{1}{n}(f + Uf + \cdots + U^{n-1} f)$

$\Rightarrow \|A_n f\| \leq \|f\| \quad \forall f \in H$
$\quad A_n f = f \quad \forall f \in H_0$

$\Rightarrow A_n f \longrightarrow \pi(f) = f \quad \forall f \in H_0$

If $f = g - Ug$ then $A_n f = \frac{1}{n}(g - U^n g)$

$\Rightarrow \|A_n f\| \leq 2\frac{\|g\|}{n} \longrightarrow 0$

$\Rightarrow A_n f \to 0 = \pi(f) \quad \forall f \in H_0^\perp = \overline{\mathrm{im}(I-U)H}$.

Altogether, $A_n f \to \pi(f) \ \forall f \in H = H_0 \oplus H_0^\perp$

Defn. An event A is invariant if $T^{-1}(A) = A$ and quasi-invariant if $P(A \Delta T^{-1}(A)) = 0$. Let

$$\mathcal{I} = \{A \in \mathcal{F} : T^{-1}(A) = A\}$$

be the σ-field of invariant sets.

Fact $\pi(f) = E[f \mid \mathcal{I}]$.

Defn. P is ergodic for T if for every $A \in \mathcal{I}$,

$$P(A) \in \{0,1\}.$$

Fact. If P is ergodic then $E[f \mid \mathcal{I}] = E[f]$.

Fact. Any product measure is ergodic.

Proof. This follows from the Kolmogorov 0-1 law. Let $A \in \mathcal{I}$. There are $A_n \in X^{\mathbb{Z}}$ depending only on coordinates $[-n,n]$ s.t. $P(A \Delta A_n) \le \frac{1}{n}$. Since $A \in \mathcal{I}$, $P(A \Delta T^{\pm 2n} A_n) \le \frac{1}{n}$ as well. Thus A is in the tail σ-field, so Kolmogorov's 0-1 law implies $P(A) \in \{0,1\}$.

<u>Cor</u>. Let $f \in L^p(P)$, $1 \le p < \infty$. Then

$$\frac{1}{n}\left(f + \cdots + U^{n-1} f\right) \longrightarrow E[f \mid \mathcal{I}] \quad \text{in } L^p.$$

<u>Proof</u>. Let $f \in L^\infty$. Then $A_n f \to \pi f$ in L^2 and thus in L^p, $1 \le p < \infty$. The extension to $f \in L^p$ follows by a limiting argument, using that A_n and $E[\cdot \mid \mathcal{I}]$ are bounded on L^p.

<u>Maximal Ergodic Lemma</u> Let $f \in L^1(P)$, and set

$$E_n^0 = \{ \omega : \sup_{1 \le j \le n} [\underbrace{f(\omega) + \cdots + f(T^{j-1}\omega)}_{S_j f(\omega)}] \ge 0 \}.$$

Then $\displaystyle\int_{E_n^0} f \, dP \ge 0$.

<u>Proof.</u> Let $S_n^* = \max_{0 \le j \le n} S_j$. Then

$$S_j = f + S_{j-1} \circ T \Rightarrow S_n^* = f + (S_{n-1}^* \circ T) \vee 0.$$

Since $S_n^* \ge 0$ on E_n^0,

$$f = S_n^* - (S_{n-1}^* \circ T) \vee 0 = (S_n^* \vee 0) - (S_{n-1}^* \circ T) \vee 0.$$

$$\Rightarrow \int_{E_n^0} f \, dP = \int_{E_n^0} [(S_n^* \vee 0) - (S_{n-1}^* \circ T) \vee 0] \, dP$$

$$\ge \int_{E_n^0} [S_n^* \vee 0 - (S_n^* \circ T) \vee 0] \, dP$$

$$= \int_{E_n^0} (S_n^* \vee 0) \, dP - \underbrace{\int_{TE_n^0} (S_n^* \vee 0) \, dP}_{\le \int_{S_n^* > 0} (S_n^* \vee 0) \, dP = \int_{E_n^0} (S_n^* \vee 0) \, dP}$$

$$\ge 0$$

Maximal ineq. Let $f \in L^1(P)$, $\ell > 0$. Then

$$P\Big[\underbrace{\{\omega : \sup_{1\le j\le n} |A_j f(\omega)| \ge \ell\}}_{E_n}\Big] \le \frac{1}{\ell}\int_{E_n} |f|\, dP.$$

Proof. WLOG $f \ge 0$. Applying the Maximal Ergodic Lemma to $f-\ell$ gives $E_n^0 \rightsquigarrow E_n$,

$$\int_{E_n} [f-\ell]\, dP \ge 0 \iff \int_{E_n} f\, dP \ge \ell P[E_n].$$

$$\Rightarrow P[E_n] \le \frac{1}{\ell}\int_{E_n} f\, dP.$$

Almost sure ergodic thm. Let $f \in L^1(P)$. Then

$$\frac{f(\omega) + \cdots + f(T^{n-1}\omega)}{n} \longrightarrow g(\omega) \quad \text{a.s.}$$

Sketch. Convergence holds if $f \in H_0$ and if $f = g - Ug$ with $g \in L^\infty$ as then

$$A_n f = \frac{1}{n}\big(g(\omega) - U^n g(\omega)\big) \le \frac{2\|g\|_{L^\infty}}{n} \to 0.$$

The functions $f_1 + f_2$, $f_1 \in H_0$, $f_2 = g - Ug$, $g \in L^\infty$ are dense in $L^1(P)$. It remains to show the space for which a.s. holds is closed. Similar to mart.

7.2. Structure of invariant measures

<u>Defn.</u> Let $\mathcal{M}$ denote the set of T-invariant prob. meas. on $(\Omega, \mathcal{F})$ and $\mathcal{M}_e$ the set of ergodic ones.

<u>Fact.</u> $\mathcal{M}$ is convex (or empty), i.e., if $P, Q \in \mathcal{M}$ then

$$tP + (1-t)Q \in \mathcal{M} \quad \forall t \in [0,1].$$

<u>Thm.</u> $P \in \mathcal{M}$ is ergodic iff P is an extreme point of $\mathcal{M}$, i.e., it cannot be written as a nontrivial convex combination of elements of $\mathcal{M}$.

<u>Proof.</u> Suppose $P = tP_1 + (1-t)P_2$, $P_1 \neq P_2$, $t \in (0,1)$, i.e. P is not extremal. To show P cannot be ergodic, assume it is: for every $A \in \mathcal{I}$,

$$P(A) = 0 \text{ resp. } 1 \quad \Rightarrow P_1(A) = P_2(A) = 0 \text{ resp. } 1.$$

$\Rightarrow P_1 = P_2$ on $\mathcal{I}$.

Claim: Since P_1 and P_2 are invariant in fact $P_1 = P_2$ on $\mathcal{F}$.

Given f bounded, $\mathcal{F}$-measurable, by the a.s ergodic theorem there is E s.t. $P_1(E) = P_2(E) = 1$, for $\omega \in E$,

$$h(\omega) = \lim_{n\to\infty} \frac{1}{n}\left(f(\omega) + f(T\omega) + \cdots + f(T^{n-1}\omega)\right) \text{ exists.}$$

Stationarity and BCT:

$$\int f \, dP_1 = \int h \, dP_1$$
$$\int f \, dP_2 = \int h \, dP_2$$

The RHS are equal since h is $\mathcal{I}$ measurable and $P_1 = P_2$ on $\mathcal{I}$. Thus $P_1 = P_2$ on $\mathcal{F}$.

For the other direction, assume P is not ergodic:

$$\exists A \in \mathcal{I} \text{ s.t. } P(A) \in (0,1).$$

Set

$$P_1(E) = \frac{P(E \cap A)}{P(A)}, \quad P_2(E) = \frac{P(E \cap A^c)}{P(A^c)}$$

$$\Rightarrow P = tP_1 + (1-t)P_2, \quad t = P(A) \in (0,1).$$

Thus P is not extremal.

<u>Fact</u>. Two distinct ergodic invariant probability measures are orthogonal on $\mathcal{I}$: There is $E \in \mathcal{I}$ s.t.

$$P_1(E) = 1, \quad P_2(E) = 0.$$

<u>Pf</u>. Otherwise $P_1(E) = P_2(E)$ $\forall E \in \mathcal{I}$. As in last proof then $P_1 = P_2$ on $\mathcal{F}$.

Thm. If $(\Omega, \mathcal{F})$ is a complete separable metric space (also called a Polish space) then every invariant measure P has an ergodic decomposition:

$$P = \int Q \, \mu_P(dQ)$$

where μ_P is a prob. measure on $\mathcal{M}_e$ the set of ergodic probability measures.

Sketch. Denote by P_ω the regular cond. prob. given $\mathcal{I}$:

$$\mathbb{E}[F \mid \mathcal{I}](\omega) = \int F \, dP_\omega$$

$$\Rightarrow \quad P = \int P_\omega \, dP$$

It then suffices to show P_ω is ergodic for P-a.e. ω. Then P_ω defines a map $\Omega \to \mathcal{M}_e$ and μ_P is the image under this map.

One then needs to check P_ω is stationary a.s. and ergodic a.s.

<u>Example</u>. Let $(\Omega, \mathcal{F}) = ([0,1], \mathcal{B})$, $Tx = x + a \mod 1$.

(i) If $a \in \mathbb{R} \setminus \mathbb{Q}$ the uniform measure is the unique invariant measure.

$$\hat{P}(2\pi n) = \int e^{2\pi i n x} \, dP = \int e^{2\pi i n Tx} \, dP = e^{2\pi i n a} \hat{P}(2\pi n)$$

$$a \in \mathbb{R} \setminus \mathbb{Q} \Rightarrow e^{2\pi i n a} = 1 \text{ iff } n = 0$$

$$\Rightarrow \hat{P}(2\pi n) = 0 \text{ iff } n \neq 0$$

$$\Rightarrow P \text{ is uniform}$$

(ii) If $a = \frac{p}{q}$ (with p, q rel. prime) then for any x, $P_x = \text{Uniform}\{x, x+a, \dots, x+(q-1)a\}$ is ergodic.

7.3. Stationary Markov chains

Let $\pi(x, dy)$ be a transition prob. on $(X, \mathcal{B})$. Let $\Omega = X^{\mathbb{Z}}$ with $\mathcal{F}$ the product σ-field. Let

$$\mathcal{F}_n^m = \sigma(X_j : m \le j \le n),$$

$$\mathcal{F}_n = \mathcal{F}_n^{-\infty} = \sigma(X_j : j \le n),$$

$$\mathcal{F}^m = \mathcal{F}_\infty^m = \sigma(X_j : m \le j).$$

<u>Defn</u>. P is a (two-sided) Markov process on Ω if

$$P(X_{n+1} \in A | \mathcal{F}_n) = \pi(X_n, A) \quad \text{a.e. } \forall A \in \mathcal{B},\ n \in \mathbb{Z}.$$

<u>Fact</u>. Let P be a Markov process that is also stationary. Then the distribution μ defined by

$$\mu(A) = P[X_n \in A]$$

is π-invariant: for every $A \in \mathcal{B}$,

$$\mu(A) = \int \pi(x, A)\, \mu(dx).$$

Conversely, given any such μ, one can construct a stationary Markov process.

Defn. Let $\tilde{M}$ be the set of π-invariant prob. distributions, i.e.,

$$\tilde{M} = \left\{ \mu : \mu(A) = \int \pi(x,A)\,\mu(dx) \quad \forall A \in \mathcal{B} \right\}.$$

$$\tilde{M}_e = \left\{ \mu \in \tilde{M} : \begin{array}{l} \mu \text{ is an extremal point in} \\ \text{the convex set } \tilde{M} \end{array} \right\}.$$

Thm. (no proof). Any $\mu \in \tilde{M}_e$ is ergodic.

7.4. Stationary Gaussian processes

Defn. The joint density on $\mathbb{R}^N$ of N Gaussian random variables with mean $\mu = (\mu_i)$ and covariance $C = (C_{ij})$ is given by

$$p(y) = \frac{1}{\sqrt{\det(2\pi C)}} \exp\left[-\frac{1}{2}\left((y-\mu, C^{-1}(y-\mu)\right)\right]$$

provided C is invertible. If C is only positive semidefinite, there is a subspace $S \subset \mathbb{R}^N$ on which C is strictly pos. def. and the Gaussian measure is defined by choosing an orthonormal basis on S and then in the same way.

Such random variables $X = (X_1, \ldots, X_N)$ are also called a Gaussian vector, $X \sim \mathcal{N}(\mu, C)$.

Fact. Let X be a Gaussian vector in $\mathbb{R}^N$ and $T : \mathbb{R}^N \to \mathbb{R}^M$ a linear map. Then TX is a Gaussian vector in $\mathbb{R}^M$ with mean $T\mu$ and covariance TCT^*.

Fact. Any pos. semi-def. matrix C can be written as TT^* for some $T : \mathbb{R}^N \to \mathbb{R}^N$. Hence $X \sim \mathcal{N}(\mu, C)$ if $X = \mu + T\xi$ with $\xi \sim \mathcal{N}(0, id)$. T can be chosen upper triangular.

Defn. $X=(X_n)_{n\in\mathbb{Z}}$ is a stationary Gaussian process if $(X_n)_{|n|\le N}$ is Gaussian for every N and

$$E[X_n]=\mu$$
$$\mathrm{Cov}(X_n,X_m)=\rho_{n-m}=\rho_{m-n}$$

for some $\mu\in\mathbb{R}$, $(\rho_n)_{n\in\mathbb{Z}}$.

Fact. $\sum_{k,j=1}^{n}\rho_{j-k}z_j\bar{z}_k\ge 0\quad\forall n,\ z\in\mathbb{C}^n$

By a version of Bochner's Thm, there is a measure μ on $S^1=[0,2\pi)/\sim$ s.t.

$$\rho_k=\int_0^{2\pi}e^{ik\theta}\,d\mu(\theta).$$

Since $\rho_k=\rho_{-k}$, μ is invariant under $\theta\to 2\pi-\theta$. The measure μ is called the spectral measure of the Gaussian process.

Question. Let $(\xi_n)_{n\in\mathbb{Z}}$ be i.i.d. standard Gaussians. When can a Gaussian process $(X_n)_{n\in\mathbb{Z}}$ be written

$$X_n=\sum_{m=-\infty}^{\infty}a_{n-m}\xi_m,\quad \sum a_n^2<\infty?$$

$$\Rightarrow\ \rho_k=\sum_j a_j a_{j+k}=(a*a)_k \Leftrightarrow d\mu(\theta)=\underbrace{\left|\sum a_j e^{ij\theta}\right|^2}_{=f\in L^1}\frac{d\theta}{2\pi}$$

Thus (X) has such a "moving average representation" if the spectral measure has density in L^1.

<u>Question</u>: When does a Gaussian process have a causal representation

$$X = \sum_{j\geq 0} a_j \xi_{n-j}, \qquad \sum_{j\geq 0} a_j^2 < \infty \ ?$$

We know $\mu = f \frac{d\theta}{2\pi}$ with $f \in L^1$. The condition is equivalent to the existence of $g \in L^2$ s.t. $f = |g|^2$,

$$g = \sum_{j\geq 0} a_j e^{ij\theta}.$$

<u>Thm.</u> Such a g exists iff

$$\int_0^{2\pi} \log f(\theta)\, d\theta > -\infty.$$

8. Further topics

8.1. Gaussian Hilbert spaces

Defn. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Then $S \subset L^2$ is a Gaussian space if S is a closed linear subspace and any $X \in S$ is a Gaussian random variable.

Example. Let $(\Omega, \mathcal{F}, P)$ be a prob. space on which there is a sequence X_i of i.i.d. $\mathcal{N}(0,1)$ r.v. Then (X_i) is orthonormal in L^2, i.e.,

$$E[X_i X_j] = \delta_{ij}$$

and $S = \overline{\mathrm{span}\{X_i\}}$ is a Gaussian space.

Prop. Let H be a separable Hilbert space and $(\Omega, \mathcal{F}, P)$ as in the example. Then there is an isometry $I : H \to S$. Thus

- $I(f) \sim \mathcal{N}(0, (f,f)_H) \quad \forall f \in H$
- $E[I(f)\, I(g)] = (f,g)_H \quad \forall f, g \in H$.

In fact, $I(\alpha f + \beta g) = \alpha I(f) + \beta I(g)$ a.s.

Defn. A white noise on $\mathbb{R}_+$ is an isometry WN from $L^2(\mathbb{R}_+)$ into a Gaussian space. For $A \subset \mathbb{R}_+$ a Borel set, write $WN(A) = WN(1_A)$.

Prop. (i) For $A \in \mathcal{B}(\mathbb{R}_+)$, $|A| < \infty$, $WN(A) \sim \mathcal{N}(0, |A|)$

(ii) For $A, B \in \mathcal{B}(\mathbb{R}_+)$, $A \cap B = \emptyset$, $WN(A)$ and $WN(B)$ are independent.

(iii) For $A = \bigcup_{i=1}^{\infty} A_i$, $A_i \in \mathcal{B}(\mathbb{R}_+)$ disjoint,

$$WN(A) = \sum_{i=1}^{\infty} WN(A_i) \quad \text{in } L^2 \text{ and a.s.}$$

Rk. WN looks like a random measure,

$$A \in \mathcal{B}(\mathbb{R}_+) \mapsto WN(A, \omega),$$

but it is not. (Nullsets depend on the A_i in (iii).)

Defn. For $t \geq 0$, define $B(t) = WN([0,t])$.

Fact. For any $t_1, \dots, t_n$, the vector $(B(t_i))_{i=1}^n$ is Gaussian and

(i) $B_0 = 0$ a.s.

(ii) $E[B_s B_t] = s \wedge t$

(iii) $B_t - B_s$ is independent of $\sigma(B_r, r < s)$ and $\sim \mathcal{N}(0, |t-s|)$.

Proof. For example,

$$E[B(s)B(t)] = E[WN([0,s])\, WN([0,t])]$$
$$= (1_{[0,s]}, 1_{[0,t]}) = s \wedge t.$$

Cor. $E\left[|B(t)-B(s)|^{2p}\right] = (2p-1)!!\, |t-s|^p$

By Kolmogorov's continuity thm (next section), it follows that there is a modification of $B(t)$ which is continuous.

Proof. $B(t)-B(s) \sim \mathcal{N}(0, |t-s|)$. For $X \sim \mathcal{N}(0,1)$,

$$E[X^{2p}] = (2p-1)!!$$

is a direct computation.

8.2. Kolmogorov's continuity theorem

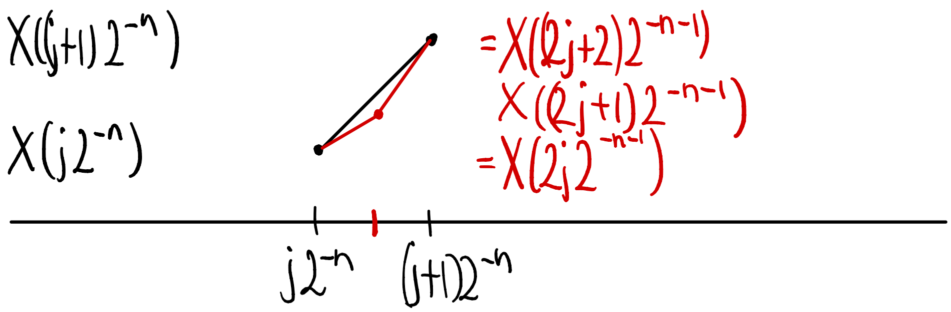
Thm. Let $X(t)$, $t\in[0,1]$ be random variables s.t. for some $\alpha,\beta>0$, $C<\infty$,

$$E\left[|X(t)-X(s)|^{\beta}\right]\le C|t-s|^{1+\alpha}\quad \forall\, 0\le s\le t\le 1.$$

Then there is a cont. version of X with the same finite dimensional distributions.

Proof. Define $X_n:[0,1]\to\mathbb{R}$ by

$$X_n(t)=2^n(t-j2^{-n})X((j+1)2^{-n}) + 2^n((j+1)2^{-n}-t)X(j2^{-n}),\quad t\in[j2^{-n},(j+1)2^{-n}]$$



$$\sup_{t\in[0,1]}|X_{n+1}(t)-X_n(t)| = \sup_{1\le j\le 2^n}\underbrace{\sup_{t\in[(j-1)2^{-n},j2^{-n}]}|X_{n+1}(t)-X_n(t)|}_{|X_{n+1}((2j-1)2^{-n-1})-\underbrace{X_n((2j-1)2^{-n-1})}_{\frac12 X_n((j-1)2^{-n})+\frac12 X_n(j2^{-n})}|}$$

$$\le \sup_{1\le j\le 2^n}\max\left\{|X((2j-1)2^{-n-1})-X((j-1)2^{-n})|,\ |X((2j-1)2^{-n-1})-X(j2^{-n})|\right\}.$$

$$\Rightarrow P\left[\sup_{0\le t\le 1} |X_{n+1}(t) - X_n(t)| \ge 2^{-n\gamma}\right]$$
$$\le 2^{n+1} \sup_j P\left[|X(j2^{-n-1}) - X((j+1)2^{-n-1})| \ge 2^{-n\gamma}\right]$$
$$\le 2^{n+1} \sup_j P\left[|X(j2^{-n-1}) - X((j+1)2^{-n-1})|^{\beta} \ge 2^{-n\gamma\beta}\right]$$
$$\le C 2^{n+1} 2^{-(n+1)(1+\alpha)} 2^{n\beta\gamma}$$

Choose γ s.t. $\beta\gamma \le \alpha$. Then

$$\sum_n P\left[\sup |X_{n+1}(t) - X_n(t)| \ge 2^{-n\gamma}\right] < \infty$$

Thus with prob. 1. X_n converges unif. on $[0,1]$.
The limit X^* takes the same values on dyadic points. It follows that in fact

$$P\left[X(t) = X^*(t)\right] = 1 \qquad \forall t \in [0,1].$$