MSH2 - PROBABILITY THEORY

ANDREW TULLOCH

1. Lecture 1 - Thursday 3 March

Definition 1.1 (σ-field)**.** Let Ω be a non-empty set. Let $\mathcal{F}$ be a collection of subsets of Ω. We call $\mathcal{F}$ a σ-field if

- $\emptyset \in \mathcal{F}$,
- $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$,
- $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$
- If $(A_i) \in \mathcal{F}$, then $\bigcup_{j=1}^{\infty} A_j \in \mathcal{F}$.

Definition 1.2 (Probability measure)**.** Let $\mathbb{P}$ be a function on $\mathcal{F}$ satisfying

- If $A \in \mathcal{F}$ then $\mathbb{P}(A) \geq 0$,
- $P(\Omega) = 1$,
- If $(A_j) \in \mathcal{F}$ and $A_i \cap A_j = \emptyset$, then $\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbb{P}(A_j)$.

Then we call $\mathbb{P}$ a **probability measure** on $\mathcal{F}$.

Definition 1.3 (σ-field generated by a set)**.** If A is a class of sets, then $\sigma(A)$ is the smallest σ-field that contains A.

Example 1.4. For a set B, $\sigma(B) = \{\emptyset, \Omega, B, B^c\}$.

Definition 1.5 (Borel σ-field)**.** Let $\mathcal{B}$ be the class of all **finite** unions of intervals of the form $(a, b]$ on $\mathbb{R}$. The σ-field $\sigma(\mathcal{B})$ is called the **Borel σ-field**.

Note that $\mathcal{B}$ itself is not a σ-field - consider $\bigcup_{j=1}^{\infty}(0, \frac{1}{2} - \frac{1}{j}] = (0, \frac{1}{2}) \notin \mathcal{B}$.

1.1. Constructing extensions of functions to form probability measures.

Lemma 1.6 (Continuity property)**.** *Let $\mathcal{A}$ be a field of subsets of Ω. Assume $\emptyset \in \mathcal{A}$ and that $\mathcal{A}$ is closed under complements and finite unions.*

If $A_j \in \mathcal{F}$ and $A_{j+1} \subset A_j$ with $\bigcap_{j=1}^{\infty} A_j = \emptyset$, then $\lim_{j \to \infty} \mathbb{P}(A_j) = 0$.

Theorem 1.7. *Let $\sigma(\mathcal{A})$ be the σ-field generated by $\mathcal{A}$. If the continuity property holds, then there is a **unique** probability measure on $\sigma(\mathcal{A})$ which is an extension of $\mathbb{P}$, i.e. the measures agree on all elements of $\mathcal{A}$.*

Definition 1.8 (Limits of sets)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and assume $(A_i) \in \mathcal{F}$. Then define $\limsup_{m\to\infty} A_n$ as

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m\geq n} A_m = \overline{\lim} A_n$$

An element $\omega \in \overline{\lim} A_n$ if and only if $\omega \in A_m$ for some $m \geq n$ for all n - that is, ω is in infinitely many of the sets A_m.

Similarly, define $\liminf_{m\to\infty} A_n$ as

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{m\geq n} A_m = \underline{\lim} A_n$$

An element $\omega \in \underline{\lim} A_n$ if and only if ω is in all but a finite number of sets A_m.

Clearly,

$$\underline{\lim} A_n \subseteq \overline{\lim} A_n$$

If $\underline{\lim} A_n$ and $\overline{\lim} A_n$ coincide we write it as $\lim A_n$.

Lemma 1.9. *Assume the continuity property holds. If* $A_n \downarrow A$ *then* $\mathbb{P}(A_n) \downarrow \mathbb{P}(A)$*, and if* $A_n \uparrow A$ *then* $\mathbb{P}(A_n) \uparrow \mathbb{P}(A)$*.*

Proof. If $A_n \downarrow A$, then $A_n \supseteq A_{n+1} \dots$ and $\bigcap_{n=1}^{\infty} A_n = A$. We can write $A_n = (A_n - A) \cup A$. Then we have

$$\mathbb{P}(A_n) = \mathbb{P}(A_n - A) + \mathbb{P}(A)$$
$$\mathbb{P}(A_n) \geq \mathbb{P}(A)$$

By the continuity property, $\mathbb{P}(A_n - A) \to 0$, and so $\mathbb{P}(A_n) \downarrow \mathbb{P}(A)$. ☐

2. Lecture 2 - Thursday 3 March

Theorem 2.1.

$$\mathbb{P}(\underline{\lim} A_n) \leq \underline{\lim} \mathbb{P}(A_n) \leq \overline{\lim} \mathbb{P}(A_n) \leq \mathbb{P}(\overline{\lim} A_n)$$

Proof. We know $A_n \downarrow \underline{\lim} A_n$, and so from Lemma 1.9 we have that a. ☐

Definition 2.2 (Measurable function)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $X : \Omega \to \mathbb{R}$ be real valued function on Ω. Then X is **measurable** with respect to $\mathcal{F}$ if $X^{-1}(B)$ is an element of $\mathcal{F}$ for every B in the Borel σ-field of $\mathbb{R}$.

Definition 2.3 (Random variable)**.** A random variable is a measurable function from Ω to $\mathbb{R}$.

Definition 2.4 (Expectation)**.** If $\int_\Omega |X(\omega)| \, d\mathbb{P} < \infty$ then we can define $\mathbb{E}(X) = \int_\Omega X(\omega) \, d\mathbb{P}$

Definition 2.5 (Distribution)**.** X induces a probability measure $\mathbb{P}_X$ on $\mathbb{R}$

$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(S)), S \in \mathcal{B}$$

P_X is called the **distribution** of X. $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$ is a probability space. The distribution function $F_X(x) = \mathbb{P}(\{\omega : X(\omega) \leq x\}) = \mathbb{P}_X((-\infty, x])$. We have $\mathbb{E}(X) = \int_{\mathbb{R}} x\, dP_X(x) = \int_{\mathbb{R}} x\, dF_X(x)$.

2.1. **Key results from Measure Theory.**

Theorem 2.6 (Monotone convergence theorem)**.** *If* $0 \leq X_n \uparrow X a.s$ *then* $0 \leq \mathbb{E}(X_n) \uparrow \mathbb{E}(X)$ *where* $\mathbb{E}(X)$ *is infinite if* $\mathbb{E}(X_n) \uparrow \infty$.

Theorem 2.7 (Dominated convergence theorem)**.** *If* $\lim X_n = X a.s.$ *and* $|X_n| \leq Y$ *for all* $n \geq 1$, *with* $\mathbb{E}(|Y|) < \infty$ *then* $\lim \mathbb{E}(X_n) = \mathbb{E}(X)$.

Theorem 2.8 (Fatau's Lemma)**.** *If* $X_n \geq Y$ *for all* n *with* $\mathbb{E}(|Y|) < \infty$ *then*

$$\mathbb{E}(\liminf X_n) \leq \liminf \mathbb{E}(X_n)$$

Theorem 2.9 (Composition)**.** *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *and* $(\Omega', \mathcal{F}')$ *be spaces. Let* $\Phi : \Sigma \to \Sigma'$ *be measurable. Define* $\mathbb{P}_\Phi$ *on* $\mathcal{F}$ *by* $\mathbb{P}_\Phi(M) = \mathbb{P}(\Phi^{-1}(M))$. *Let* X' *be a measurable function from* Σ' *to* $\mathbb{R}$. *Then* $X(\omega) = X'(\Phi(\omega))$ *is a measurable function. Then we have*

$$\mathbb{E}(X) = \int_{\Omega'} X'\, d\mathbb{P}_\varphi$$

Proof. Suppose X' is an indictor function for $A \in \mathcal{F}'$. Then

$$\int_{\Omega'} X'\, d\mathbb{P}_\varphi = \int_A d\mathbb{P}_\varphi = \mathbb{P}_\varphi(A) = \mathbb{P}(\varphi^{-1}(A)) = \mathbb{E}(X)$$

So the result is true for simple functions.

Now, suppose $X' \geq 0$. Then there exists a pointwise increasing sequence of simple functions X'_n such that $X'_n \to X'$. By the monotone convergence theorem, we know

$$\lim_{n\to\infty} \int_{\Omega'} X'_n\, d\mathbb{P}_\varphi = \int_{\Omega'} X'\, d\mathbb{P}_\varphi$$

But $X_n(\omega) = X'_n(\Phi(\omega))$ are also simple functions increasing to X. Hence, we know that $\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$. □

3. Lecture 3 - Thursday 10 March

Theorem 3.1 (Jensen's inequality)**.** *Let* $\varphi(x)$ *be a convex function on* $\mathbb{R}$. *Let* X *be a random variable. Assume* $\mathbb{E}(X) < \infty$, $\mathbb{E}(\varphi(X)) < \infty$. *Then*

$$\varphi(E(X)) \leq \mathbb{E}(\varphi(X))$$

Theorem 3.2 (Hölder's inequality)**.** *Let* $1 < p < \infty$ *and* $\frac{1}{p} + \frac{1}{q} = 1$. *Then we have*

$$|\mathbb{E}(XY)| \leq \mathbb{E}(|XY|) \leq (\mathbb{E}(|X|^p))^{1/p} (\mathbb{E}(|Y|^q))^{1/q}$$

If $p = q = 2$ *we obtain the Cauchy-Swartz inequality* $\mathbb{E}(|XY|) \leq \left(\mathbb{E}(|X|^2)\right)^{1/2} \left(\mathbb{E}(|Y|^2)\right)^{1/2}$.

If $Y = 1$ *then* $\mathbb{E}(|X|) \leq (\mathbb{E}(|X|^p))^{1/p}$.

Proof. Let W be a random variable taking values a_1 with probability $1/p$, a_2 with probability $1/q$, with $1/p + 1/q = 1$. Applying Jensen's inequality with $\varphi(x) = -\log(x)$ gives

$$\begin{aligned}
\mathbb{E}(-\log W) &\geq -\log \mathbb{E}(W) \\
\frac{1}{p}(\log a_1) + \frac{1}{q}(-\log a_2) &\geq -\log(\frac{a_1}{p} + \frac{a_2}{q}) \\
-\log(a_1^{1/p} \cdot a_2^{1/q}) &\geq -\log(\frac{a_1}{p} + \frac{a_2}{q}) \\
a_1^{1/p} \cdot a_2^{1/q} &\leq \frac{a_1}{p} + \frac{a_2}{q}
\end{aligned}$$

Where the inequality is trivial if a_1 or a_2 is zero.

Setting $a_1 = |x|^p$ and $a_2 = |y|^q$, we obtain

$$|xy| \leq \frac{|x|^p}{p} + \frac{|y|^q}{q}.$$

Let $x = \frac{X}{(\mathbb{E}(|X|^p))^{1/p}}$ and $y = \frac{Y}{(\mathbb{E}(|Y|^q))^{1/q}}$ or take expectations across the inequality, we obtain

$$\mathbb{E}(|XY|) \leq (\mathbb{E}(|X|^p))^{1/p} (\mathbb{E}(|Y|^q))^{1/q}$$

∎

Example 3.3. If $1 < r < r'$ then $\frac{r'}{r} > 1$. Then

$$\mathbb{E}(|X|^r) \leq (\mathbb{E}((|X|^r)^{r'/r}))^{1/(r'/r)} = (\mathbb{E}(|X|^{r'}))^{r'/r}$$

Theorem 3.4 (Liapounov's inequality)**.**

$$(\mathbb{E}|X|^r)^{1/r} \leq (\mathbb{E}(|X|^{r'}))^{1/r'}$$

Corollary 3.5. *Thus if* $\mathbb{E}(|X|^r) < \infty$ *then* X *has all moments of lower order finite i.e.* $\mathbb{E}(|X|^p) < \infty$ *for all* $1 \leq p \leq r$

Theorem 3.6 (Minkowski's inequality)**.** *If* $p \geq 1$, *then*

$$(\mathbb{E}(|X+Y|^p))^{1/p} \leq (\mathbb{E}(|X|^p))^{1/p} + (\mathbb{E}(|Y|^p))^{1/p}$$

Proof.

$$\begin{aligned}
\mathbb{E}(|X+Y|^p) &\leq \mathbb{E}(|X| \cdot |X+Y|^{p-1} + \mathbb{E}(|Y| \cdot |X+Y|^{p-1}) \\
&= \mathbb{E}(|X|^p)^{1/p}(\mathbb{E}(|X+Y|^{p-1})^q)^{1/q} + \mathbb{E}(|Y|^p)^{1/p}(\mathbb{E}(|X+Y|^{p-1})^q)^{1/q}
\end{aligned}$$

Let $1/p + 1/q = 1$. Then from Hölder,

$$\mathbb{E}(|X+Y|^p) \leq (\mathbb{E}(|X+Y|^p))^1/q \cdot ((\mathbb{E}(|X|^p))^{1/p} + (\mathbb{E}(|Y|^p))^{1/p}$$

and so

$$(\mathbb{E}(|X+Y|^p))^{1/p} \leq (\mathbb{E}(|X|^p))^{1/p} + (\mathbb{E}(|Y|^p))^{1/p}$$

□

3.1. **Modes of Convergence.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X_n(\omega), n \geq 1$ is a sequence of random variables.

Definition 3.7 (Almost surely convergence)**.** We say X_n converges almost surely if

$$\mathbb{P}(\{\omega \mid X_n(\omega) \text{ has a limit}\}) = 1$$

We write $X_n \overset{a.s.}{\to} X$ where X denotes the limiting random variable.

Definition 3.8 (Convergence in probability)**.** X_n converges in probability to X

$$X_n \overset{p}{\to} X$$

if for all $\epsilon > 0$,

$$\mathbb{P}(\{\omega \mid |X_n(\omega) - X(\omega)| > \epsilon\}) \to 0$$

or alternatively,

$$\mathbb{P}(|X_n - X| > \epsilon) \to 0$$

Definition 3.9 (Convergence in mean)**.** X_n converges to X in mean of order p (or in L^p) if

$$\mathbb{E}(|X_n - X|^p) \to 0$$

We write $X_n \overset{L^p}{\to} X$. We note that for convergence of order L^p, we need $\mathbb{E}(|X_n|^p) < \infty$.

Theorem 3.10. *If $X_n \overset{L^p}{\to} X$ then $X_n \overset{p}{\to} X$ for any $p > 0$.*

4. Lecture 4 - Thursday 10 March

Lemma 4.1. *Let $C_1, C_2, \dots$ be sets in $\mathcal{F}$ and $\sum_n \mathbb{P}(C_n) < \infty$. Then $\mathbb{P}(\overline{\lim} C_n) = 0$*

Proof. Since $\overline{\lim} C_n = \bigcap_{n=1}^\infty \bigcup_{m \geq n} C_m$, we have

$$\mathbb{P}(\overline{\lim} C_n) \leq \mathbb{P}(\bigcup_{m \geq n} C_m) \leq \sum_{m \leq n} \mathbb{P}(C_m) \to 0$$

□

Theorem 4.2. *If there exists a sequence of positive constants $\{\epsilon_n\}$ with $\sum_n \epsilon_n < \infty$ and*

$$\sum_n \mathbb{P}(|X_{n+1} - X_n| > \epsilon_n) < \infty$$

then X_n converges almost surely to some limit X.

Proof. Let $A_n = \{|X_{n+1} - X_n| > \epsilon_n$. So from the above Lemma, $\mathbb{P}(\overline{\lim} A_n) = 0$. We also have that $\omega \in \overline{\lim} A_n$ if and only if ω is in infinitely many A_m. For $\omega \notin \overline{\lim} A_n$, then there is a last set containing ω. Define $N(\omega) = n$ if $\omega \in \bigcup_{m \geq n} A_m - \bigcup_{m > n} A_m$, and zero if $\omega \in (\bigcup m \geq 1 A_m)^c$.

For $\omega \notin \overline{\lim} A_n$, we have $\sum_{n=1}^{\infty} X_{n+1}(\omega) - X_n(\omega)$ exists as $\sum_n \epsilon_n < \infty$. Since

$$X_n(\omega) = X_1(\omega) + (X_2(\omega) - X_1(\omega)) + \cdots + (X_n(\omega) - X_{n-1}(\omega))$$

we know $\lim X_n(\omega)$ exists - i.e. $\mathbb{P}(\lim X_n(\omega))$ exists$) = 1$. □

Theorem 4.3. *Every sequence of random variables X_n that converges almost surely converges in probability. Conversely, if $X_n \xrightarrow{p} X$ then there exists a subsequence $\{X_{n_k}\}$ which converges almost surely.*

Proof. Assume $X_n \xrightarrow{a.s.} X$. Let $\epsilon > 0$. Consider $\overline{\lim}\mathbb{P}(|X_n - X| > \epsilon) \leq \mathbb{P}(\limsup\{|X_n - X| > \epsilon\})$ by a previous theorem (Theorem 2 in Lecture Notes). We have

$$\begin{aligned} \limsup\{|X_n - X| > \epsilon\} &= \{\omega \mid |X_n(\omega) - X(\omega)| > \epsilon \text{ infinitely often}\} \\ &\subseteq \{\omega \mid \lim X_n(\omega) \neq X(\omega)\} \end{aligned}$$

Hence, we have

$$\mathbb{P}(\overline{\lim}|X_n - X| > \epsilon) \leq 1 - P(\lim X_n(\omega) = X(\omega)) = 0 \quad \text{as } X_n \xrightarrow{a.s.} X$$

since $\lim \mathbb{P}(|X_n - X| > \epsilon) = 0$.

Conversely, assume $X_n \xrightarrow{p} X$. Given $\epsilon > 0$, consider $\mathbb{P}(|X_n - X_m| > \epsilon) \leq \mathbb{P}(|X - X_m| > \epsilon/2 + \mathbb{P}(|X - X_n| > \epsilon/2)$ (If $|X - X_n| \leq \epsilon/2$ and $|X - X_m| \leq \epsilon/2$, then $|X_n - X_m| \leq \epsilon$ by the triangle inequality). Thus, $\mathbb{P}(|X_m - X_n| > \epsilon) \to 0$ as m and $n \to 0$. Set $n_1 = 1$ and define n_j to be the smallest integer $N > n_{j-1}$ such that

$$\mathbb{P}(|X_r - X_s| > 2^{-j}) < 2^{-1} \quad \text{when } r, s > N$$

Then apply Theorem 4.2, and as

$$\sum_j \mathbb{P}(|X_{n_{j+1}} - X_{n_j}| > 2^{-j}) < \sum 2^{-j} = 1 < \infty$$

we know that X_{n_j} converges almost surely.

□

Example 4.4. We now construct an example where $X_n \xrightarrow{p} 0$ but X_n does not converge almost surely to 0.

Let $\Omega = [0,1], \mathcal{F}$ the Borel σ-field, and $\mathbb{P}$ the Lebesgue measure. Let $\varphi_{kj} = \mathbb{I}_{[j-1/k, j/k]}$ for $j = 1, \ldots, k$ and $k = 1, 2, \ldots$.

Let $X_1 = \varphi_{11}, X_2 = \varphi_{21}, X_3 = \varphi_{22}$, etc. For any $p > 0$,

$$\mathbb{E}(|X_n|^p) = \int X_n \, d\mathbb{P} = [j_n - 1/k_n, j_n/k_n] \to 0$$

and so $X_n \xrightarrow{L^p} 0$.

However, for each $\omega \in \Omega$ and each k there are some j such that $\varphi_{kj}(\omega) = 1$. Thus $X_n(\omega) = 1$ infinitely often. Similarly $X_n(\omega) = 0$ infinitely often. Hence X_n does not converge almost surely to 0.

5. Lecture 5 - Thursday 17 March

Following from the previous lecture, we now modify the examples to show convergence in probability does not imply convergence in L^p even when $\mathbb{E}(|X_n|^p) < \infty$.

From 4.4, replace φ_{kj} by $k^{1/p}\varphi_{kj}$. Then

$$\mathbb{P}(|X_n| > 0) = 1/k_n \to 0$$

as $n \to \infty$. Similar,y

$$\mathbb{E}(|X_n|^p) = (k_n^{1/p})^p \mathbb{P}(X_n \neq 0) = 1$$

and so

$$\lim_{n\to\infty} \mathbb{E}(|X_n|^p) = 1$$

and thus X_n does not converge in L^p to zero. Thus convergence in probability does not imply convergence in L^p.

Next define $X_1 = \varphi_{11}$, $X_n = \varphi_{n1} n^{1/p}$. Then

$$X_n(\omega) \to 0$$

for $\omega > 0$ so $X_n \stackrel{a.s.}{\to} 0$. We also have

$$\mathbb{E}(|X_n|^p) = (n^{1/p})^p \frac{1}{n} = 1$$

and so X_n does not converge in L^p to zero.

Definition 5.1 (Uniform integrability)**.** A sequence $\{X_n\}$ is uniformly integrable if

$$\lim_{y\to\infty} \sup_n \int_{|X_n| \geq y} |X_n| \, d\mathbb{P} = 0$$

Theorem 5.2 (Convergence in probability and uniform integrability imply convergence in L^p)**.** *If $X_n \stackrel{p}{\to} X$ and $\{|X_n|\}$ is uniformly integrable, then $X_n \stackrel{L^p}{\to} X$.*

Definition 5.3 (Independence)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $A_1, A_2, \ldots, A_n \in \mathcal{F}$ The events are said to be independent if

$$\mathbb{P}(A_{i_1}, \ldots, A_{i_k}) = \mathbb{P}(A_{i_1}) \ldots \mathbb{P}(A_{i_k})$$

for all $1 \leq i_1 < \cdots < i_k \leq n$, $k = 2, 3, \ldots, n$.

In the infinite case, let $\{A_\alpha, \alpha \in I\}$, I an index set, is a set of independent events if each finite subset is independent.

Definition 5.4 (Independence of random variables)**.** Let $X_1, \ldots, X_n$ be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. $X_1, \ldots, X_n$ are independent if $A_i = \{X_i \in S_i\}$ are independent for every set of Borel sets, $S_i \in \mathcal{B}$.

Alternatively, let X and Y be random variables. Let $\mathcal{B}_2$ be the Borel σ-field on $\mathbb{R}^2$. $Z(\omega) = (X(\omega), Y(\omega))$ is then a map form Ω to $\mathbb{R}^2$. Z is Borel measurable if

$$Z^{-1}(S) \in \mathcal{F}$$

for all $S \in \mathcal{B}_2$. $\mathbb{P}_{X,Y}$ is the induced measure on B_2, and $F_{X,Y}$ is the joint distribution of (X, Y). Let

$$F_{X,Y}(x, y) = \mathbb{P}_{X,Y}((-\infty, x], (-\infty, y]) = \mathbb{P}(\{\omega : X(\omega) \leq x, Y(\omega) \leq y\})$$

Theorem 5.5. *If X and Y are independent then*

$$F_{X,Y}(x, y) = F_X(x)F_Y(y)$$

Theorem 5.6. *Let X and Y be independent, with $\mathbb{E}(|X|) < \infty$ and $\mathbb{E}(|Y) < \infty$. Then*

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

Proof. Start with simple functions. Then

$$X(\omega) = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}(\omega)$$

with $\{A_i\}$ disjoint. Let

$$Y(\omega) = \sum_{j=1}^{m} b_j \mathbf{1}_{B_j}(\omega)$$

with $\{B_j\}$ disjoint.

Independence implies $\mathbb{P}(A_i B_j) = \mathbb{P}(A_i)\mathbb{P}(B_j)$.

Then

$$\begin{aligned}
\mathbb{E}(XY) &= \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \mathbb{E}(\mathbf{1}_{A_i} \mathbf{1}_{B_j}) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \mathbb{P}(A_i B_j) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \mathbb{P}(A_i)\mathbb{P}(B_j)
\end{aligned}$$

by independence.

Now extend to non-negative random variables X, Y by constructing sequences of simple functions using monotone convergence theorem. Let

$$X_n(\omega) = \frac{i}{2^n} \quad \text{if } \frac{i}{2^n} < X(\omega) \leq \frac{i+1}{2^n}, i = 0, 1, \ldots, n2^n$$

and zero if $X(\omega) > n$.

For simple functions, we have

$$\mathbb{E}(X_n Y_n) = \mathbb{E}(X_n)\mathbb{E}(Y_n)$$

and so by the monotone convergence theorem,

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

□

Theorem 5.7. *Let X and Y be independent random variables. Then*

$$\mathbb{E}(|X+Y|^r) < \infty$$

if and only if

$$\mathbb{E}(|X|^r) < \infty \text{ and } \mathbb{E}(|Y|^r) < \infty$$

for any $r > 0$.

Lemma 5.8 (c_r inequality)**.** *We have*

$$|x+y|^r \leq c_r\left(|x|^r + |y|^r\right)$$

for x, y real, c_r constant, $r \geq 0$.

Proof. If $r = 0$, trivial.

If $r = 1$, we obtain the triangle inequality.

If $r > 1$, we have

$$\begin{aligned} |x+y|^r &\leq [2\max(|x|, |y|)]^r \\ &= 2^r \max(|x|^r, |y|^r) \\ &\leq 2^r(|x|^r, |y|^r) \end{aligned}$$

and setting $c_r = 2^r$ proves for $r > 1$.

If $0 < r < 1$, consider $f(t) = 1 + t^r - (1+t)^r$, with $f(0) = 0$. Differentiating, we have $f'(t) = rt^{r-1} - r(1+t)^{r-1} \geq 0$ for $t > 0$. Thus $f(t)$ is increasing for $t > 0$. Hence

$$\begin{gathered} f(t) > f(0) = 0 \\ 1 + t^r \geq (1+t)^r. \end{gathered}$$

Using $t = \frac{|y|}{|x|}$, we obtain

$$(|x| + |y|)^r \leq |x|^r + |y|^r$$

□

6. Lecture 6 - Thursday 17 March

Lemma 6.1. *For any $\alpha > 0$ and distribution function F,*

$$\int_0^\infty x^\alpha \, dF(x) = \alpha \int_0^\infty x^{\alpha-1}[1 - F(x)] \, dx$$

Proof. Consider. Integrating by parts, we have that this is equal to

$$\begin{aligned}\int_0^b x^\alpha \, dF(x) &= -\int_0^b x^\alpha \, d(1 - F(x)) \\ &= [-x^\alpha](1 - F(x))|_0^b + \int_0^b \alpha x^{\alpha-1}(1 - F(x)) \, dx \\ &= -b^\alpha(1 - F(b)) + \int_0^b \alpha x^{\alpha-1}(1 - F(x)) \, dx\end{aligned}$$

We also have

$$0 \leq b^\alpha(1 - F(b)) \leq \int_b^\infty x^\alpha \, dF(x)$$

If the LHS converges then $\lim_{b\to\infty} \int_0^\infty x^\alpha \, dF(x) \to 0$. Thus the term $b^\alpha(1 - F(b))$ is squeezed to zero.

Conversely,

$$\int_0^b x^\alpha \, dF(x) \leq \int_0^b \alpha x^{\alpha-1}(1 - F(x)) \, dx$$

and so

$$\int_0^\infty \alpha x^{\alpha-1}(1 - F(x)) \, dx < \infty \Rightarrow \int_0^\infty x^\alpha \, dF(x) < \infty.$$

□

Theorem 6.2. *Let X, Y independent and $r > 0$. Then*

$$\mathbb{E}(|X + Y|^r) < \infty \iff \mathbb{E}(|X|^r)\infty, \mathbb{E}(|Y|^r) < \infty$$

Proof. If $\mathbb{E}(|X|^r) < \infty$, $\mathbb{E}(|Y|^r) < \infty$. Then

$$\mathbb{E}(|X + Y|^r) \leq c_r(\mathbb{E}(|X|^r) + \mathbb{E}(|Y|)^r) < \infty$$

Assume $\mathbb{E}(|X + Y|^r) < \infty$. Assume X and Y have median 0 (without loss of generality). Then

$$\mathbb{P}(X \leq 0) \geq \frac{1}{2}, \mathbb{P}(X \geq 0) \geq \frac{1}{2}$$

Similarly for Y.

Now,

$$\begin{aligned}\mathbb{P}(|X| > t) &= P(X < -t) + P(X > t), t > 0\\ &= \frac{P(X < -t, Y \leq 0)}{P(Y \leq 0)} + \frac{P(X > t, Y \geq 0)}{P(Y \geq 0)}\\ &= 2P(X + Y \leq -t) + 2P(X + Y > t)\\ &= 2P(|X + Y| > t)\end{aligned}$$

by independence.

Using the previous lemma, we have

$$\begin{aligned}\mathbb{E}(|X|^r) \int_0^\infty x^r \, dF(x) &= r \int_0^\infty x^{r-1} P(|X| > x) \, dx\\ &\leq 2r \int_0^\infty x^{r-1} P(|X + Y| > x) \, dx\end{aligned}$$

$$= 2r\mathbb{E}(|X + Y|^r).$$

So $\mathbb{E}(|X + Y|^r) < \infty \Rightarrow \mathbb{E}(|X|^r) < \infty$. Similarly for $\mathbb{E}(|Y|^r) < \infty$.

□

Theorem 6.3. *If X and Y are independent with distribution functions F and G respectively, then*

$$\begin{aligned}P(X + Y \leq x) &= \int_{\mathbb{R}} F(x - y) \, dG(y)\\ &= \int_{\mathbb{R}} G(x - y) \, dF(y)\end{aligned}$$

Proof. This is just a simple statement of Fubini's theorem. □

Corollary 6.4. *Suppose that X has an absolutely continuous distribution function*

$$F(x) = \int_{-\infty}^x f(u) \, du$$

for some density function f with $\int_{\mathbb{R}} f(x) \, dx = 1$ and $f \geq 0$.

Let Y be independent of X. Then $X + Y$ has an absolutely continuous distribution with density

$$\int_{\mathbb{R}} f(x - y) \, dG(y)$$

Thus we have

$$\begin{aligned}P(X + Y \leq x) &= \int_{\mathbb{R}} \int_{-\infty}^x f(t - y) \, dt \, dG(y)\\ &= \int_{-\infty}^x \int_{\mathbb{R}} f(t - y) \, dG(y) \, dt\end{aligned}$$

Definition 6.5. A distribution function F that can be represented in the form

$$F(x) = \sum_j b_j \mathbf{1}_{[a_j,\infty]}(x)$$

with a_j real, $b_j \geq 0$, $\sum_{b_j} = 1$ is called **discrete**.

If a distribution function is continuous then it may be:

(1) **Absolutely continuous**, in which case there is a density function $f \geq 0$ such that $F(b) - F(a) = \int_a^b f(u)\,du$. f is called the density.
(2) **Singular**, in which case $F'(x)$ exists and equal zero almost everywhere with respect to the Lebesgue measure (see Chung §1.3)

Theorem 6.6. *Any distribution function F can be written uniquely as a convex combination of a discrete, an absolutely continuous, and a singular distribution. By convex, we mean a linear combination with non-negative coefficients summing to one.*

Theorem 6.7 (Chebyshev's inequality)**.** *Let X be a random variable and g an increasing, non-negative function. If $g(a) > 0$, then*

$$P(X \geq a) \leq \frac{\mathbb{E}(g(X))}{g(a)}.$$

Proof. We have

$$\begin{aligned}\mathbb{E}(g(X)) &= \int_{\mathbb{R}} g(x)\,dF(x)\\ &\geq \int_a^\infty g(x)\,dF(x)\\ &\geq g(a)\int_a^\infty dF(x)\\ &= g(a)P(X \geq a)\end{aligned}$$

□

Corollary 6.8. *Let $g(x) = x^2$. Then*

$$P(|X - \mathbb{E}(X)| > a) \leq \frac{Var(X)}{a^2}$$

Let $g(x) = e^{ax}$. Then

$$P(X \geq a) \leq \frac{\mathbb{E}(e^{cX})}{e^{ca}} = e^{-ca}\mathbb{E}(e^{cX})$$

Let $g(x) = |x|^k, k > 0$. Then

$$P(|X| \geq a) \leq \frac{\mathbb{E}(|X|^k)}{a^k}.$$

7. Lecture 7 - Thursday 24 March

Definition 7.1 (Weak law of large numbers)**.** Let $X_1, \dots, X_n \dots$ be IID random variables with $\mathbb{E}(X_i) = \mu$,$\mathrm{Var}(X_i) = \sigma^2 < \infty$. Then

$$\overline{X}_n \xrightarrow{p} X$$

Proof.

$$\begin{aligned} P(|\overline{X}_n) &\leq \frac{\mathbb{E}(\overline{X}_n - \mu)^2}{\epsilon^2} \\ &= \frac{\sigma^2/n}{\epsilon^2} \to 0 \end{aligned}$$

as $n \to \infty$.

We have

$$\mathbb{E}(|\overline{X}_n - \mu|^2) = \sigma^2/n \to 0$$

and so $\overline{X}_n$ converges to μ in L^2 □

We can relax the assumptions to $E(|X) < \infty$ (no need to have finite variance). See Chung (1974) p.109, Theorem 5.2.2.

Theorem 7.2. *Let X_i be uncorrelated, and $\mathbb{E}(X_i) = \mu_i$, Var$(X_i) = \sigma_i^2 < \infty$ with*

$$\frac{1}{n^2} \sum_{i=1}^{n} \sigma_i^2 \to 0$$

then we have

$$\overline{X}_n - \frac{1}{n} \sum_{i=1}^{n} \mu_i \xrightarrow{p} 0$$

Proof.

$$\begin{aligned} P(|\overline{X}_n - \frac{1}{n} \sum_{i=1}^{n} \mu_i| > \epsilon) &= P(\frac{1}{n} \sum_{i=1}^{n} (X_i - \mu_i)| > \epsilon) \\ &\leq \frac{\mathrm{Var}(\frac{1}{n} \sum_{i=1}^{n} (X_i - \mu_i))}{\epsilon^2} \to 0 \end{aligned}$$

as $\sum_1 n^2 \sum_{i=1}^{n} \sigma_i^2 \to 0$. □

Theorem 7.3 (Borel-Cantelli lemma)**.** *Let $A_1, \dots$ be events in a probability space. Let $B = \limsup A_n = \bigcap_{n \geq 1} \bigcup_{m \geq n} A_m$. Then*

(i) $\sum_n P(A_n) < \infty$ then $P(B) = 0$.

(ii) If A_i are independent and $\sum_n P(A_n) \to \infty$ then $P(B) = 1$.

For (ii) we need independence. Consider $A_i = A$ where $P(A) = \frac{1}{3}$. Then

$$B = \limsup A_n = A$$

and $P(B) = \frac{1}{3}$

Proof. Preliminary lemma - if $0 < x < 1$, then $\log(1-x) < -x$. We can then show that if $\sum_n a_n \to \infty$ then $\prod_n (1 - a_n) \to 0$.

(i)
$$P(B) \leq P(\bigcup_{m \geq n} A_n) \leq \sum_{m \geq n} P(A_m) \to 0$$
and so $P(B) = 0$.

(ii) We will prove $P(\bigcup_{m \geq n} A_m) = 1$ for all n. Take $K > n$. Then
$$\begin{aligned} 1 - P(\bigcup_{m \geq n} A_m) &\leq 1 - P(\bigcup_{m=n}^{K} A_m) \\ &= P((\bigcup_{m=n}^{K} A_n)^c) \\ &= P(\bigcup_{m=n}^{K} A_m^c) \\ &= \prod_{m=n}^{K} (1 - P(A_m)) \quad \text{by independence} \\ &\to 0 \end{aligned}$$
as $\sum_n P(A_n) \to \infty$ as $K \to \infty$. Thus
$$P(\bigcup_{m \geq n} A_m) = 1$$
for all m, and so $P(B) = 1$.

□

Theorem 7.4 (Strong law of large numbers)**.** *Let* $X_1, \dots$ *be IID random variables. Let* $\mathbb{E}(X_1) = \mu$, $\mathbb{E}(X_1^4) < \infty$. *Let* $S_n = \sum_{j=1}^{n} X_j$. *Then*
$$\overline{X}_n = \frac{1}{n} S_n \overset{a.s.}{\to} \mu$$

Proof.
$$\begin{aligned} \mathbb{E}(\sum_{i=1}^{n} (X_i - \mu))^4 &= \sum_{i=1}^{n} E(X_i - \mu)^4 + 6\binom{n}{2}\sigma^4 \\ &= n\mathbb{E}(X_1 - \mu)^4 + 3n(n-1)\sigma^4 \\ &\leq Cn^2. \end{aligned}$$

From Chebyshev, we have

$$\begin{aligned} P(|S_n - \mu n| > \epsilon n) &\leq \frac{E(S_n - \mu n)^4}{)(\epsilon n)^4} \\ &\leq \frac{cn^2}{\epsilon^4 n^4} = \frac{k}{n^2} \end{aligned}$$

and so

$$\sum_n P(|S_n - n\mu| > n\epsilon) < \infty,$$

and so $P(\limsup\{|\frac{S_n}{n} - \mu| > \epsilon\}) = 0$. Letting $A_\epsilon = \{|\frac{S_n}{n} - \mu| > \epsilon\}$. Then

$$\begin{aligned} P(|\frac{S_n}{n} - \mu| \text{ does not converge to zero}) &= P(\bigcup_k A_{1/k}) \\ &\leq \sum_k P(A_{1/k}) \\ &= 0 \end{aligned}$$

by Borel-Cantelli. □

8. Lecture 8 - Thursday 24 March

Let $X_1, \ldots$ be IID random variables with mean μ. Then

$$P(\lim_{n\to\infty} \frac{S_n}{n} = \mu) = 1$$

Conversely, if $\mathbb{E}(|X|)$ does not exist, then

$$P(\limsup |\frac{S_n}{n}| = \infty) = 1$$

Theorem 8.1. *If $E(X^2) < \infty$, and $\mu = 0$ (WLOG),*

$$\begin{aligned} P(|n^{-\alpha} S_n| \geq \epsilon) &\leq \frac{E(S_n^2)}{n^{2\alpha}\epsilon^2} \\ &= n^{1-2\alpha}\sigma^2/\epsilon^2 \to 0 \end{aligned}$$

provided $S \geq \frac{1}{2}, n^{-\alpha} S_n \xrightarrow{p} 0$.

Theorem 8.2 (Hausdorff (1913))**.** $|S_n| = \mathcal{O}(n^{\frac{1}{2}+\epsilon})$ *a.s for any $\epsilon > 0$.*

Assumes $\mathbb{E}(|X_i|^r) < \infty$ for $r = 1, 2, \ldots$.

Proof. Previously, we showed $\mathbb{E}(S_n^4) \leq Cn^2$ for some $C > 0$. Then we can extend this to

$$\mathbb{E}(S_n^{2k}) \leq c_k n^k, k = 1, 2, \ldots$$

Then

$$P(n^{-\alpha}|S_n| > a) \geq \frac{c_k n^k}{(an^\alpha)^{2k}}$$
$$= c_k a^{-2k} n^{k(1-2\alpha)}$$

and so

$$\sum P(n^{-\alpha}|S_n| > a) < \infty$$

if $k(1-2\alpha) > -1$ i.e. $\alpha \geq \frac{1}{2} + \frac{1}{2k}$.

By Borel-Cantelli, $P(|S_n| > an^\alpha \; i.o.) = 0$ if $\alpha > \frac{1}{2} + \frac{1}{2k}$. $\square$

Theorem 8.3 (Hardy and Littlewood (1914))**.** $|S_n| = \mathcal{O}(\sqrt{n \log n})$ *a.s.*

Lemma 8.4. *Suppose* $|X_i| \leq M$ *a.s. (*X_i *is bounded). Then for any* $x \in [0, \frac{2}{M}]$*, we have*

$$\mathbb{E}(e^{xS_n}) \leq \exp[\frac{nx^2\sigma^2}{2}(1 + xM)]$$

Proof. The random variables e^{xX_i} are independent, so $\mathbb{E}(e^{xS_n}) = \left[\mathbb{E}(e^{xX_1})\right]^n$. We can then evaluate

$$\begin{aligned}
\mathbb{E}(e^{xX_1}) &= \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{(xX_1)^k}{k!}\right] \\
&= 1 + 0 + x^2\sigma^2/2 + \mathbb{E}(\sum_{k=3}^{\infty} \frac{(xX_1)^k}{k!}) \\
&\leq 1 + x^2\sigma^2/2 + \sum_{k=3}^{\infty} \frac{x^k M^{k-2}\sigma^2}{k!} \\
&\leq 1 + x^2\sigma^2/2 + \sigma^2 M^{-2}/3! \sum_{k=3}^{\infty} \frac{x^k M^k}{3^{k-3}} \\
&= 1 + x^2\sigma^2/2 + \sigma^2 M^{-2}/6 \frac{(xM/3)^3}{(1 - xM/3)} \\
&= 1 + x^2\sigma^2/2 = \frac{\sigma^2 M x^3}{6(1 - xM/3)}.
\end{aligned}$$

If $0 \leq x \leq 2/M$, we have

$$\begin{aligned}
\mathbb{E}(e^{xX_1}) &\leq 1 + \sigma^2 x^2/2 + \sigma^2 x^2/2(xM) \\
&= 1 + \sigma^2 x^2/2(1 + xM) \\
&\leq \exp(\sigma^2 x^2/2(1 + xM))
\end{aligned}$$

$\square$

Corollary 8.5. *For* $0 < a < \frac{2\sigma^2 n}{M}$*, under the conditions of the above Lemma,*

$$P(S_n \geq a) \leq e^{-\frac{a^2}{2n\sigma^2})1-\frac{Ma}{n\sigma^2}}$$

Proof.

$$P(S_n \geq a) \leq \frac{E(e^{xS_n})}{e^{ax}}$$
$$\leq \exp(\frac{n\sigma^2 x^2}{2}(1+xM) - ax) \quad 0 < x \leq \frac{2}{M}$$

Put $x = \frac{a}{n\sigma^2}$. THen

$$P(S_n \geq a) \leq \exp(\frac{a^2}{2n\sigma^2}(1 = \frac{aM}{n\sigma^2}) - \frac{a^2}{n\sigma^2})$$
$$= \exp(\frac{-a^2}{2n\sigma^2}(1 - \frac{aM}{n\sigma^2}))$$

□

We can now prove the Hardy-Littlewood result. If $|X_i| \leq M$ almost surely then $|S_n| = \mathcal{O}(\sqrt{n \log n})$ a.s.

Proof. Put $a = c\sqrt{n \log n}$. Then

$$P(S_n \geq c\sqrt{n \log n}) \leq \exp(\frac{c^2 \log n}{2\sigma^2}(1 - \frac{Mc\sqrt{\log n}}{\sqrt{n}\sigma^2}))$$
$$= n^{-c^2/2\sigma^2} \exp(\frac{Mc^3}{2\sigma^4} \frac{\log n \sqrt{\log n}}{\sqrt{n}})$$

If $c^2 > 2\sigma^2$ then $\sum_n P(S_n > c\sqrt{n \log n}) < \infty$. By Borel-Cantelli, we then have

$$P(S_n > c\sqrt{n \log n}\, i.o.) = 0$$

Now apply the argument to $-X_i$. Then

$$P(-S_n > c\sqrt{n \log n}\, i.o.) = 0$$

□

Theorem 8.6 (Khintchine (1923))**.** $|S_n| = \mathcal{O}(\sqrt{n \log \log n})$ *a.s.*

Theorem 8.7 (Khintchine (1924))**.** *Let $X_i = \pm 1$ with probability $\frac{1}{2}$. Then*

$$\limsup \frac{|S_n|}{\sqrt{n \log \log n}} = \sqrt{2} a.s.$$

9. Lecture 9 - Thursday 31 March

Definition 9.1 (Induced σ-field)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let Y be a set of random variables on $(\Omega, \mathcal{F})$. Then $\sigma(Y)$ is the smallest σ-field contained in $\mathcal{F}$ with respect to which each $X \in Y$ is measurable.

That is, for each $B \in \mathcal{B}$, the Borel σ-field on $\mathbb{R}$, we have

$$X^{-1}(B) \in \sigma(Y)$$

Thus $\sigma(Y)$ is the intersection of all σ-fields which contain every set of the form $X^{-1}(B)$ for all $B \in \mathcal{B}, X \in Y$.

Definition 9.2 (Independent σ-fields)**.** If $X_1, \dots$ are independent random variables and $A_i \in \sigma(X_i)$, then

$$P(\bigcap_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i) \tag{$\star$}$$

If $\mathcal{F}_1, \mathcal{F}_2, \dots$ are σ-fields contained in $\mathcal{F}$ and ($\star$) holds for any $A_i \in \mathcal{F}_i$ then we sa the σ-fields are independent.

Theorem 9.3. *Let $\mathcal{F}_0, \mathcal{F}_1, \dots$ be independent σ-fields and let $\mathcal{G}$ be σ-fields generated by any subset of $\mathcal{F}_1, \mathcal{F}_2, \dots$. Then $\mathcal{F}_0$ is independent of $\mathcal{G}$.*

Proof. **Outline.** Take $\mathcal{G}$ to be the smallest σ-field containing $\mathcal{F}_1, \mathcal{F}_2, \dots$.

If $A \in \mathcal{F}_0$, $B \in \mathcal{G}$, then we need to show

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

(1) Assume $P(A) > 0$.
(2) If $B = A_1 \cap A_2 \dots A_n$ then the result is true.
(3) Let $\mathcal{G}_a$ be the class of **finite** unions of B. Then $\mathcal{G}_a$ is a finitely additive field, and $G \in \mathcal{G}_a$ can be written as $G = \bigcup_{i=1}^{k} G_i$ where G_i has the form of B above. Then

$$\begin{aligned} P(A \cap G) &= P(\bigcup_{i=1}^{k} A \cap G_i) \\ &= \sum P(A \cap G_i) = \sum_{i=j} P(A \cap G_i \cap G_j) + \dots \\ &= P(A)P(G) \end{aligned}$$

by the inclusion-exclusion formula and independence of A and G_i.
(4) Now, let $P_A(B) = \frac{P(A \cap B)}{P(A)}$. Then P_A and P are measures on $\mathcal{F}$, and P and P_A agree on $\mathcal{G}_a$. Thus by the extension theorem they agree on the σ-field generated by $\mathcal{G}_a$ which includes $\mathcal{G}$.

□

Definition 9.4 (Tail σ-field)**.** Let $X_1, X_2, \dots$ be a sequence of random variables and let

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

be the σ field generated by X_n, X_{n+1}. Then

$$\mathcal{F}_n \supseteq F_{n+1} \supseteq F_{n+2} \dots$$

and let

$$\mathcal{T} = \bigcap n \mathcal{F}_n$$

be the **tail σ-field**.

$\mathcal{T}$ is the collection of events defined in terms of $X_1, X_2, \dots$ not affected by altering a finite number of the random variables.

Theorem 9.5 (The $0-1$ law)**.** *Any set belonging to the tail σ-field of a sequence of independent random variables has probability* 0 *or* 1*.*

Proof. We have $\sigma(X_n)$ is independent of $\sigma(\{X_{n+1}, X_{n+2}, \dots\}) = \mathcal{F}_{n+1} \supseteq \mathcal{T}$ and so $\mathcal{T}$ is independent of $\sigma(X_n)$ for every n. By the previous theorem, it follows that $\mathcal{F}$ is independent of $\mathcal{G} = \sigma(\{X_1, X_2, \dots\})$ but as $\mathcal{T} \subseteq \mathcal{G}$, we know that $\mathcal{T}$ is independent of itself. Thus, for any $A \in \mathcal{T}$,

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

and so $P(A) = 0$ or 1. ◻

9.1. **Martingales.**

Definition 9.6 (Martingale)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\{\mathcal{F}_n\}$ be an increasing sequence of σ-fields.

$$\mathcal{F}_n \subseteq \mathcal{F}_{n+1} \subseteq \cdots \subseteq \mathcal{F}.$$

Let $\{S_n\}$ be a sequence of random variables on Ω. Then $\{S_n\}$ is a **martingale** with respect to $\{\mathcal{F}_n\}$ if

(1) S_n is measurable with respect to $\mathcal{F}_n$.
(2) $\mathbb{E}(|S_n|) < \infty$.
(3) $\mathbb{E}(S_n \,|\, \mathcal{F}_m) = S_m$ almost surely for all $m \leq n$.

10. Lecture 10 - Thursday 31 March

Definition 10.1 (Supermartingale)**.** $\{S_n\}$ is a **supermartingale** with respect to $\{\mathcal{F}_n\}$ if

(1) S_n is measurable with respect to $\mathcal{F}_n$.
(2) $\mathbb{E}(|S_n|) < \infty$.
(3) $\mathbb{E}(S_n \,|\, \mathcal{F}_m) \leq S_m$ almost surely for all $m \leq n$.

Definition 10.2 (Submartingale)**.** $\{S_n\}$ is a **submartingale** with respect to $\{\mathcal{F}_n\}$ if

(1) S_n is measurable with respect to $\mathcal{F}_n$.
(2) $\mathbb{E}(|S_n|) < \infty$.
(3) $\mathbb{E}(S_n \,|\, \mathcal{F}_m) \geq S_m$ almost surely for all $m \leq n$.

Definition 10.3 (Regular martingale)**.** Let X is a random variable $\mathbb{E}(|X|) < \infty$, $S_n = \mathbb{E}(X \,|\, \mathcal{F}_n)$ and assume $\{S_n\}$ is a martingale with respect to $\{F_n\}$.

If a martingale can be written in this way for some X then it is **regular**.

Not every martingale is a regular martingale.

Example 10.4. Assume $P(X_i = 1) = p$, $P(X_i = -1) = 1 - p$, and let $S_n = \sum_{i=1}^n X_i$. If $p \neq \frac{1}{2}$ then

$$Y_n = \left(\frac{1-p}{p}\right)^{S_n}$$

is a martingale with respect to $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$, since

$$\begin{aligned}
\mathbb{E}(Y_n \,|\, \mathcal{F}_{n-1}) &= \mathbb{E}\left(\left(\frac{1-p}{p}\right)^{S_n + X_n} \,|\, \mathcal{F}_{n-1}\right) \\
&= \left(\frac{1-p}{p}\right)^{S_n}\left[\left(\frac{1-p}{p}\right)p + \left(\frac{1-p}{p}\right)^{-1}(1-p)\right] \\
&= Y_{n-1}
\end{aligned}$$

10.1. **Conditional expectations.** If $\mathcal{G} \subseteq \mathcal{F}$ then

$$L^2(\mathcal{G}) = \{X \,|\, \mathbb{E}(X^2) < \infty, X \text{ is } \mathcal{G}\text{-measurable}\}$$

If $Y \in L^2$ define $Z = \mathbb{E}(Y \,|\, \mathcal{G})$ to be the projection of Y onto $L^2(\mathcal{G})$, where

$$\mathbb{E}(Y - Z)^2 = \inf_{U \in L^2(\mathcal{G})} \mathbb{E}(Y - U)^2$$

Then $Y - Z$ will be orthogonal to the elements of $L^2(\mathcal{G})$. That is,

$$\int (Y - Z)X \, dP = 0$$

for all $X \in L^2(\mathcal{G})$. If $A \in \mathcal{G}$, then letting $X = \mathbf{1}_A$, we have

$$\boxed{\int_A Y \, dP = \int_A \mathbb{E}(Y \,|\, \mathcal{G}) \, dP}$$

If $Y \geq 0$ construct $\{Y_n\}$ with $Y_n \in L^2$ such that $Y_n \uparrow Y$. Define

$$\mathbb{E}(Y \,|\, \mathcal{G}) = \lim_{n \to \infty} \mathbb{E}(Y_n \,|\, \mathcal{G}).$$

The limit exists as

$$\mathbb{E}(Y_n \,|\, \mathcal{G}) \geq \mathbb{E}(Y_m \,|\, \mathcal{G}), n \geq m.$$

We still have

(1) $\mathbb{E}(Y \,|\, \mathcal{G})$ is $\mathcal{G}$-measurable, and
(2) For all $A \in \mathcal{G}$,

$$\int_A Y \, dP = \int_A \mathbb{E}(Y \,|\, \mathcal{G}) \, dP$$

as

$$\int_A \mathbb{E}(Y \,|\, \mathcal{G}) \, dP = \lim_{n \to \infty} \int_A \mathbb{E}(Y_n \,|\, \mathcal{G}) \, dP = \lim_{n \to \infty} \int_A Y_n \, dP = \int_A Y \, dP$$

by the monotone convergence theorem.

If $Y \in L^1$, defining $Y = Y^+ - Y^-$, we define

$$\mathbb{E}(Y \,|\, \mathcal{G}) = \mathbb{E}(Y^+ \,|\, \mathcal{G}) - \mathbb{E}(Y^- \,|\, \mathcal{G}).$$

10.2. **Stopping times.**

Definition 10.5. A map

$$\nu : \Omega \to \overline{\mathbb{N}} = \{0, 1, 2, \ldots, \infty\}$$

is called a **stopping time** with respect to $\{\mathcal{F}_n\}$, an increasing sequence of σ-fields, if

$$\{\nu = n\} \in \mathcal{F}_n.$$

and thus

$$\{\nu \leq n\}, \{\nu > n\} \in \mathcal{F}_n$$

Theorem 10.6 (Properties of stopping times)**.** *Let $\mathcal{F}_\infty = \vee_{n=1}^{\infty} \mathcal{F}_n$, the σ-field generated by all $\mathcal{F}_n$. Then we have*

(1) For all stopping times ν, ν is $\mathcal{F}_\infty$-measurable.

$$\{\nu = n\} \in \mathcal{F}_n, \{\nu = \infty\} = \{\bigcup_n \{v = n\}\}^c \in \mathcal{F}_\infty$$

(2) The minimum and maximum of a countable sequence of stopping times is a stopping time. To prove this, let $\{v_k\}$ be a sequence of stopping times. Then

$$\{\max_k v_k \leq n\} = \bigcap_k \{v_k \leq n\} \in \mathcal{F}_n$$

$$\{\min_k v_k > n\} = \bigcap_k \{v_k > n\} \in \mathcal{F}_n$$

Lemma 10.7. *Let $\{Y_n^1\}$ and $\{Y_n^2\}$ be two positive supermartingales with respect to $\{\mathcal{F}_n\}$, an increasing sequence of σ-fields. Let ν be a stopping time. If $Y_n^1 \geq Y_n^2$ on $[\nu = n]$, then*

$$Z_n = Y_n^1 \mathbf{1}_{\{\nu > n\}} + Y_n^2 \mathbf{1}_{\{\nu \leq n\}}$$

is a positive supermartingale.

Proof. We have that Z_n is $\mathcal{F}_n$-measurable and positive. We then have

$$\begin{aligned}
\mathbb{E}(Z_n \,|\, \mathcal{F}_{n-1}) &= \mathbb{E}(Y_n^1 \mathbf{1}_{\{\nu > n\}} + Y_n^2 \mathbf{1}_{\{\nu \leq n\}} \,|\, \mathcal{F}_{n-1}) \\
&= \mathbb{E}(Y_n^1 \mathbf{1}_{\{\nu > n-1\}} - Y_n^1 \mathbf{1}_{\{\nu = n\}} + Y_n^2 \mathbf{1}_{\{\nu \leq n-1\}} + Y_n^2 \mathbf{1}_{\{\nu = n\}} \,|\, \mathcal{F}_{n-1}) \\
&\leq Y_n^1 \mathbf{1}_{\{\nu > n-1\}} + Y_n^2 \mathbf{1}_{\{\nu \leq n-1\}} + \mathbb{E}((Y_n^2 - Y_n^1) \mathbf{1}_{\{\nu = n\}} \,|\, \mathcal{F}_{n-1}) \\
&\leq Z_{n-1}
\end{aligned}$$

as $Y_n^2 - Y_n^1 < 0$ on $\{\nu = n\}$. ∎

11. Lecture 11 - Thursday 7 April

Theorem 11.1 (Maximal inequality for positive supermartingales)**.** *Let $\{Y_n\}$ be a positive supermartingale with respect to $\{\mathcal{F}_n\}$. Then*

$$\sup_n Y_n < \infty a.s$$

on $[Y_0 < \infty]$ and

$$P(\sup_n Y_n > a \,|\, \mathcal{F}_0) \leq \min(1, \frac{Y_0}{a})$$

Proof. Fix $a > 0$ and let $\nu_a = \inf\{n : Y_n > a\} = \infty$ if $\sup_n Y_n \leq a$. Then the sequence $Y_n(2) = a$ is a positive supermartingale, and so

$$Z_n = Y_n \mathbf{1}_{\{\nu_a > n\}} + a \mathbf{1}_{\{\nu_a \leq n\}}$$

is a positive supermartingale by the previous lemma. Then we have

$$\mathbb{E}(Z_n \,|\, \mathcal{F}_0) \leq Z_0 = \begin{cases} Y_0 & Y_0 \leq a \\ a & Y_0 > a \end{cases}$$

Thus $Z_n \geq a \mathbf{1}_{\{\nu_a \leq n\}}$ and so

$$aP(va \leq n \,|\, \mathcal{F}_0) \leq \min(Y_0, a)$$

for all a. THus

$$P(\sup_n Y_n > a \,|\, \mathcal{F}_0) = P(\nu_a < \infty \,|\, \mathcal{F}_0) \leq \min(1, \frac{Y_0}{a})$$

□

Write

$$\begin{aligned}
P(Y_0 < \infty, \sup_n Y_n > a) &= \mathbb{E}(\mathbf{1}_{\{Y_0 < \infty\}} \mathbf{1}_{\{\sup_n Y_n > a\}}) \\
&= \mathbb{E}(\mathbf{1}_{\{Y_0 < \infty\}}) \mathbb{E}(\mathbf{1}_{\{\sup_n Y_n > a\}} \,|\, \mathcal{F}_0) \\
&\leq \int_{Y_0 < \infty} \min(1, \frac{Y_0}{a}) \, dP \\
&\to 0
\end{aligned}$$

as $a \to \infty$ by the dominated convergence theorem.

Thus, we have

$$P(Y_0 < \infty, \sup_n Y_n < \infty) = 1 a.s.$$

Fix $a < b \in \mathbb{R}$. For any process Y_n, define the following random variables

$$\nu_1 = \min(n \geq 0, Y_n \leq a)$$
$$\nu_2 = \min(n > \nu_1, Y_n \geq b)$$
$$\nu_3 = \min(n > \nu_2, Y_n \leq a)$$

and so on. If any v_i is undefined it is subsequently set to infinity.

Define $\beta_{ab} = \max\ p : \nu v_{2p} < \infty$, equal to the number of upcrossings of (a, b) by Y_n. We have $\beta_{ab} = \infty$ if and only if $\liminf y_n \leq a < b \leq \limsup y_n$. We also have Y_n converges if and only if $\beta_{ab} < \infty$ for all rationals a, b, $a < b$.

Theorem 11.2 (Dubin's inequality)**.** *If Y_n is a positive supermartingale, then $\beta_{ab}(\omega)$ are random variables and for each integer $k \geq 1$, we have*

$$P(\beta_{ab} \geq k \,|\, \mathcal{F}_0) \leq (\frac{a}{b})^k \min(1, \frac{Y_0}{a}), 0 < a < b.$$

Proof. The v_k defined above are stopping times with respect to $\mathcal{F}_n$, as

$$[\nu_{2p} = n] = \bigcup_{m=0}^{n-1} [\nu_{2p-1} = m, Y_{m+1} \leq b, \ldots, Y_{n-1} < b, Y_n \geq b]$$

and as ν_1 is a stopping time, we then use induction.

We then have $[\beta_{ab} \geq k] = [\nu_{2k} < \infty]$. Then define

$$Z_n = \mathbf{1}_{\{0 \leq n < \nu_1\}} + \sum_{k=1}^{K} (\frac{b}{a})^{k-1} \frac{Y_n}{a} \mathbf{1}_{\{\nu_{2k-1} \leq n \leq \nu_{2k}\}}$$
$$+ \left(\frac{b}{a}\right)^k \mathbf{1}_{\{\nu_{2k} \leq n < \nu_{2k+1}\}} + \left(\frac{b}{a}\right)^K \mathbf{1}_{\{n \geq \nu_{2K+1}\}}$$

i.e. $\mathbf{1}_{\{0 \leq n < \nu_1\}} + \frac{Y_n}{a} \mathbf{1}_{\{\}} \nu_1 \leq n < \nu_2 + \frac{b}{a} \mathbf{1}_{\{\nu_2 \leq n < \nu_3\}} + \cdots + \left(\frac{b}{a}\right)^K \mathbf{1}_{\{\nu_{2K} \leq n\}}$.

We now apply the previous lemma to show $\{Z_n\}$ is a positive supermartingale. We have

$$\left(\frac{b}{a}\right)^k, \left(\frac{b}{a}\right)^{k-1} \frac{Y_n}{a}$$

are positive supermartingales. On $[\nu_1 = n]$, we have $1 \geq \frac{Y_n}{a}$. On $[\nu_{2k-1} = n]$ we have

$$\left(\frac{b}{a}\right)^{k-1} \geq \left(\frac{b}{a}\right)^{k-1} \frac{Y_n}{a}$$

On the even stopping times, we have $\left(\frac{b}{a}\right)^{k-1} \frac{Y_n}{a} \geq \left(\frac{b}{a}\right)^k$. Thus

$$\mathbb{E}(Z_n \,|\, \mathcal{F}_0) \leq Z_0$$

as Z_n is a positive supermartingale. d

Since $Z_n \geq \frac{b}{a}^K \mathbf{1}_{\{\nu_{2k} \leq n\}}$, we have

$$P(\nu_{2k} \leq n \,|\, \mathcal{F}_0) \leq \frac{a}{b}^K \min(1, \frac{Y_0}{a})$$

Letting $n \to \infty$, we have

$$\begin{aligned} P(\beta_{ab} \geq k \,|\, \mathcal{F}_0) &= P(\nu_{2k} < \infty \,|\, \mathcal{F}_0) \\ &\leq \left(\frac{a}{b}\right)^K \min(1, \frac{Y_0}{a}). \end{aligned}$$

□

12. Lecture 12 - Thursday 7 April

Theorem 12.1. *Let $\{Y_n\}$ be a positive supermartingale. Then there exists a random varaible Y_∞ such that $Y_n \overset{a.s.}{\to} Y_\infty$ and $\mathbb{E}(Y_\infty \,|\, \mathcal{F}_n) \leq Y_n$ for all n.*

Proof. From Durbin's inequality,

$$P(\beta_{ab} \geq k) \leq \left(\frac{a}{b}\right)^k$$

By Borel-Cantelli, as we have a summable sequence of probabilities, $\beta_{ab} < \infty$ almost surely. Hence

$$P(Y_n \text{ converges}) = P\left(\bigcap_{\substack{a<b \\ a,b \in \mathbb{Q}}} \beta_{ab} < \infty\right) = 1$$

Let $\lim_{n\to\infty} Y_n = Y_\infty$. If $p < n$, then

$$\mathbb{E}\left(\inf_{m \geq n} Y_n \,|\, \mathcal{F}_p\right) \leq \mathbb{E}(Y_n \,|\, \mathcal{F}_p) \leq Y_p.$$

Furthermore, $\inf_{m \geq n} Y_m \uparrow Y_\infty$ so by the monotone convergence theorem, we have

$$\mathbb{E}(Y_\infty \,|\, \mathcal{F}_p) = \lim_{n\to\infty} \mathbb{E}\left(\inf_{m \geq n} Y_m \,|\, \mathcal{F}_p\right) \leq Y_p.$$

□

Theorem 12.2. *Let Z be a positive random variable with $\mathbb{E}Z^p < \infty$, $p \geq 1$. Then*

$$Y_n = \mathbb{E}(Z_n \,|\, \mathcal{F}_n) \overset{a.s.}{\to}, \overset{L^p}{\to} \mathbb{E}(Z \,|\, \mathcal{F}_\infty),$$

Note that almost sure convergence does not, in general, imply L^p convergence, although they both imply convergence in probability.

Proof. Suppose $Z \leq a$ almost surely. Then there exists Y_∞ such that $Y_n \overset{a.s.}{\to} Y_\infty$ (as Y_n are positive martingales). Fix n and let $B \in \mathcal{F}_n$. Then

$$\lim_{n\to\infty} \int_B Y_{m+n} \, dP = \int_B Z \, dP$$

by definition of conditional expectation. Now $0 \le Y_n \le a$ so by the dominated convergence theorem,

$$\int_B Y_\infty \, dP = \int_B Z \, dP$$

and hence

$$Y_\infty = \mathbb{E}(Z \mid \mathcal{F}_\infty)$$

and so the random variable Y_∞ can be identified as the conditional expectation.

Since $|Y_n| \le a$, the $\{Y_n^p\}$ are uniformly integrable, and so $Y_n \xrightarrow{L^p} Y_\infty$. This follows from noting that $Y_n \xrightarrow{a.s.} Y_\infty$, and using that if $X_n \xrightarrow{p} X$ and $\{|X_n|^p\}$ is uniformly integrable then $X_n \xrightarrow{L^p} X$.

Now remove the assumption that $Z \le a$. Taking the L^p norm of the conditional expectations gives

$$\|E(Z \mid \mathcal{F}_n) - \mathbb{E}(Z \mid \mathcal{F}_\infty)\|_p \le \|E(Z \wedge a \mid \mathcal{F}_n) - \mathbb{E}(Z \wedge a \mid \mathcal{F}_\infty)\|_p + 2\|(Z-a)^+\|_p.$$

Now we know that $\|(Z-a)^+\|_p \to 0$ as $a \to \infty$, as $\mathbb{E}(Z^p) < \infty$. Hence we have

$$Y_n \xrightarrow{L^p} \mathbb{E}(Z \mid \mathcal{F}_\infty).$$

By uniqueness of limits, we obtain our required result. □

Corollary 12.3. *If $Z \in L^p$ and $Y_n = \mathbb{E}(Z \mid \mathcal{F}_n)$ then $Y_n \xrightarrow{a.s.}, \xrightarrow{L^p} \mathbb{E}(Z \mid \mathcal{F}_\infty)$*

Theorem 12.4. *Martingale convergence theorem*

(a) If $\{Y_n\}$ is an integrable submartingale and $\sup_n \mathbb{E}(Y_n^+) < \infty$ then there exists an integrable Y_∞ such that

$$Y_n \xrightarrow{a.s.} Y_\infty$$

(b) If $\{Y_n\}$ is an integrable martingale satisfying $\sup_n \mathbb{E}|Y_n| < \infty$ then there exists an integrable Y_∞ such that

$$Y_n \xrightarrow{a.s.} Y_\infty.$$

Proof.

(a) $\{Y_n^+\}$ is a positive submartingale as

$$\mathbb{E}(Y_{n+1}^+ \mid \mathcal{F}_n) \ge \mathbb{E}(Y_{n+1} \mid \mathcal{F}_n) \ge Y_n$$

If $p > n$, then

$$\begin{aligned}\mathbb{E}(Y_{p+1}^+ \mid \mathcal{F}_n) &= \mathbb{E}(\mathbb{E}(Y_{p+1}^+ \mid \mathcal{F}_p) \mid \mathcal{F}_n) \\ &\ge \mathbb{E}(Y_{p^+} \mid \mathcal{F}_n).\end{aligned}$$

Hence $M_n = \lim_{p\to\infty} \mathbb{E}(Y_p^+ \mid \mathcal{F}_n)$ as we have a monotone sequence.

Now,

$$\begin{aligned}\mathbb{E}(M_n) &= \mathbb{E}\left(\lim_{p\to\infty}\mathbb{E}(Y_p^+ \mid \mathcal{F}_n)\right)\\ &= \lim_{p\to\infty}\mathbb{E}(\mathbb{E}(Y_p^+ \mid \mathcal{F}_n)) \quad \text{MCT}\\ &= \lim_{p\to\infty}\mathbb{E}(Y_p^+) < \infty\end{aligned}$$

so M_n is positive and integrable. M_n is a martingale as

$$\begin{aligned}\mathbb{E}(M_{n+1} \mid \mathcal{F}_n) &= \mathbb{E}\left(\lim_{p\to\infty}\mathbb{E}(Y_p^+ \mid \mathcal{F}_{n+1}) \mid \mathcal{F}_n\right)\\ &= \lim_{p\to\infty}\mathbb{E}(Y_p^+ \mid \mathcal{F}_n) \quad \text{MCT}\\ &= M_n.\end{aligned}$$

Let $Z_n = M_n - Y_n$. Then Z_n is integrable as $M_, Y_n$ are, and Z_n is a positive supermartingale, as

$$\begin{aligned}\mathbb{E}(Z_{n+1} \mid \mathcal{F}_n) &= \mathbb{E}(M_{n+1} \mid \mathcal{F}_n) - \mathbb{E}(Y_{n+1} \mid \mathcal{F}_n)\\ &\le M_n - Y_n \quad \text{as } Y_n \text{ is submartingale}\\ &= Z_n\end{aligned}$$

and so Z_n is a positive supermartingale. Note that $M_n \ge Y_n^+$ and so

$$M_n - Y_n = M_n - (Y_n^+ - Y_n^-) \ge Y_n^+ - (Y_n^+ - Y_n^-) = Y_n^-$$

Thus Z_n and M_n converge almost surely to Z_∞ and M_∞ respectively, and so

$$Y_n = M_n - Z_n \overset{a.s.}{\to} M_\infty - Z_\infty = Y_\infty \in L^1.$$

(b) Note that $|Y_n| = 2Y_n^+ - Y_n$, and if $\{Y_n\}$ is a martingale, then

$$\begin{aligned}\mathbb{E}|Y_n| &= 2\mathbb{E}Y_n^+ - \mathbb{E}Y_n\\ &2\mathbb{E}Y_n^+ - \mathbb{E}Y_0\end{aligned}$$

and so $\sup \mathbb{E}Y_n^+ < \infty$ if and only if $\sup_n \mathbb{E}|Y_n| < \infty$.

$\square$

Theorem 12.5 (Martingale convergence theorem (restated))**.** *Let $\{Y_n\}$ be an integrable (sub/super) martingale, that is, $\sup_n \mathbb{E}|Y_n| < \infty$. Then there exists an almost sure limit*

$$\lim_{n\to\infty} Y_t = Y_\infty$$

and Y_∞ is an integrable random variable.

13. Lecture 13, 14 - Thursday 14 April

Definition 13.1 (Reverse martingale). $\{Y_n, \mathcal{G}_n\}$ is a reverse martingale if $\{\mathcal{G}_n\}$ is a decreasing sequence of σ-fields,

$$\mathcal{G}_n \supseteq \mathcal{G}_{n+1}$$

Y_n is $\mathcal{G}_n$-measurable, $\mathbb{E}(|Y_n|) < \infty$, and

$$\mathbb{E}(Y_n \,|\, \mathcal{G}_n) = Y_m \text{ a.s for } m \geq n$$

Proposition 13.2. *We have*

$$\begin{aligned} \mathbb{E}(|Y_n|) &= \mathbb{E}(\mathbb{E}(|Y_n| \,|\, \mathcal{G}_{n+1})) \\ &\geq \mathbb{E}(|\mathbb{E}(Y_n \,|\, \mathcal{G}_{n+1})|) \\ &= \mathbb{E}(|Y_{n+1}|) \end{aligned}$$

and so $\mathbb{E}(|Y_n|) \leq \mathbb{E}(|Y_0|)$ *for all* n, *and*

$$Y_n = \mathbb{E}(Y_0 \,|\, \mathcal{G}_n).$$

Theorem 13.3. *If* $\{Y_n\}$ *is a reverse martingale with respect to* $\{\mathcal{G}_n\}$, *then there exists a random variable* Y_∞ *such that*

$$Y_n \stackrel{a.s.}{\to} Y_\infty, Y_n \stackrel{L^1}{\to} Y_\infty = \mathbb{E}(Y_0 \,|\, \mathcal{G}_\infty)$$

where $\mathcal{G}_\infty = \bigcap \mathcal{G}_n$.

Proof. We have $Y_n = \mathbb{E}(Y_0 \,|\, \mathcal{G}_n)$ and so $\{Y_n\}$ is uniformly integrable. Hence if $Y_n \stackrel{a.s.}{\to} Y_\infty$ it also converges in L^1. Let

$$Z_n = \mathbb{E}(Y_0^+ \,|\, \mathcal{G}_n) - Y_n.$$

Note that $Z_n \geq 0$. Then

$$\mathbb{E}(Z_n \,|\, \mathcal{G}_{n+1}) = Z_{n+1}$$

and so we only need to consider convergence for positive reverse martingales.

Let $\beta_{a,b}^{(n)}$ be the number of upcrossings of $[a, b]$ by $\{Y_0, Y_1, \dots, Y_n\}$. Applying Dubin's inequality to the martingale

$$\{Y_n, Y_{n+1}, \dots, Y_1, Y_0\}$$

Then

$$P(\beta_{a,b}^{(n)} \geq k \,|\, \mathcal{G}_n) \leq \left(\frac{a}{b}\right)^k$$

which is independent of n, and thus

$$P(\beta_{a,b}^{(n)} \geq k \,|\, \mathcal{G}_\infty) \leq \left(\frac{a}{b}\right)^k$$

for all n, and so

$$P(\beta_{a,b} \geq k \,|\, \mathcal{G}_\infty) \leq \left(\frac{a}{b}\right)^k.$$

where $\beta_{a,b}$ is the number of upcrossings for $\{Y_n\}$, which implies

$$\beta_{a,b} < \infty \,\text{a.s.}$$

Arguing as in the positive supermartingale case, we have $\{Y_n\}$ converges almost surely, and we have $Y_\infty = \limsup Y_n$ is $\mathcal{G}_n$ measurable for all n and so is $\mathcal{G}_\infty$ measurable. □

Theorem 13.4 (Strong law of large numbers)**.** *Let $X_1, X_2, \ldots$ be IID with $\mathbb{E}(|X_1|) < \infty$. Let $\mathbb{E}(X_1) = \mu$. Let $S_n = \sum_{i=1}^n X_i$. Then*

$$\frac{1}{n} S_n \overset{a.s.}{\to} \mu.$$

Proof. Let $\mathcal{G}_n = \sigma\{S_n, S_{n+1}, S_{n+2}, \ldots\} = \sigma\{S_n, X_{n+1}, X_{n+2}, \ldots\}$. We then have $\mathcal{G}_n \supseteq \mathcal{G}_{n+1}$. We have

$$\begin{aligned}
\frac{1}{n} S_n &= \mathbb{E}(\frac{1}{n} S_n \,|\, \mathcal{G}_n) \\
&= \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i \,|\, \mathcal{G}_n) \\
&= \mathbb{E}(X_1 \,|\, \mathcal{G}_n),
\end{aligned}$$

as

$$\mathbb{E}(X_1 \,|\, \mathcal{G}_n) = \mathbb{E}(X_2 \,|\, \mathcal{G}_n) = \ldots \mathbb{E}(X_n \,|\, \mathcal{G}_n)$$

by IID/symmetry.

Thus $\frac{1}{n} S_n$ is a reverse martingale with respect to $\{G_n\}$. From above, we have have

$$\frac{1}{n} S_n = \overline{X}_n \overset{a.s.}{\to}, \overset{L^1}{\to} \mathbb{E}(X \,|\, \mathcal{G}_\infty).$$

We have $\lim_{n\to\infty} \sum_{i=1}^n X_i$ is in the tail σ-field of the sequence of $\{X_n\}$ and X_i are IID and so the limiting random variable is degenerate.

Consider $\overline{X}_\infty = \mathbb{E}(X \,|\, \mathcal{G}_\infty)$. By the Kolmogorov 0-1 law, we have

$$P(\{\overline{X}_\infty \leq a\}) = 0 \text{ or } 1.$$

Thus $\overline{X}_\infty$ is a constant with probability one. Since

$$\mathbb{E}(X_1 \,|\, \mathcal{G}_n) \overset{L^1}{\to} \mathbb{E}(X_1 \,|\, \mathcal{G}_\infty)$$

we have

$$\lim_{n\to\infty} \mathbb{E}(\frac{1}{n} S_n) = \mathbb{E}(\mathbb{E}(X_1 \,|\, \mathcal{G}_\infty)) = \mathbb{E}(X_1) = \mu.$$

Thus $\overline{X}_\infty = \mu$ almost surely, that is,

$$\frac{1}{n} S_n \overset{a.s.}{\to}, \overset{L^1}{\to} \mu.$$

□

13.1. **Characteristic functions.** Following **Fallow Volume 2**.

Definition 13.5 (Characteristic function)**.** Let X be a random variable. Then the characteristic function is defined by

$$\varphi(t) = \mathbb{E}(e^{itX}).$$

$\varphi(t)$ is always defined (unlike moment generating function (MGF), probability generating function (PGF)).

Proof. Let $\varphi(t)$ be the characteristic function of the random variable X. Then

(i) $|\varphi(t)| \leq \mathbb{E}(|e^{itX}|) = 1 = \varphi(0)$.
(ii) $\varphi(-t) = \mathbb{E}(e^{-itX}) = \overline{\varphi(t)}$.
(iii) If X is symmetric about 0 then $\varphi(t)$ is real.
(iv) $\varphi(t)$ is uniformly continuous in t.

Proof.

$$\begin{aligned} |\varphi(t+h) - \varphi(t)| &= \left| \int e^{i(t+h)X} - e^{itX} \, dF(x) \right| \\ &= \left| \int e^{itX} (e^{ihX} - 1) \, dF(x) \right| \\ &\leq \int \left| e^{ihX} - 1 \right| \, dF(x) \\ &= \int \sqrt{\cos^2(xh-1) + \sin^2(xh)} \, dF(x) \\ &= \int \sqrt{2 - 2\cos hx} \, dF(x) \to 0 \end{aligned}$$

as $h \to 0$ by the dominated convergence theorem. $\square$

(v) If X and Y are independent random variables with characteristic functions φ and ψ respectively, then $X + Y$ has characteristic function

$$\chi(t) = \varphi(t) \cdot \psi(t)$$

(vi) If X has a characteristic function φ then $aX + b$ has a characteristic function $e^{itb}\varphi(at)$.
(vii) If φ is a characteristic function the so is $|\varphi|^2$.

Proof. Let X and Y have the same distribution, with X independent of Y. Then $Z = X - Y$ has a characteristic function $\varphi(t)\varphi(-t) = |\varphi(t)|^2$. $\square$

(viii) Let X have a MGF $M(t)$. Then $\varphi(t) = M(it)$.

$\square$

Example 13.6. (i) Let $X \sim N(0,1)$. Then
$$\varphi(t) = e^{-\frac{1}{2}t^2}.$$

(ii) Let $Y \sim N(\mu, \sigma^2)$. Then
$$\varphi(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2}.$$
as $Y = \mu + \sigma Z$ with $Z \sim N(0,1)$.

(iii) Let $X \sim \text{Poisson}(\lambda)$ Then
$$\varphi(t) = e^{\lambda(e^{it}-1)}.$$

(iv) Let $P(X = 1) = \frac{1}{2} = P(X = -1)$. Then
$$\varphi(t) = \frac{1}{2}\left(e^{it} + e^{-it}\right) = \cos t.$$

(v) Let $X \sim \text{Exp}(\lambda)$. Then
$$\begin{aligned}\varphi(t) &= \int_0^\infty e^{itx} \lambda e^{-\lambda x}\, dx \\ &= \int_0^\infty \lambda e^{-x(\lambda - it)}\, dx \\ &= \frac{\lambda}{\lambda - it}\end{aligned}$$

Theorem 13.7 (Parseval's relation)**.** *Let F and G be distribution functions with associate characteristic functions φ and ψ. Then*
$$\int e^{-izt}\varphi(z)\, dG(z) = \int \psi(x-t)\, dF(x)$$

Proof.
$$\begin{aligned}\int e^{-izt}\varphi(z)\, dG(z) &= \int e^{-izt}\left(\int e^{izt}\, dF(x)\right) dG(z) \\ &= \int\int e^{iz(x-t)}\, dF(x)\, dG(x) \\ &= \int \left(\int e^{iz(x-t)}\, dG(z)\right) dF(x) \quad \text{by Fubini's theorem} \\ &= \int \psi(x-t) dF(x)\end{aligned}$$
∎

Corollary 13.8. *If G is the distribution function of a $N(0, \frac{1}{\sigma^2})$ random variable. Then $\psi(t) = e^{-\frac{1}{2\sigma^2}t^2}$, and so the above relationship becomes*
$$\int e^{izt}\varphi(z)\frac{\sigma}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2\sigma^2}\, dz = \int e^{-\frac{1}{2\sigma^2}(x-t)^2}\, dF(x).$$

Rearranging, we obtain

$$\frac{1}{2\pi}\int e^{-izt}\varphi(z)e^{-\frac{1}{2}z^2\sigma^2}\,dz = \int \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(x-t)^2}\,dF(x)$$

Then the right hand side is the density of the convolution of F and a $N(t,\sigma^2)$ distribution. Call the convolution distribution F_σ. Then

$$\begin{aligned} F_\sigma(\beta) - F_\sigma(\alpha) &= \int_\alpha^\beta \left(\frac{1}{2\pi}\int e^{-izt}\varphi(z)e^{-\frac{1}{2}z^2\sigma^2}\,dz\right)dt \\ &= \frac{1}{2\pi}\int \varphi(z)e^{-\frac{1}{2}z^2\sigma^2}\frac{e^{-iz\beta}-e^{-iz\alpha}}{-iz}\,dz \end{aligned}$$

If α and β are continuity points of F, then

$$F(\beta) - F(\alpha) = \lim_{\sigma\to 0}\frac{1}{2\pi}\int \varphi(z)e^{-\frac{1}{2}z^2\sigma^2}\frac{e^{-iz\beta}-e^{-iz\alpha}}{-iz}\,dz \tag{$\star$}$$

for as $\sigma \to 0$, $F_\sigma \to F$.

Since a function has only countably many points of discontinuity, we can then derive the following theorem.

Theorem 13.9. *Let X be a random variable with distribution function F and characteristic function φ. Assume*

$$\int |\varphi(t)|\,dt < \infty.$$

Then F has a bounded, continuous density f given by

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

Proof. From ($\star$) apply DCT. Then

$$\begin{aligned} F(\beta) - F(\alpha) = F(\beta) - F(\alpha) &= \lim_{\sigma\to 0}\int_\alpha^\beta \left(\frac{1}{2\pi}\int \varphi(z)e^{-\frac{1}{2}z^2\sigma^2}\,dz\right)dt \\ &= \int_\alpha^\beta \left(\frac{1}{2\pi}\int e^{-izt}\varphi(z)\,dz\right)dt \end{aligned}$$

$\square$

Corollary 13.10. *If $\varphi(t)$ is non-negative and integrable continuous function associated with a distribution function F. Then $\frac{\varphi(t)}{2\pi F'(0)}$ is a density function with characteristic function $\frac{F'(x)}{F'(0)}$.*

Proof. We have

$$\begin{aligned} F'(x) &= \frac{1}{2\pi}\int e^{-izx}\varphi(z)\,dz \\ &= \frac{1}{\pi}\int_0^\infty \cos(xz)\varphi(z)\,dz \quad \text{as } \varphi(z) \text{ is real} \end{aligned}$$

Thus

$$\begin{aligned} F'(0) &= \frac{1}{\pi}\int_0^\infty \varphi(z)\,dz \\ 1 &= \frac{1}{2F'(0)\pi}\int \varphi(z)\,dz \end{aligned}$$

and thus

$$\frac{F'(x)}{F'(0)} = \int \cos(xz)\frac{\varphi(z)}{2\varphi F'(0)}\,dz$$

$\square$

14. Lecture 14 - Thursday 14 April

15. Lecture 15 - Thursday 21 April

Example 15.1. X has density $f(x) = \frac{1}{2}e^{-|x|}$. Then

$$\begin{aligned} \varphi(t) &= \frac{1}{2}\int e^{itx}e^{-|x|}\,dx \\ &= \int_0^\infty \cos txe^{-x}\,dx \\ &= \int_0^\infty \frac{1}{2}\left(e^{itx}+e^{-itx}\right)e^{-x}\,dx \\ &= \frac{1}{2}\int_0^\infty e^{-x(1-it)} +^{-x(1+it)}\,dx \\ &= \frac{1}{2}\left[\frac{-1}{1-it}e^{-x(1+it)} + \frac{-1}{1+it}e^{-x(1+it)}\right]_0^\infty \\ &= \frac{1}{1+t^2} \end{aligned}$$

Thus $\varphi(t) = \frac{1}{1+t^2}$ which is a non-negative, integrable characteristic function. Thus,

$$\frac{\varphi(t)}{2\pi f(0)} = \frac{1}{\pi(1+t^2)}$$

which is the Cauchy distribution. We then know that the characteristic function of the Cauchy distribution is

$$\gamma(t) = \frac{F'(x)}{F'(0)} = \frac{f(x)}{f(0)} = e^{-|t|}$$

from Corollary 13.10.

Theorem 15.2 (Moment theorem)**.** *Let F be the distribution function of X. Assume X has finite moments up to order n, i.e. $\mathbb{E}(|X|^n) < \infty$. Then the characteristic function $\varphi(t)$ has uniformly continuous derivatives up to order n, and*

$$\varphi^{(k)}(t) = i^k\mathbb{E}(|X|^k), k = 1, 2, \ldots, n$$

and

$$\varphi(t) = 1 + \sum_{k=1}^{n} \mathbb{E}(X^k)\frac{(it)^k}{k!} + o(t^n)$$

as $t \to 0$.

Conversely, if φ *can be written as*

$$\varphi(t) = 1 + \sum_{k=1}^{n} a_k \frac{(it)^k}{k!} + o(t^n)$$

as $t \to 0$*, then the associated density function has finite moments up to order* n *if* n *is even, and up to order* $n-1$ *if* n *is odd, with* $a_k = \mathbb{E}(|X|^k)$.

Proof.

Lemma 15.3. *For any* $t \in \mathbb{R}$,

$$\left| e^{it} - 1 - it \cdots - \frac{(it)^{n-1}}{(n-1)!} \right| \leq \frac{|t|^n}{n!}.$$

Proof. Taylor's Theorem. □

Suppose $\mathbb{E}(|X|^k) < \infty$ for $k = 1, 2, \ldots, n$. Then

$$|x^k e^{itx}| \leq |x|^k$$

, so

$$\int x^k e^{itx}\, dF(x)$$

exists. Now

$$\begin{aligned} \frac{\varphi(t+h) - \varphi(t)}{h} &= \left| \int \frac{e^{i(t+h)x} - e^{itx}}{h}\, dF(x) \right| \\ &= \left| \int e^{itx} \cdot \frac{e^{ihx} - 1}{h}\, dF(x) \right| \\ &\leq \int |x|\, dF(x) < \infty \end{aligned}$$

from Lemma 15.3.

So by DCT,

$$\varphi'(t) = \lim_{h \to 0} \frac{\varphi(t+h) - \varphi(t)}{h} = i \int x e^{itx}\, dF(x)$$

and thus

$$\varphi'(0) = i\mathbb{E}(X).$$

Using induction, we obtain

$$\varphi^{(k)}(t) = i^k \int x^k e^{itx}\, dF(x)$$

and $\varphi^{(k)}(0) = i^k \mathbb{E}(X^k)$ for $k = 1, 2, \ldots, n$.

Arguing as in the proof of characteristic functions uniform continuity.

Expanding $\varphi(t)$ about $t = 0$ in a Taylor series, we have

$$\varphi(t) = 1 + \sum_{k=1}^{n} \varphi^{(k)}(0) \frac{[}{t}^{k}]k! + R_n(t), t > 0.$$

with

$$R_n(t) = \frac{t^n}{n!} \left[\varphi^{(n)}(\theta t) - \varphi^{(n)}(0)\right], 0 < \theta < 1.$$

We then have

$$\begin{aligned} \left|\frac{R_n(t)}{t^n}\right| &\leq \frac{1}{n!} \int |x|^n |e^{i\theta t x} - 1| \, dF(x) \\ &\leq \frac{2}{n!} \int |x|^n \, dF(x). \end{aligned}$$

and so by the DCT,

$$\lim_{t \to 0} \left|\frac{R_n(t)}{t^n}\right| = 0,$$

and thus $R_n(t) = o(t^n)$.

Conversely, suppose φ has an expansion up to order $2k$. Then φ has a finite derivative of order $2k$ at $t = 0$ Then

$$\begin{aligned} -\varphi^{(2)}(0) = -\lim_{h \to 0} \frac{\varphi(h) - 2\varphi(0) - \varphi(-h)}{h^2} & \\ &= \lim_{h \to 0} 2 \int \frac{1 - \cos hx}{x^2} \, dF(x) \\ &\geq 2 \int \lim_{h \to 0} \frac{1 - \cos hx}{h^2} \, dF(x) \text{ by Fatau} \\ &= \int x^2 dF(x) = E(X^2) \end{aligned}$$

and so $\varphi^{(2)}(0) < \infty \Rightarrow \mathbb{E}(X^2) < \infty$.

Using induction, assume finite $2(k-1)^{\text{th}}$ derivative at $0 \Rightarrow \mathbb{E}(X^{2(k-1)}) < \infty$. Then from the first part,

$$\varphi^{(2(k-2))}(t) = (-1)^{k-1} \int x^{2k-2} e^{itx} \, dF(x)$$

Suppose $\varphi^{2k}(0) < \infty$. Then let

$$G(x) = \int_{-\infty}^{x} y^{2k-2} \, dF(y).$$

so $\frac{G(x)}{G(\infty)}$ is a distribution function with characteristic function

$$\begin{aligned}\psi(t) &= \frac{1}{G(\infty)} \int e^{itx} x^{2k-2}\, dF(x) \\ &= \frac{(-1)^{k-1}\varphi^{(2k-2)}(t)}{G(\infty)}\end{aligned}$$

As $\varphi^{(2k-2)}(t)$ is twice differentiable at $t = 0$. So

$$\psi^{(2)}(0) \geq \int y^2 y^{2k-2}\, \frac{dF(y)}{G(\infty)}$$

and thus $\mathbb{E}(X^{2k}) < \infty$. as required. □

16. Lecture 16 Thursday 21 April

Corollary 16.1. *Let φ be a characteristic function associated with a random variable X. Then φ has continuous derivatives of all orders if and only if X has finite moments of all orders.*

Corollary 16.2. *The function $\varphi(t) = e^{-|t|^\alpha}$ is not a characteristic function if $\alpha > 2$. Note that $\alpha = 1$ was the Cauchy distribution, $\alpha = 2$ is the Normal distribution.*

Proof. If $\alpha > 2$ then

$$\lim_{t \to 0} \varphi^{(2)}(t) = 0 \Rightarrow \mathbb{E}(X^2) = 0$$

which implies X is degenerate. But if X is degenerate at b, then

$$\varphi(t) = e^{itb} \neq e^{-|t|^\alpha}$$

Thus by uniqueness of characteristic functions, $e^{-|t|^\alpha}$ is not a characteristic function. □

16.1. **Lattice distributions.**

Theorem 16.3 (Lattice distributions)**.** *Let X be a random variable with distribution function F, characteristic function φ. If $c \neq 0$ then the following are equivalent.*

(i) X has a lattice distribution whose range is continued in $0, \pm b, \pm 2b, \ldots,\ b = \frac{2\pi}{c}$.
(ii) $\varphi(t + nc) = \varphi(t)$ for $n = \pm 1, \pm 2, \ldots,$ that is, φ is periodic with period c.
(iii) $\varphi(c) = 1$.

Proof. (1) $\Rightarrow$ (2).

$$\begin{aligned}\varphi(t) &= \sum_{k=-\infty}^{\infty} P(X = kb) e^{itkb} \\ &= \sum_{k=-\infty}^{\infty} P(X = kb) e^{2\pi itk/c}\end{aligned}$$

which implies

$$\varphi(t+nc)=\varphi(t)$$

as $e^{2\pi inck/c}=1$.

(2) $\Rightarrow$ (3). Simply set $t=0, n=1$. Then $\varphi(0)=\varphi(c)=1$.

(3) $\Rightarrow$ (1).

$$1-\mathbb{E}(\cos cX)=0$$
$$\mathbb{E}(1-\cos cX)=0$$

but as $1-\cos cX\geq 0$, X must have probability components on points where $\cos cX=1$, that is, cX takes on the values $0, \pm\pi, \pm 2\pi, \dots$. □

Corollary 16.4. *X is degenerate if and only if $|\varphi(t)|=1$ for all t.*

Proof. If $P(X=b)=1$, then $\varphi(t)=e^{itb}$, and so $|\varphi(t)|=1$ for all t.

If $|\varphi(c)|=1$ for $c\neq 0$, then $\varphi(c)=e^{i\theta}$ for some θ. Let $\varphi_1(t)=\varphi(t)e^{-i\theta t/c}$ is characteristic function of $X-\frac{\theta}{c}$. Then $\varphi_1(c)=1$, thus $X-\frac{\theta}{c}$ is a lattice taking values in $0, \pm\frac{2\pi}{c}, \pm\frac{4\pi}{c}, \dots$.

Now, pick some $b\in\mathbb{R}$ with $\frac{b}{c}$ irrational. Then $|\varphi(b)|=1$, and then $X-a_2$ is a lattice taking values in $0, \pm\frac{2\pi}{b}, \pm\frac{4\pi}{b}, \dots$. Then

(i) $|\varphi(t)|<1$ for $t\neq 0$ (e.g. Normal, $e^{-\frac{1}{2}t^2}$).
(ii) $|\varphi(\lambda)|=1$ and $|\varphi(t)|<1$ on $0<t<\lambda$ (e.g. discrete ± 1, $\cos t$).
(iii) $|\varphi(t)|=1\forall t$, degenerate distributions.

□

Example 16.5. We can construct 3 nontrivial distribution functions $\varphi_1, \varphi_2, \varphi_3$ such that

(i) $\varphi_1(t)=\varphi_2(t), \forall t\in[-1,1]$.
(ii) $|\varphi(t)|=|\varphi_3(t)|, \forall t$.

Consider $g(x)=1-|x|, x\in[-1,1]$. This has characteristic function $\varphi(t)=\frac{2(1-\cos t)}{t^2}$. But the characteristic function is positive and integrable, and so

$$\varphi_1(t)=\begin{cases}1-|t| & |t|\leq 1\\ 0 & |t|>1\end{cases}$$

is the characteristic function of the density

$$f(x)=\frac{1-\cos x}{\pi x^2}.$$

We can express $\varphi_1(t)$ as the trigonometric series,

$$\varphi_1(t)=1-|t|=\frac{1}{2}+\sum_{k=1}^{\infty}a_k\cos(k\pi t)$$

with

$$a_k = 2\int_0^1 (1-t)\cos(k\pi t)\,dt = \begin{cases} \frac{4}{k\pi^2} & k \text{ odd} \\ 0 & k \text{ even} \end{cases}$$

We can thus write

$$\varphi_1(t) = \frac{[}{1}]2 + \sum_{k=1}^{\infty} \frac{4}{(2k-1)^2\pi^2}\cos(2k-1)\pi t.$$

Let V be a random variable, with

$$P(V=0) = \frac{1}{2}, P(V=\nu) = \frac{2}{\nu^2}, \nu = \pm\pi, \pm 3\pi, \pm 5\pi, \dots$$

Then V is a lattice distribution, with characteristic function

$$\varphi_2(t) = \frac{1}{2} + \frac{4}{\pi^2}\left(\cos\pi t + \frac{\cos 3\pi t}{9} + \frac{\cos 5\pi t}{25} + \dots\right)$$

and thus $\varphi_1(t) = \varphi_2(t)$ on $[-1,1]$, but have different density functions.

Finally, let U be a lattice random variable with distribution

$$P(U = \pm\frac{(2k+1)\pi}{2}) = \frac{4}{\pi^2(2k+1)^2}, k = 0, 1, 2, \dots$$

Then U has a characteristic function $\varphi_3(t) = 2\left[\varphi_2(\frac{t}{2}) - \frac{1}{2}\right]$. Thus

$$|\varphi_3(t)| = |\varphi_2(t)| \quad \forall t.$$

17. Lecture 17 - Thursday 5 May

17.1. **Sequences of characteristic functions.**

Lemma 17.1 (Helly selection theorem)**.** *Given a sequence of distribution functions $\{F_n\}$ then there exists a sequence $\{n_k\}$ and a non decreasing right continuous function F such that*

$$F_{n_k}(x) \to F(x)$$

at all continuity points x of F.

Proof. First order the rationals to get a sequence $\{r_k\}$. From $\{F_n(r_1)\}$ we choose a subsequence $\{F_{n_{1k}}(r_1)\}$ which converges.

Now from the sequence $\{n_{1k}\}$ choose a subsequence $\{n_{2k}\}$ such that $\{F_{n_{2k}}(r_2)\}$ converges, etc.

Now let $n_k = n_{kk}$. Then for each rational number r, the limit $F_{n_k}(r)$ exists as $n \to \infty$. Define $L(R) = \lim F_{n_k}(r), r \in \mathbb{Q}$. Then $L(r)$ is non-decreasing and takes values in $[0,1]$. Let $F(x) = \inf_{r \le x} L(r)$. Then F is non-decreasing, and right continuous, and $F_{n_k}(x) \to F(x)$ for all $x \in \mathbb{Q}$ and at all points of continuity of F. □

Lemma 17.2 (Extended Helly-Bragg theorem)**.** *If a sequence of distribution functions $\{F_n\}$ converges to a function F at all continuity points of F and g is a* ***bounded, continuous, real valued***

function *then*

$$\int_{\mathbb{R}} g\,dF_n \to \int_{\mathbb{R}} g\,dF$$

Proof. Let $M = \sup_x |g(x)|$, and let a, b be continuity points of F. Then

$$\begin{aligned}\left|\int_{\mathbb{R}} g\,dF_n - \int_{\mathbb{R}} g\,dF\right| &\le \left|\int_{\mathbb{R}} g\,dF_n - \int_a^b g\,dF_n\right| + \left|\int_a^b g\,dF_n - \int_a^b g\,dF\right| + \left|\int_a^b g\,dF - \int_{\mathbb{R}} g\,dF\right| \\ &\le M[F_n(a) - F_n(-\infty) + F_n(\infty) - F_n(b)] + \left|\int_a^b g\,dF_n - \int_a^b g\,dF\right| \\ &\quad + M[F(a) - F(-\infty) + F(\infty) - F(b)]\end{aligned}$$

Since

$$F_n(a) \to F(a), F_n(b) \to F(b)$$

as a, b are continuity points, we can choose a, b large enough to make the 3rd term small ($< \frac{\epsilon}{3}$ for arbitrary $\epsilon > 0$), and then N large enough to make the first term small.

Now we deal with the middle term. Let $a = x_{0N} < x_{1N} < \cdots < x_{\nu_N, N} = b$ be a sequence of subdivisions of $[a, b]$, such that $\Delta_n \to 0$ (partition width) as $n \to \infty$. Then

$$g_N(x) = \sum_{\nu=1}^{\nu_N} g(x_\nu, N)\mathbf{1}_{\{x_{\nu-1,N} \le x \le x_{\nu,N}\}}$$

Then $\sup_{x\in[a,b]} |g_N(x) - g(x)| \to 0$ as $N \to \infty$ (as g is bounded and continuous.) Then by DCT we have

$$\int_a^b g\,dF_n = \lim_{N\to\infty} \int_a^b g_N\,dF_n$$
$$\int_a^b g\,dF = \lim_{N\to\infty} \int_a^b g_N\,dF$$

Next, we will show

$$\lim_{n\to\infty} \int_a^b g_N\,dF_n = \int_a^b g_N\,dF$$

Let $x_{\nu,N}$ be continuity points of F so

$$F_n(x_{\nu,N}) - F_n(x_{\nu-1,N}) \to F(x_{\nu,N}) - F(x_{\nu-1,N}).$$

Hence

$$\begin{aligned}\lim_{n\to\infty} \int_a^B g_N(x)\,dF_n &= \lim_{n\to\infty} \sum_{\nu=1}^{\nu_N} g(x_{\nu,N})(F_n(x_{\nu,N}) - F_n(x_{\nu-1,N})) \\ &= \int_a^b g_N(x)\,dF(x)\end{aligned}$$

If $M_N = \sup_{x\in[a,b]} |g_N(x) - g(x)|$, then

$$\left|\int_a^b g\,dF_N - \int_a^b g\,dF\right| \le \int_a^b |g-g_n|\,dF_n + \left|\int_a^b g_n\,dF_n - \int_a^b g_N\,dF\right| + \int_a^b |g-g_N|\,dF$$
$$\le M_N[F_n(b)-F_n(a)] + \left|\int_a^b g_N\,dF_n\right.$$
$$\left.\int_a^b g_N\,dF\right| + M_N[F(b)-F(a)]$$

Since $M_N \to 0$ as $N \to \infty$. Then choosing N large enough to make M_N small enough, for a large N fixed, N_2 say, we have

$$\left|\int_a^b g_{N_2}\,dF_n - \int_a^b g_{N_2}\,dF\right| \le \frac{\epsilon}{9}$$

The result then follows. □

Lemma 17.3. *Let $\{F_n\}$ be a sequence of distribution functions with associated characteristic function $\{\varphi_n\}$. Assume $\varphi_n(t) \to \varphi(t)$ as $n \to \infty$ for all $t \in \mathbb{R}$. Then there exists a non-decreasing right continuous function F such that $F_n(x) \to F(x)$ at all continuity points x of F.*

Proof. From Lemma 17.1 there exists a subsequence $\{n_k\}$ and a non-decreasing continuous function F such that $F_{n_k}(x) \to F(x)$ at all continuity points of F. Using Parseval's relation on $\{F_{n_k}, \varphi_{n_k}\}$, we have

$$\frac{1}{2\pi}\int_{\mathbb{R}} e^{-izt}\varphi_{n_k}(z)e^{-\frac{1}{2}\sigma^2z^2}\,dz = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-t}{\sigma}\right)^2}\,dF_{n_k}(x)$$

Let $k \to \infty$. Then the LHS becomes

$$\frac{1}{2\pi}\int_{\mathbb{R}} e^{-izt}\varphi(z)e^{-\frac{1}{2}\sigma^2z^2}\,dz$$

by the dominated convergence theorem.

The RHS becomes

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-t}{\sigma}\right)^2}\,dF(x)$$

by an application of Lemma 17.2. Thus φ determines F uniquely (as before), so the limit F must be the same for all convergent subsequences. □

Theorem 17.4 (Continuity theorem)**.** *Let $\{F_n\}$ be a sequence of distribution functions converging to a distribution function F at all continuity points x of F. This happens if and only if $\varphi_n(t) \to \varphi$ pointwise and φ is continuous in the neighbourhood of the origin. If this is the case then φ is the characteristic function associated with F, and is continuous everywhere.*

Proof. If $\{F_n\}$ converges to F, use Lemma 17.2, with $g(x) = \cos(xt) + \sin(xt)$. □

18. Lecture 18 - Thursday 12 May

Theorem 18.1. *Assume $F_n \to F$ at continuity points of F, and associated characteristic function $\varphi_n \to \varphi$ pointwise. If $\varphi_n \to \varphi$ and φ is continuous in a neighbourhood of 0, then $F_n \to F$ and F is distribution function associated with φ.*

Proof. From previous lemma, there exists a non-decreasing, right continuous non-negative function F such that $F_n \to F$. We need to show F is a distribution function, that is $F(+\infty) - F(-\infty) \geq 1$. By Parseval's relation, we have

$$\frac{\sigma}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-izt} \varphi(z) e^{-\frac{1}{2}\sigma^2 z^2}\, dz = \int_{\mathbb{R}} e^{-\frac{1}{2}\left(\frac{x-t}{\sigma}^2\right)}\, dF(x) \leq F(+\infty) - F(-\infty)$$

The left hand side is equal to

$$\mathbb{E}(e^{-iN_\sigma t}\varphi(N_\sigma))$$

where $N_\sigma \sim N(0, \frac{1}{\sigma^2})$. Since

$$\left|e^{-izt}\varphi(t)\right| \leq 1$$

Assume φ is continuous on $|t| < A$. Then

$$\begin{aligned}\mathbb{E}(e^{-iN_\sigma t}\varphi(N_\sigma)) &= \mathbb{E}(e^{-iN_\sigma t}\varphi(N_\sigma) \,|\, |N_\sigma| \geq A) \cdot P(|N_\sigma| \geq A) \\ &\quad + \mathbb{E}(e^{-iN_\sigma t}\varphi(N_\sigma) \,|\, |N_\sigma| < A) P(|N_\sigma| < A).\end{aligned}$$

The first term tends to zero as $\sigma \to \infty$, as $P(|N_\sigma| \geq A) \to 0$ on $|N_\sigma| < A$. Then the distribution function tends to

$$G(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

as $\sigma \to \infty$.

$$\lim_{\sigma \to \infty} \frac{\sigma}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-izt}\varphi(z) e^{-\frac{1}{2}\sigma^2 z^2}\, dz = \int_{\mathbb{R}} e^{-izt}\varphi(z)\, dG(z) = 1$$

by the extended Helly-Bragg theorem. $\square$

Corollary 18.2. *If X_n has distribution function F_n and characteristic function φ_n, and X has distribution function F and characteristic function φ. Then the following are equivalent.*

i) $F_n(x) \to F(x)$ *at all continuity points x of F.*
ii) $\varphi_n(t) \to \varphi(t)$ *for all t,*
iii) $\mathbb{E}(g(X_n)) \to \mathbb{E}(g(X))$ *for all real, bounded, continuous functions g.*

In these cases we write $X_n \xrightarrow{d} X$ (X_n converges in distribution to X)

Corollary 18.3. *Suppose $X_n \xrightarrow{d} X$. If h is any continuous real valued function, then $h(X_n) \xrightarrow{d} h(X)$.*

Proof. $X_n \xrightarrow{d} X$ if and only if $\mathbb{E}(g(X_n)) \to \mathbb{E}(g(X))$. Then $g(h(x))$ is real, bounded, and continuous. Then

$$\mathbb{E}(g(h(X_n))) \to \mathbb{E}(g(h(X))) \Rightarrow h(X_n) \xrightarrow{d} h(x)$$

for all g real, bounded, continuous. □

Theorem 18.4 (Slutsky's theorem). *IF $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} a$, then*

$$X_n + Y_n \xrightarrow{d} X + a$$

Proof. Given $\epsilon > 0$, choose x such that $x, x - a \pm \epsilon$ are continuity points of $F(x) = P(X \leq x)$. Then

$$\begin{aligned} P(X_n + Y_n \leq x) &= P(X_n + Y_n \leq x, |Y_n - a| > \epsilon) + P(X_n + Y_n \leq x, |Y_n - a| \leq \epsilon) \\ &\leq P(|Y_n - a| > \epsilon) + P(X_n \leq x - a + \epsilon) \\ P(X_n \leq x - a - \epsilon) &= P(X_n \leq x - a - \epsilon, |Y_n - a| > \epsilon) + P(X_n \leq x - a - \epsilon, |Y_n - a| \leq \epsilon) \\ &\leq P(|Y_n - a| > \epsilon) + P(X_n + Y_n \leq x) \end{aligned}$$

Taking limits as $n \to \infty$, we have

$$P(X \leq x - a - \epsilon) \leq \lim_{n\to\infty} P(X_n + Y_n \leq x) \leq P(X \leq x - a + \epsilon)$$

Since $x - a \pm \epsilon$ are continuity points of F, we have

$$\lim_{n\to\infty} P(X - n + Y_n \leq x) = P(X \leq x - a).$$ □

18.1. **Central limit theorem.**

Note (Notation). Let $X_1, X_2, \ldots$ are independent random variables with characteristic functions $\varphi_1, \varphi_2, \ldots$ and distribution functions $F_1, F_2, \ldots$. Let $\mathbb{E}(X_i) = 0, \operatorname{Var}(X_i) = \sigma_i^2 < \infty$, $i = 1, 2, \ldots$. Let

$$S_n = \sum_{i=1}^n X_i, \qquad s_n^2 = \operatorname{Var}(S_n) = \sum_{i=1}^n \sigma_i^2$$

Theorem 18.5 (Lindeberg conditions). *Let $\epsilon > 0$. Then*

$$\begin{aligned} L_n(\epsilon) &= \frac{1}{s_n^2} \sum_{i=1}^n \mathbb{E}\left(X_i^2 \mathbf{1}_{\{(\}} |X_i| > \epsilon s_n)\right) \\ &= \frac{1}{s_n^2} \sum_{i=1}^n \int_{|x| > \epsilon s_n} x^2 \, dF_i(x) \end{aligned}$$

*Then the **Lindeberg condition** is*

$$\forall \epsilon > 0, \quad L_n(\epsilon) \to 0 \quad \textit{as } n \to \infty$$

Example 18.6. Assume $\mathbb{E}(|X_i|^3) < \infty$. Then

$$\begin{aligned} L_n(\epsilon) &\leq \frac{1}{s_n^2}\sum_{i=1}^n \mathbb{E}(X_i^2 \frac{|X_i|}{\epsilon s_n}) \\ &= \frac{1}{\epsilon}\frac{1}{s_n^3}\sum_{i=1}^n \mathbb{E}(|X_i|^3) \end{aligned}$$

Theorem 18.7 (Liapounov's condition)**.**

$$\frac{1}{s_n^3}\sum_{i=1}^n \mathbb{E}(|X_i|^3) \to 0 \quad \textit{as } n \to \infty$$

From above, Liapounov's condition implies Lindeberg's condition.

Theorem 18.8 (Central limit theorem)**.** *If for all $\epsilon > 0$, $L_n(\epsilon) \to 0$ as $n \to \infty$, then*

$$\frac{S_n}{s_n} \xrightarrow{d} N(0,1)$$

Proof. Preliminaries.

(i) If $|a_k| <\leq 1$ and $|b_k| \leq 1$ for all k, then

$$\left|\prod_{i=1}^n a_i - \prod_{i=1}^n b_i\right| \leq \sum_{i=1}^n |a_i - b_i|$$

as $a_1a_2 - b_1b_2 = (a_1 - b_1)a_2 - b_1(a_1 - b_2)$ and use induction.

(ii) $|e^z - 1 - z| \leq \delta|z|, \delta > 0$, for $|z|$ sufficiently small.

It is sufficient to prove

$$\varphi_{S_n/s_n}(t) = \prod_{k=1}^n \varphi_k(t/s_n) \to e^{-\frac{1}{2}t^2}) \tag{‡}$$

for all t.

Now

$$\begin{aligned} |\varphi_k(t/s_n) - 1| &= \left|\int (e^{\frac{itx}{s_n}} - 1 - \frac{itx}{s_n}) dF_k(x)\right| \quad \text{as } \mathbb{E}(X_k) = 0 \\ &\leq \int \frac{t^2}{x^2} 2s_n^2 \, dF_k(x) \\ &= \frac{1}{2}\frac{\sigma_k^2}{s_n^2}t^2 \end{aligned} \tag{$\star$}$$

Now

$$\begin{aligned} \sigma_k^2 &= \mathbb{E}(X_k^2 \mathbf{1}_{\{|X_k| \leq us_n\}}) + \mathbb{E}(X_k^2 \mathbf{1}_{\{|X_k| > us_n\}}) \\ &\leq (us_n)^2 + s_n^2 L_n(u) \end{aligned}$$

Hence
$$\frac{\sigma_k^2}{s_n^2} \le u^2 + L_n(u)$$
and since there are no k on the RHS, we have
$$\max_{k\le n} \frac{\sigma_k^2}{s_n^2} \le u^2 + L_n(u)$$
By Lindenberg's condition, we have $L_n(u) \to 0$ as $n \to \infty$, and as u was arbitrary, we have
$$\max_{k\le n} \frac{\sigma_k^2}{s_n^2} \to 0$$
From Assignment 5, we know
$$\exp(\varphi_k(t) - 1)$$
is a characteristic function. Let $\delta \to 0$. Then
$$\begin{aligned}
\left|\exp(\sum_{k=1}^n(\varphi_k(t/s_n)) - 1) - \prod_{k=1}^n \varphi_k(t/s_n)\right| &\le \sum_{k=1}^n \left|e^{\varphi_k(t/s_n)-1} - \varphi_k(t/s_n)\right| \quad \text{by (i)} \\
&\le \delta \sum_{k=1}^n |\varphi_k(t/s_n) - 1| \quad \text{by (ii)} \\
&\le \frac{\delta t^2}{2} \sum_{k=1}^n \frac{\sigma_k^2}{s_n^2} \quad \text{by } (\star) \\
&= \frac{\delta t^2}{2}. \quad \text{if } n \text{ is sufficiently large}
\end{aligned}$$

By (‡), we must show
$$\sum_{k=1}^n (\varphi_k(t/s_n) - 1) + \frac{1}{2}t^2 \to 0$$
that is,
$$\sum_{k=1}^n \int \left(e^{itx/s_n} - 1 - \frac{itx}{s_n} + \frac{1}{2}\frac{t^2x^2}{s_n^2}\right) dF_k(x) \to 0 \tag{†}$$
The modulus of the integral in (†) is bounded by
$$\frac{1}{6}\left|\frac{tx}{s_n}\right|^3 \le u\frac{|t|^3x^2}{6s_n^2}$$
if $|x| \le us_n$ and
$$\frac{x^2t^2}{2s_n^2} + \frac{x^2t^2}{2s_n^2}$$
when $|x| > us_n$. Hence the integral of (†) is bounded above by
$$\frac{u|t|^3}{6} + \frac{t^2}{s_n^2}\sum_{k=1}^n \int_{|x|>us_n} x^2\, dF_k(x) = \frac{u|t|^3}{6} + L_n(u)t^2$$

as the integral is the Lindeberg's condition.

Given $t, \epsilon > 0$, choose u such that $\frac{u|t|^3}{6} < \frac{\epsilon}{2}$, and N_0 large enough such that $L_n(u)t^2 < \frac{\epsilon}{2}$ for $n > N_0$. So the left hand side of (†) is bounded above by ϵ, and so the result follows. $\square$

Theorem 18.9 (Partial converse of the central limit theorem)**.** *Suppose that $s_n \to \infty$ and $\frac{\sigma_n}{s_n} \to 0$ as $n \to \infty$. Then the Lindeberg condition is necessary for*

$$\frac{S_n}{s_n} \xrightarrow{d} N(0,1).$$

Proof. By assumption, given $\epsilon > 0$ there exists $N_1 > 0$ such that

$$\frac{\sigma_k}{\sigma_n} < \frac{\sigma_k}{\sigma_k} < \epsilon$$

for $N_1 \leq k \leq n$ as $s_n^2 \leq s_k^2 (k \leq n)$. We also have

$$\frac{\sigma_k}{s_n} < \epsilon, k = 1, 2, \ldots, N_1$$

for $n > N_1$ as $s_n^2 \to \infty$. Hence

$$\max_{1 \leq k \leq n} \frac{\sigma_k}{s_k} \to 0$$

as $n \to \infty$. Assume $\frac{S_n}{s_n} \xrightarrow{d} N(0,1)$. If (5) holds then this convergence is equivalent to (1) $\iff$ (3)($\Rightarrow$ (4)) as (1) $\iff$ (3) requiring (5), to ensure

$$\left| \varphi_k(\frac{t}{s_n}) - 1 \right|$$

can be made uniformly small.

The real part of (4),

$$\sum_{k=1}^{n} \int \left(\cos(\frac{xt}{s_n}) - 1 + \frac{x^2 t^2}{2s_n^2} \right) dF_k(x) \geq \sum_{k=1}^{n} \int_{|x > us_n} \left(\cos(\frac{xt}{s_n}) - 1 + \frac{x^2 t^2}{2s_n^2} \right) dF_k(x)$$

For any $u > 0$, choose t such that $\frac{x^2 t^2}{2s_n^2} - 2 > 0$ if $|x| > us_n$ (i.e. $t^2 > \frac{4}{n^2}$). Continuing, we have

$$\begin{aligned}
&\geq \sum_{k=1}^{n} \int_{|x|>us_n} \left(\frac{x^2 t^2}{2s_n^2} - 2 \right) dF_k(x) \\
&\geq \sum_{k=1}^{n} \int_{|x|>us_n} \left(\frac{x^2 t^2}{2s_n^2} - 2\frac{x^2}{u^2 s_n^2} \right) dF_k(x) \\
&= \left(\frac{t^2}{2} - \frac{2}{u^2} \right) \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{|x|>us_n} x^2 \, dF_k(x) \\
&= \left(\frac{t^2}{2} - \frac{2}{u^2} \right) L_n(u)
\end{aligned}$$

Thus $L_n(u) \to 0$ as $n \to \infty$, that is, Lindeberg's condition holds. $\square$

Corollary 18.10. *Let $X_1, X_2, \dots$ IID with $\mathbb{E}(X_1) = 0$, Var$(X_1) = \sigma^2$. Then*

$$\frac{S_n}{\sqrt{n\sigma^2}} \xrightarrow{d} N(0,1)$$

Let $\bar{X}_k = \frac{S_n}{n}$.

Proof. We have $s_n^2 = n\sigma^2$. For $\epsilon > 0$, we have

$$\begin{aligned} L_n(\epsilon) &= \frac{1}{n\sigma^2} \sum_{k=1}^{n} \mathbb{E}(X_k^2 \mathbf{1}_{\{|X_k| > \epsilon\sigma\sqrt{n}\}}) \\ &= \frac{1}{\sigma^2} \mathbb{E}(X_1^2 \mathbf{1}_{\{|X_1| > \epsilon\sigma\sqrt{n}\}}) \to 0 \end{aligned}$$

as $n \to \infty$ as $\mathbb{E}(X_1^2) < \infty$. □

19. Lecture 19 - Thursday 19 May

The central limit theorem is about distribution functions. It is not an automatic consequence that the derivatives (densities) converge.

If $\frac{S_n}{s_n}$ has density $f_n(x)$ we need further conditions to ensure $f_n(x) \to \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ as $n \to \infty$.

Theorem 19.1. *If X_i are IID with characteristic functions $\varphi(t)$ and $|\varphi(t)|$ is integrable then*

$$f_n(x) \to \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

Example 19.2 (Densities not converging)**.** Let X_i have density

$$f(x) = \frac{C}{x(\log x)^2}, \quad 0 < x < \frac{1}{2}.$$

Then $\mathbb{E}(X^2) < \infty$ but $\sum_{i=1}^n X_i$ has an unbounded density on $(0, \frac{1}{2})$.

19.1. Stable Laws.

Definition 19.3 (Stable distribution)**.** A distribution F is said to be stable if it is not concentrated at one point, and when X_1 and X_2 are independent with distribution function F and a_1, a_2 are arbitrary constants there exists some $\alpha > 0, \beta$ such that

$$\frac{\alpha_1 X_1 + \alpha X_2 - \beta}{\alpha}$$

has distribution function F.

Example 19.4. If X_1 has a characteristic function $\varphi(t)$ then

$$\begin{aligned} \alpha X_3 + \beta &= a_1 X_1 + a_2 X_2 \\ e^{i\beta t}\varphi(\alpha t) &= \varphi(a_1 t)\varphi(a_2 t) \end{aligned}$$

If $\varphi(t) = e^{-c|t|^\gamma}$, $0 < \gamma \leq 2$, then

$$\varphi(a_1 t)\varphi(a_2 t) = e^{-c(|a_1|^\gamma + |a_2|^\gamma)|t|^\gamma}.$$

As these distributions are symmetric, we have $\beta = 0$, and so setting $\alpha = (|a_1|^\gamma + |a_2|^\gamma)$. Thus distributions with characteristic functions of the form $e^{-c|t|^\gamma}$ are stable. Hence the Cauchy distribution is stable ($\gamma = 1$), and the normal distribution is stable ($\gamma = 2$).

Theorem 19.5. *If φ is the characteristic function of a symmetric random variable ($X \stackrel{d}{=} -X$) with a stable distribution then $\varphi(t) = e^{-c|t|^\gamma}$ for some $c > 0$, $\gamma \in (0, 2]$.*

Recall that a distribution is symmetric if and only if φ is real.

Partial. $\varphi(t)\varphi(t) = \varphi(\alpha t)$ used to show that $\varphi(t) \neq 0$. (Since $\varphi(0) = 1$ and $\varphi(t)$ is continuous).

Then build up properties of φ. □

Theorem 19.6 (Lèvy)**.** *Let $X_1, X_2, \dots$ be independent and identically distributed random variables with distribution functions G. Let $S_n = \sum_{i=1}^n X_i$. Suppose that there exists a sequence of constants (a_n, b_n) with $b_n > 0$, such that*

$$\frac{S_n - a_n}{b_n} \xrightarrow{d} X$$

where X is not a constant. Then X is stable.

Definition 19.7 (Domain of attraction)**.** If X has distribution function F then we say G is in the **domain of attraction** of F.

Corollary 19.8. *If X has finite variance then G is in the domain of attraction of the normal distribution.*

Corollary 19.9. *If G satisfies $\lim_{x\to\infty} x(1 - G(x)) = c > 0$ then G is in the domain of attraction of the Cauchy distribution, that is,*

$$x\,\mathbb{P}(X > x) \to c.$$

A necessary and sufficient condition to be in the domain of attraction for the Cauchy distribution is

$$1 - G(x) = P(X_1 > x) = \frac{L(x)}{x}$$

*where $L(x)$ is a **slowly varying function**. $L(x)$ is a slowly varying function if for all $C > 0$,*

$$\lim_{x\to\infty} \frac{L(Cx)}{L(x)} = 1.$$

For example, $L(x) = 1, L(x) = \log x$ are slowly varying functions.

Theorem 19.10. *All stable laws are absolutely continuous and the distribution functions have derivatives of all orders.*

Theorem 19.11. *The normal distribution is the only stable law with finite variance.*

Theorem 19.12. *It can be shown that the canonical form of the characteristic function of a stable law is*

$$\varphi(t) = \exp\left[i\gamma t - c|t|^\gamma \left\{1 + \frac{i\beta t}{|t|}\omega(t,\alpha)\right\}\right]$$

where

$$\gamma \in \mathbb{R}, \alpha \in (0,2], c > 0, |\beta| \leq 1, \omega(t,\alpha) = \begin{cases} \tan\frac{\pi\alpha}{2} & \alpha \neq 1 \\ \frac{2}{\pi}\log|t| & \alpha = 1 \end{cases}$$

If φ is real, then $\beta = \gamma = 0$.

20. Lecture 20 - Thursday 19 May

20.1. **Infinitely divisible distributions.** Consider a triangular array $\{X_{nk}\}_{k=1}^n$ where for each n, $X_{n1}, X_{n2}, \ldots, X_{nn}$ are independent random variables. We assume that the distribution are identically distributed for each n.

$$\begin{matrix} X_{11} & & \\ X_{21} & X_{22} & \\ X_{31} & X_{32} & X_{33} \\ \vdots & & & \ddots \end{matrix}$$

Example 20.1. Let $X_{nk} \sim B(1, p_n)$. Then $S_n = \sum_{k=1}^n X_{nk} \sim B(n, p_n)$. We know that if $np_n \to \lambda$ as $n \to \infty$, then

$$S_n \xrightarrow{d} \text{Poisson}(\lambda).$$

Note that the Poisson distribution is not continuous, nor is it stable. Consider X_1, X_2 Poisson distributed, and let $Y = 2X_1 + 3X_2$. Then Y is not in the Poisson family as $P(Y = 1) = 0$.

Definition 20.2 (Infinitely divisible)**.** A distribution function F is infinitely divisible if for every positive integer k, F is the k-fold convolution of some distribution G_k with itself.

Example 20.3. (1) The Poisson distribution is infinitely divisible, as

$$\varphi(t) = e^{\lambda(e^{it}-1)} = \left[e^{\frac{\lambda}{k}(e^{it}-1)}\right]^k$$

(2) Symmetric stable laws are infinitely divisible, as

$$\varphi(t) = e^{-c|t|^\alpha} = \left(e^{-\frac{c}{k}|t|^\alpha}\right)^k$$

Lemma 20.4. *Assume $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{d} Y$, $\{X_n\}, \{Y_n\}$ independent. Then*

$$X_n + Y_n \xrightarrow{d} X + Y.$$

Proof. X_n has a characteristic function $\varphi_n(t) \to \varphi(t)$. Y_n has characteristic function $\psi(t) \to \psi(t)$. Then

$$\varphi_n(t)\psi_n(t) \to \varphi(t)\psi(t).$$

□

Theorem 20.5. *Given the array* $\{X_{nk}\}$*, letting* $S_n = \sum_{k=1}^n X_{nk}$*. If* $P(S_n \leq x) \to F(x)$ *then* F *is infinitely divisible.*

Proof. Fix k. We must show that F is the k-fold convolution of some G_k. Let $n' = mk$, $m = 1, 2, \dots,$ and let

$$Y_i^{(m)} = X_{n',(i-1)m+1} + \cdots + X_{n',im}, \quad i = 1, \dots, k.$$

Then

$$S_{mn} = Y_1^{(m)} + \cdots + Y_k^{(m)}$$

and $Y_f^{(m)}$ are IID.

If $P(Y_1^{(m)} \leq x) \to G_k(x)$ as $m \to \infty$ then

$$G_k^{*k} = F$$

So we need to show that G_k is a well defined distribution. We have $Y_1^{(m)}$ is the sum of m iid random variables, and

$$H_m(x) = P(Y_1^{(m)} \leq x).$$

We need to ensure "no probability escapes to infinity." Given a convergent subsequence of distribution functions, we know that the limit satisfies $G_k(x), G_k$ right continuous, non-decreasing. We need to show $G(+\infty) = 1$. Suppose that there exits $\epsilon > 0$ such that for any $M > 0$ we can find a subsequence m'_n such that

$$P(|Y_1^{(m'_n)}| > M) > \epsilon$$

There is a subsequence of $\{m'_n\}$, $\{m''_n\}$ say, such that

$$P(Y_1^{m''_n} > M) > \frac{\epsilon}{2} \quad \text{or} \quad P(Y_1^{m''_n} < -M) > \frac{\epsilon}{2}$$

So

$$P(Y_1^{m''_n} + \cdots + Y_k^{m''_n} > kM) > \left(\frac{\epsilon}{2}\right)^k$$

and so $F(km) \leq 1 - \left(\frac{\epsilon}{2}\right)^k$ (modulo choosing continuity points kM of F). Now, since we know that our limiting distribution F is a proper distribution function, we obtain our contradiction (no such $\epsilon > 0$ exists).

Hence G_k is a proper distribution function, and so $G_k^{*k} = F$. □

Definition 20.6 (Compound Poisson distribution)**.** Let $X_1, X_2, \dots$ IID random variables. and let $N \sim \textsc{Poisson}(\lambda)$. Then let $S_N = X_1 + \cdots + X_N$. Then S_N has a compound Poisson distribution.

If X has characteristic function φ, then S_N has characteristic function

$$\begin{aligned}\mathbb{E}(e^{itS_N}) &= \sum_{n=0}^{\infty} \mathbb{E}(e^{itS_N} \mid N = n) P(N = n) \\ &= \sum_{n=0}^{\infty} \varphi(t)^n e^{-\lambda} \frac{\lambda^n}{n!} \\ &= e^{-\lambda} e^{\lambda \varphi(t)} \\ &= e^{\lambda(\varphi(t) - 1)}.\end{aligned}$$

The compound Poisson distribution is clearly infinitely divisible.

Theorem 20.7. *A distribution function F is infinitely divisible if and only if it is the weak limit of a sequence of distributions, each of which is compound Poisson.*

21. Lecture 21 - Thursday 26 May

Theorem 21.1. *A distribution function is infinitely divisible if and only if it is the weak limit (limit in distribution) of a sequence of distribution functions each of which is compound Poisson.*

Lemma 21.2. *The weak limit of a sequence of infinitely divisible distributions is infinitely divisible.*

Proof. Let $F_n(x)$ be a sequence of distribution functions that are infinitely divisible with

$$F_n(x) \to F(x)$$

at all continuity points x of F. Form an array $\{X_{nk}\}_{k=1}^n$ where for a given n, $X_{n1}, X_{n2}, \dots, X_{nn}$ are IID with distribution function ${}_nF_n(x)$, the n^{th} *root of* F_n. Then

$$S_n = \sum_{k=1}^{n} X_{nk}$$

has distribution function F_n.

We know $F_n(x) \to F(x)$ so from the previous result F is infinitely divisible as it is the limit of the row sums of a triangular arrow of row-wise infinitely divisible random variables. □

Lemma 21.3. *The characteristic function of an infinitely divisible distribution is never zero.*

Proof. If $\varphi(0) = 1$ and φ is continuous, without loss of generality assume φ is real (if not, consider $|\varphi|^2 = \varphi\overline{\varphi}$ which is real and infinitely divisible.)

Let $\varphi_k(t)^k = \varphi(t)$. Assume $\varphi(t) > 0$ for $|t| \leq a$. Then for $t \in (-a, a)$, $\varphi_k(t) \to 1$ as $k \to \infty$.

Now note that

$$1 - \varphi(2t) \leq 4(1 - \varphi(t)), \qquad (\star)$$

as

$$\begin{aligned}
1-\varphi(2t) &= \int (1-\cos 2tx\, dF(x) \quad \text{as } \varphi \text{ is real} \\
&= \int (2-2\cos^2 tx)\, dF(x) \quad \cos 2\theta = 2\cos^2\theta - 1 \\
&= 2\int (1-\cos tx)(1+\cos tx)\, dF(x) \\
&\leq 4\int (1-\cos tx)\, dF(x) \quad 1-\cos tx \geq 0 \\
&= 4(1-\varphi(t))
\end{aligned}$$

as required.

Then we have $1-|\varphi(2t)| \leq 1-\varphi(2t)|^2 \leq 4(1-|\varphi(t)|^2) \leq 8(1-|\varphi(t)|)$. If $\varphi(t) \neq 0$ on $0 < t < a$ and $\epsilon > 0$ arbitrary, we can find k large enough such that

$$1-|\varphi_k(t)| < \frac{\epsilon}{8}$$

which implies $1-|\varphi_k(2t)| < \epsilon$ and so $|\varphi_k(2t)| \neq 0$ on $|t| < a$. So $\varphi_k(t)| \neq 0$ on $|t| < 2a$, and hence $|\varphi(t)| \neq 0$ on $|t| < 2a$.

Iterating this argument, we have that $|\varphi(t)| > 0$ for all t. □

Lemma 21.4. *For each k, let φ_k be a characteristic function such that $\varphi_k^k(t) = \varphi(t)$. $\varphi(t)$ is a characteristic function of an infinitely divisible distribution. Then $lim_{k\to\infty}\varphi_k(t) = 1$ for all t.*

Proof. Since φ is continuous and $\varphi(0) = 1$, we have

$$|\varphi_k(t)| = |\varphi(t)|^{1/k} \to 1$$

as $k \to \infty$.

We have $k \arg\varphi_k(t) = \arg\varphi(t) + 2\pi j, j = 0, 1, \ldots, k-1$. Since

$$\begin{aligned}
\arg\varphi_k(0) = \arg(1) = 0 \quad &\text{so } j = 0 \\
\arg\varphi_k(t) = \frac{1}{k}\arg\varphi(t) &\to 0
\end{aligned}$$

as $k \to \infty$, and so $\varphi_k(t) \to 1$ as $k \to \infty$. □

Proof of theorem. Let φ be the characteristic function of an infinitely divisible distribution F. Let $\varphi_k^k(t) = \varphi(t)$. Then

$$\begin{aligned}
\log\varphi(t) &= k\log\varphi_k(t) \\
&= k\log(1-(1-\varphi_k(t)))
\end{aligned}$$

Since $1-\varphi_k(t)\to 0$ as $k\to\infty$, we have

$$\begin{aligned}\log\varphi(t) &= -k[1-\varphi_k(t)+\frac{(1-\varphi_k(t))^2}{2}+\dots]\\ &= -k[1-\varphi_k(t)](1+\frac{1-\varphi_k(t)}{2}+\dots)\\ &= -k[1-\varphi_k(t)]+o(1)\end{aligned}$$

and so $\varphi(t)\sim e^{-k(1-\varphi_k(t))}$ which is a compound Poisson characteristic function. $\square$

Example 21.5. Show that the $U([-1,1])$ distribution is not infinitely divisible. This has associated characteristic function $\frac{\sin t}{t}$. Then $\varphi(\frac{\pi}{2})=0$, and so the distribution is not infinitely divisible.

22. Exam material

- Borel-Cantelli lemma.
- Martingales, central limit theorems, strong law of large numbers.
- Inequalities of random variables.

Example 22.1 (Q2b) of 2010 Exam)**.** Let (X_j) be IID. Then

$$\mathbb{E}|X_1|<\infty \iff \mathbb{P}(|X_n|\geq n \text{ i.o })=0$$

We have

$$\begin{aligned}\mathbb{E}|X_1|<\infty &\iff \sum_{j=1}^{\infty}\mathbb{P}(|X_1|\geq j)<\infty\\ &\iff \sum_{j=1}^{\infty}\mathbb{P}(|X_j|\geq j) \quad \text{by IID}\\ &\iff \mathbb{P}(|X_j|\geq j \text{ i.o })=0\end{aligned}$$

by Borel-Cantelli lemma.

Example 22.2 (Q7 of 2010 Exam)**.** Let $\{X_n\}$ be a sequence of IID random variables on a probability space $(\Omega,\mathcal{F},P)$ with

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

Let $\mathcal{F}_n=\sigma(X_1,X_2,\dots,X_n)$ and let $\{B_n\}$ be a sequence of events with $B_n\in\mathcal{F}_n$, satisfying

$$B_1=\Omega, \lim_{n\to\infty}P(B_n)=0, P(\limsup B_n)=1.$$

Define $Y_1=0$ and

$$Y_{n+1}=Y_n(1+X_{n+1})+\mathbf{1}_{B_n}X_{n+1}, n=1,2\dots.$$

(a) Show that $\{Y_n\}$ is a martingale.
(b) Show that Y_n converges in probability to 0.

(c) Show that $\limsup B_n \subseteq \limsup\{Y_n \neq 0\}$ and hence show that $\{Y_n\}$ does not converge almost surely.

Proof.

(a) Note that Y_1 is $\mathcal{F}_1$-measurable. By induction, we have that $Y_n + 1$ is $\mathcal{F}_{n+1}$-measurable.

We have

$$\mathbb{E}|Y_{n+1}| \leq 2\mathbb{E}|Y_n| + P(B_n) \quad \text{as } |X_{n+1} \leq 1$$

as $\mathbb{E}|Y_1| = 0$, $P(B_n) \leq 1$, so by induction, $\mathbb{E}|Y_n| < \infty$ for all n.

Finally,

$$\begin{aligned}\mathbb{E}(Y_{n+1} \,|\, \mathcal{F}_n) &= Y_n\mathbb{E}(1 + X_{n+1} \,|\, \mathcal{F}_n) + \mathbf{1}_{B_n}\mathbb{E}(X_{n+1} \,|\, \mathcal{F}_n) \\ &= Y_n \quad \text{as } \mathbb{E}(X_{n+1} \,|\, \mathcal{F}_n) = \mathbb{E}(X_{n+1}) = 0.\end{aligned}$$

Hence Y_n is a martingale.

(b) Let $\epsilon > 0$. We must show $P(|Y_n| > \epsilon) \to 0$ as $n \to \infty$. Consider $P(Y_{n+1} \neq 0)$. We have

$$\begin{aligned}P(Y_{n+1} \neq 0) &\leq P(B_n \text{ occurs or } Y_n \neq 0 \text{ and } X_{n+1} = 1) \\ &= P(B_n) + \frac{1}{2}P(Y_n \neq 0).\end{aligned}$$

Hence

$$\lim_{n\to\infty} P(Y_{n+1} \neq 0) \leq 2 \lim_{n\to\infty} P(B_n) = 0$$

and so $Y_n \xrightarrow{p} 0$.

(c) If $Y_n \xrightarrow{a.s.} Y$ almost surely then by uniqueness of limits in probability $Y = 0$ almost surely. We have

$$Y_{n+1} = \begin{cases} 2Y_n + \mathbf{1}_{B_n} & X_{n+1} = 1 \\ -\mathbf{1}_{B_n} & X_{n+1} = -1 \end{cases}$$

Hence $B_n \subseteq \{\omega : Y_{n+1}(\omega) \neq 0\}$. Thus

$$\begin{aligned}\limsup B_n = \bigcap_m^{\infty} \bigcup_{n=m}^{\infty} B_n \subseteq \bigcap_m^{\infty} \bigcup_{n=m}^{\infty} \{Y_{n+1} \neq 0\} \\ = \limsup\{Y_n \neq 0\}\end{aligned}$$

Hence $1 = P(\limsup B_n) \leq P(\limsup\{Y_n \neq 0\})$ and so $P(Y_n \neq 0 \text{ i.o.}) = 1$, and so Y_n does not converge almost surely.

□