Lecture Notes in Probability

Kaizhao Liu

April 2022

Contents

1 Independence

1.1 Basic Definitions

Definition 1.1 (independence: events)**.** Two events A, B are independent if $P(AB) = P(A)P(B)$.

Definition 1.2 (independence: random variables)**.** Two random variables X, Y are independent if for all $C, D \in \mathcal{R}$, the events $A = \{X \in C\}$ and $B = \{Y \in D\}$ are independent.

Definition 1.3 (independence: σ-fields)**.** Two σ-fields $\mathcal{F}$ and $\mathcal{G}$ are independent if for all $A \in \mathcal{F}$ and $B \in \mathcal{G}$, the events A and B are independent.

Actually, the first definition is a special case of the second, which is a special case of the third. This can be summarized in the following theorem.

Theorem 1.1.
(i) If A and B are independent, then so are A^c and B , A and B^c, and A^c and B^c.
(ii) Events A and B are independent if and only if 1_A and 1_B are independent.
(iii) If X and Y are independent then $\sigma(X)$ and $\sigma(Y)$ are.
(iv) If $\mathcal{F}$ and $\mathcal{G}$ are independent, $X \in \mathcal{F}$, and $Y \in \mathcal{G}$, then X and Y are independent.

We can extend this definition in an evident way for finitely many objects. Then, an infinite collection of objects is said to be independent if every finite subcollection is.

1.2 Sufficient Conditions for Independence

Theorem 1.2 (π-λ theorem)**.** *If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system that contains $\mathcal{P}$, then $\sigma(\mathcal{P}) \subset \mathcal{L}$.*

2 LLN

2.1 Stochastic Orders

In calculus, two sequence of real numbers, $\{a_n\}$ and $\{b_n\}$, satisfy $a_n = O(b_n)$ if and only if $|a_n| \leq c\,|b_n|$ for all n and a constant c; and $a_n = o(b_n)$ if and only if $\frac{a_n}{b_n} \to 0$ as $n \to 0$.

Definition 2.1. Let $X_1, X_2, \cdots$ be random vectors and $Y_1, Y_2, \cdots$ be random variables defined on a common probability space.
(i) $X_n = O(Y_n)$ a.s. if and only if $P(\|X_n\| = O(|Y_n|)) = 1$.
(ii) $X_n = o(Y_n)$ a.s. if and only if $\frac{X_n}{Y_n} \to 0$ a.s..
(iii) $X_n = O_p(Y_n)$ if and only if for any $\epsilon > 0$, there is a constant $C_\epsilon > 0$ s.t. $\sup_n P(\|X_n\| \geq C_\epsilon\,|Y_n|) < \epsilon$.
(iv) $X_n = o_p(Y_n)$ if and only if $\frac{X_n}{Y_n} \to_p 0$.

2.2 WLLN

In this section we study convergence in probability and the laws of large numbers associated with this type of convergence.

Lemma 2.1 (moments and tails)**.** *Let $\xi > 0$ be an random variable with $E\xi \in (0, \infty)$. Then*

$$(1-r)^2 \frac{(E\xi)^2}{E\xi^2} \leq P(\xi > rE\xi) \leq \frac{1}{r}, \quad r > 0$$

Theorem 2.1 (convergence in L^p implies convergence in probability). *If $p > 0$, then*

$$E|Z_n|^p \to 0 \Longrightarrow Z_n \longrightarrow 0 \text{ in probability.}$$

Proof. $P(|Z_n| \geq \epsilon) \leq \frac{E|Z_n|^p}{\epsilon^p} \to 0$ □

Theorem 2.2 (L^2 weak law). *Let $X_1, X_2, ...,$ be uncorrelated random variables with $EX_i = \mu$ and $var(X_i) < C < \infty$. If $S_n = X_1 + \cdots + X_n$, then as $n \to \infty$, $\frac{S_n}{n} \longrightarrow \mu$ in L^2.*

Proof. $E(\frac{S_n}{n} - \mu)^2 = \text{var}(\frac{S_n}{n}) = \frac{1}{n^2}(\sum \text{var}(X_i)) \leq \frac{Cn}{n^2} \to 0$ □

2.3 Borel-Cantelli Lemmas

Borel-Cantelli lemmas are the ladders from convergence in probability to a.s. convergence if the sequence of events are not decreasing. If the sequence of events are decreasing, then convergence in probability is the same as a.s. convergence, and there is no need for Borel-Cantelli lemma.

Theorem 2.3 (Borel-Cantelli lemma). $\sum_{n=1}^{\infty} P(A_n) < \infty \Longrightarrow P(A_n \ i.o.) = 0.$

Proof. Let $N = \sum_{n=1}^{\infty} 1_{A_n}$. $EN < \infty$ implies $N < \infty$ a.s. □

Theorem 2.4 (The second Borel-Cantelli lemma). *If the events A_n are independent, then*

$$\sum_{n=1}^{\infty} P(A_n) = \infty \Longrightarrow P(A_n \ i.o.) = 1$$

Proof. Let $M < N < \infty$. $1-x \leq e^{-x}$ and independence imply $P(\bigcap_{n=M}^{N} A_n^c) = \prod_{n=M}^{N}(1-P(A_n)) \geq \exp(-\sum_{n=M}^{N} P(A_n)) \to 0$ as $N \to \infty$, so $P(\bigcup_{n=M}^{N} A_n) = 1, \forall M$. Therefore $P(\limsup A_n) = 1$. □

Theorem 2.5 (Kochen-Stone lemma). *Suppose $\sum_{n=1}^{\infty} P(A_n) = \infty$. If*

$$\limsup_{n\to\infty} \frac{(\sum_{k=1}^{n} P(A_k))^2}{(\sum_{1\leq i,j\leq n} P(A_i \cap A_k))} = \alpha > 0$$

then $P(A_n \ i.o.) \geq \alpha$.

Remark. *This is a generalization of 2.4.*

Theorem 2.6. *If $A_1, A_2, ...$ are pairwise independent and $\sum_{n=1}^{\infty} P(A_n) < \infty$, then* ♡

2.4 SLLN

2.5 0-1 Laws

Theorem 2.7 (Kolmogorov's 0-1 law). *If $X_1, X_2, \cdots$ are independent and $A \in \mathcal{T}$, then $P(A) = 0$ or 1.*

Proof. The key point is to show that A is independent of itself.
To show this, we can procede by two limiting steps. □

Theorem 2.8 (Hewitt-Savage 0-1 law). *If $X_1, X_2, \cdots$ are i.i.d. and $A \in \mathcal{E}$, then $P(A) = 0$ or 1.*

Lemma 2.2.

2.6 Convergence of Random Series

Theorem 2.9 (Kolmogorov's maximal inequality). *Suppose $X_1, \cdots, X_n$ are independent with $EX_i = 0$ and $Var(X_i) < \infty$. If $S_n = X_1 + \cdots + X_n$, then*

$$P(\max_{1\leq k\leq n} |S_k| \geq x) \leq \frac{Var(S_n)}{x^2}$$

Proof. There is a proof by Doob's inequality. □

Theorem 2.10. *Suppose $X_1, X_2, \cdots$ are independent and have $EX_n = 0$. If*

$$\sum_{n=1}^{\infty} Var(X_n) < \infty$$

then with probability one $\sum_{n=1}^{\infty} X_n(\omega)$ converges.

Proof. Let $S_N = \sum_{n=1}^{N} X_n$. From Kolmogorov's maximal inequality, we get

$$P(\max_{M \le m \le N} |S_m - S_M| > \epsilon) \le \epsilon^{-2} \text{Var}(S_N - S_M) = \epsilon^{-2} \sum_{n=M+1}^{N} \text{Var}(X_n)$$

Letting $N \to \infty$, we get

$$P(\sup_{M \le m} |S_m - S_M| > \epsilon) \le \epsilon^{-2} \sum_{n=M+1}^{\infty} \text{Var}(X_n)$$

If we let $w_M = \sup_{m,n \ge M} |S_m - S_n|$, then

$$P(w_M > 2\epsilon) \le P(\sup_{M \le m} |S_m - S_M| > \epsilon) \to 0 \quad \text{as } M \to \infty$$

As w_M decreases as M increases, $w_M \downarrow 0$ a.s.. But $w_M(\omega) \downarrow 0$ implies $S_n(\omega)$ is a Cauchy sequence and hence $\lim_{n\to\infty} S_n(\omega)$ exists. □

Theorem 2.11 (Kolmogorov's three series theorem)**.** *Let $X_1, X_2, \cdots$ be independent. Let $A > 0$ and let $Y_i = X_i 1_{|X_i| \le A}$. In order that $\sum_{n=1}^{\infty} X_n$ converges a.s., it is necessary and sufficient that:*
(i) $\sum_{n=1}^{\infty} P(|X_n| > A) < \infty$
(ii) $\sum_{n=1}^{\infty} EY_n$ converges
(iii) $\sum_{n=1}^{\infty} Var(Y_n) < \infty$

Proof. To prove sufficiency, let $\mu_n = EY_n$. By the above theorem, $\sum_{n=1}^{\infty}(Y_n - \mu_n)$ converges a.s.. Using (ii), $\sum_{n=1}^{\infty} Y_n$ converges a.s.. (i) and Borel-Cantelli lemma imply $P(X_n \ne Y_n \, i.o.) = 0$, so $\sum_{n=1}^{\infty} X_n$ converges a.s..
For necessity, if the sum of (i) is infinite, $P(|X_n| > A \, i.o.) > 0$ and $\lim_{m\to\infty} \sum_{n=1}^{m} X_n$ can not converge. Suppose next (i) is finite but the sum □

One of the advantage of the random series proof is that it provides estimates on the rate of convergence.

Theorem 2.12. *Let $X_1, X_2, \cdots$ be i.i.d. random variables with $EX_i = 0$ and $EX_i^2 = \sigma^2 < \infty$. Let $S_n = X_1 + \cdots + X_n$. If $\epsilon > 0$ then*

$$\frac{S_n}{\sqrt{n(\log n)^{1+\epsilon}}} \to 0 \;\; a.s.$$

The next result, show that when $E|X_1| = \infty$, $\frac{S_n}{a_n}$ cannot converge almost surely to a nonzero limit.

Theorem 2.13. *Let $X_1, X_2, \cdots$ be i.i.d. with $E|X_1| = \infty$ and let $S_n = X_1 + \cdots + X_n$. Let a_n be a sequence of positive numbers with $\frac{a_n}{n}$ increasing*

2.7 Large Deviations

Let $X_1, X_2, \cdots$ be i.i.d. and let $S_n = X_1 + \cdots + X_n$. We will investigate the rate at which $P(S_n \ge na) \to 0$ for $x > \mu = EX_i$.

Lemma 2.3. *If $\gamma_{m+n} \ge \gamma_m + \gamma_n$, then as $n \to \infty$, $\frac{\gamma_n}{n} \to \sup_m \frac{\gamma_m}{m}$.*

Theorem 2.14. $\gamma(x) = \lim_{n\to\infty} \frac{\log P(S_n \ge nx)}{n}$ *exists* ≤ 0.

Proof. Let $\pi_n = P(S_n \ge nx)$, then $\pi_{m+n} \ge P(S_m \ge mx, S_{n+m} - S_m \ge nx) = \pi_m \pi_n$. Therefore, letting $\gamma_n = \log \pi_n$, from the lemma we conclude the existence of the limit. □

Next we want to determine the limit function $h(x)$. To do this, we need to introduce the cumulant-generating function of a random varaible ξ.

$$\phi(t) = \log Ee^{t\xi}, \quad t \in \mathbb{R}$$

and the Legendre transform of ϕ, given by

$$\phi^*(x) = \sup_{t\in\mathbb{R}}(tx - \phi(t)), \quad x \in \mathbb{R}$$

Lemma 2.4. *$\phi(t)$ and $\phi^*(x)$ are convex.*

Proof. The convexity of $\phi(t)$ comes from Holder's inequality, and the convexity for $\phi^*(x)$ is a property of Legendre transform. □

3 CLT

3.1 Distributions

Definition 3.1 (distribution)**.** If X is a random variable, then X induces a probability measure on $\mathbb{R}$ called its **distribution**.

Remark. *The distribution of a random variable X is usually described by giving its **distribution function** $F(x) = P(X \leq x)$.*

Theorem 3.1 (properties of distribution fucntions)**.**
(i) F is nondecreasing
(ii) $\lim_{x\to-\infty} F(x) = 0, \lim_{x\to\infty} F(x) = 1$
(iii) F is right continuous
(iv) If $F(x-) = \lim_{y\to x^-}$, then $F(x-) = P(X < x)$
(v) $P(X = x) = F(x) - F(x-)$

Proof. Directly follows from the definitions and the inclusion of sets. □

Theorem 3.2. *If F satisfies (i),(ii),(iii) in 3.1, then it is the distribution function of some random variable.*

Proof. Let $\Omega = (0,1), \mathcal{F}$=the Borel sets, and P=Lebesgue measure. If $\omega \in (0,1)$, construct

$$X(\omega) = \sup\{y : F(y) < \omega\}$$

We need to show:

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

For $\{\omega : X(\omega) \leq x\} \supseteq \{\omega : \omega \leq F(x)\}$, observe if $\omega \leq F(x)$, then $X(\omega) \leq x$.

For $\{\omega : X(\omega) \leq x\} \subseteq \{\omega : \omega \leq F(x)\}$, observe if $\omega > F(x)$, then since F is right continuous, $\exists \epsilon > 0$ s.t. $F(x+\epsilon) < \omega$. Therefore, $X(\omega) \geq x + \epsilon > x$. □

3.2 weak convergence

Definition 3.2 (weak convergence: distribution functions)**.** A sequence of distribution functions F_n is said to **converge weakly** to a limit F if $F_n(y) \to F(y)$ for all y that are continuity points of F.

Remark. *Denoted by $F_n \Longrightarrow F$.*

Definition 3.3 (weak convergence: random variable)**.** A sequence of random variables X_n is said to **converge weakly (converge in distribution)** to a limit X_∞ if their distribution functions converge weakly.

Remark. *Denoted by $X_n \Longrightarrow X_\infty$.*

Theorem 3.3 (Skorokhod)**.** *If $F_n \Longrightarrow F_\infty$, then $\exists$ r.v. Y_n with distribution F_n s.t. $Y_n \longrightarrow Y_\infty$ a.s.*

Proof. As in the proof of 3.2, let $\Omega = (0,1), \mathcal{F}$=the Borel sets, and P=Lebesgue measure. If $\omega \in (0,1)$, construct

$$Y_n(\omega) = \sup\{y : F_n(y) < \omega\}$$

We want to show:

$$Y_n(x) \longrightarrow Y_\infty(x)$$

for all but a countable number of x.

We begin by identifying the exceptional set. Let $a_x = \sup\{y : F_\infty(y) < x\}, b_x = \inf\{y : F_\infty(y) > x\}$, and $\Omega_0 = \{x : (a_x, b_x) = \emptyset\}$. Then $\Omega - \Omega_0$ is countable. If $x \in \Omega_0$, then $F_\infty(y) < x$ for $y < Y_\infty(x)$ and $F_\infty(y) > x$ for $y > Y_\infty(x)$.

Now we show $\liminf_{n\to\infty} Y_n(x) \geq Y_\infty(x)$. Choose $y < Y_\infty(x)$ s.t. F_∞ is continuous at y. Then $F_\infty(y) < x$ and $F_n(y) \longrightarrow F_\infty(y)$, so $F_n(y) < x$ for n sufficient large, that is, $Y_n(x) \geq y$. This is true for all such y's so the result follows.

The reverse inequality $\limsup_{n\to\infty} Y_n(x) \leq Y_\infty(x)$ is true by symmetry. □

Theorem 3.4. $X_n \Longrightarrow X_\infty \Longleftrightarrow \forall$ *bounded continuous function* $g, Eg(X_n) \longrightarrow Eg(X_\infty)$

Proof. $\Longrightarrow$: By 3.3, let Y_n have the same distribution as X_n and converge a.s. Since g is continuous, $g(Y_n) \longrightarrow g(Y_\infty)$ a.s. so by the bounded convergence theorem $Eg(X_n) \longrightarrow Eg(X_\infty)$.

$\Longleftarrow$: construct a bounded and continuous function

$$g_{x,\epsilon}(y) = \begin{cases} 1 & y \leq x \\ 0 & y \geq x+\epsilon \\ \text{linear} & x < y < x+\epsilon \end{cases}$$

Therefore, $\limsup_{n\to\infty} P(X_n \leq x) \leq \limsup_{n\to\infty} Eg_{x,\epsilon}(X_n) = Eg_{x,\epsilon}(X_\infty) \leq P(X_\infty \leq x+\epsilon)$. Letting $\epsilon \to 0$ gives $\limsup_{n\to\infty} P(X_n \leq x) \leq P(X_\infty \leq x)$. The reverse inequality can be proved in the same way. □

Theorem 3.5 (continuous mapping theorem)**.**

Theorem 3.6. *TFAE:*
(i) $X_n \Longrightarrow X_\infty$
(ii) For all open sets G, $\liminf_{n\to\infty} P(X_n \in G) \geq P(X_\infty \in G)$
(iii) For all closed sets K, $\limsup_{n\to\infty} P(X_n \in K) \leq P(X_\infty \in K)$
(iv) For all Borel sets A *with* $P(X_\infty \in \partial A) = 0$, $\lim_{n\to\infty} P(X_n \in A) = P(X_\infty \in A)$

Theorem 3.7 (Helly's selection theorem)**.** *For every sequence* F_n *of distribution functions, there is a subsequence* $F_{n(k)}$ *and a right continuous nondecreasing function* F *s.t.* $F_{n(k)} \Longrightarrow_v F$.

Remark. *The limit may not be a distribution function. This type of convergence is called vague convergence.*

Proof. To construct the function F, we adopt the standard diagonal argument. Let $\{q_i\}$ be an enumeration of the rationals. Since $F_m(q_k) \in [0,1]$ is bounded for all m, there is a subsubsequence $m_k(i)$ that is a subsequence of $m_{k-1}(i)$ s.t. $F_{m_k(i)}(q_k) \longrightarrow G(q_k)$. Select the diagonal sequence $n(k) = m_k(k)$, then by construction, $F_{n(k)}(q) \longrightarrow G(q)$ for all rational q.

Now we need to consruct F from G. Let

$$F(x) = \inf\{G(q) : q \in \mathbb{Q}, q > x\}$$

then $F(x)$ is right continuous and nondecreasing.

Let x be a continuity point of F. Pick rational $s > x$ s.t. $F(x) \leq F(s) < F(x)+\epsilon$, then as $F_{n(k)}(s) \longrightarrow G(s) \leq F(s)$, for k sufficient large, we have $F_{n(k)}(x) \leq F_{n(k)}(s) < F(x)+\epsilon$. On the other hand, pick rational $r_1 < r_2 < x$ s.t. $F(x)-\epsilon < F(r_1) \leq F(r_2) \leq F(x)$, then as $F_{n(k)}(r_2) \longrightarrow G(r_2) \geq F(r_1)$, so $F_{n(k)}(x) \geq F_{n(k)}(r_2) > F(x)-\epsilon$ for k sufficient large. Thus as $\epsilon \to 0$, we have the weak convergence. □

Theorem 3.8. *Every subsequential limit is the distribution function of a probability measure* $\Longleftrightarrow$ *the sequence is* ***tight****, i.e.* $\forall\epsilon > 0, \exists M_\epsilon$ *s.t.*

$$\limsup_{n\to\infty} 1 - F_n(M_\epsilon) + F_n(-M_\epsilon) \leq \epsilon$$

Proof. First note that for vague convergence $0 \le F(x) \le 1$.

$\Longleftarrow$: Suppose the sequence is tight and $F_{n(k)} \Longrightarrow_v F$. Let $r < -M_\epsilon, s > M_\epsilon$ be continuity points of F, then $1-F(s)+F(r) = \lim_{k\to\infty} 1-F_{n(k)}(s)+F_{n(k)}(r) \le \limsup_{n\to\infty} 1-F_n(M_\epsilon)+F_n(M_\epsilon) \le \epsilon$. Letting $r \to -\infty$ and $s \to \infty$ gives $\limsup_{n\to\infty} 1-F(x)+F(-x) \le \epsilon$.

$\Longrightarrow$: Suppose F_n is not tight. Then there is an $\epsilon > 0$ and a subsequence $n(k) \to \infty$ s.t.

$$1 - F_{n(k)}(k) + F_{n(k)}(-k) \ge \epsilon$$

for all k. By passing to a further subsequence $F_{n(k_j)}$ we can suppose $F_{n(k_j)} \Longrightarrow_v F$. Let $r < 0 < s$ be continuity points of F. Then $1-F(s)+F(r) = \lim_{j\to\infty} 1-F_{n(k_j)}(s)+F_{n(k_j)}(r) \ge \liminf_{j\to\infty} 1-F_{n(k_j)}(k_j)+F_{n(k_j)}(-k_j) \ge \epsilon$. Letting $s \to \infty$ and $r \to -\infty$, we see that F is not the distribution function of a probability measure. □

Corollary 3.8.1. *If there is a $\varphi \ge 0$ s.t. $\varphi(x) \to \infty$ as $|x| \to \infty$ and*

$$\sup_n \int \varphi(x)\mathrm{d}F_n(x) = C < \infty$$

then F_n is tight.

Proof. $C \ge \int \varphi(x)\mathrm{d}F_n(x) \ge \inf_{|x|\ge M} \varphi(x)(F_n(-M)+1-F_n(M))$ □

Lemma 3.1. *If $X_n \longrightarrow X$ in probability, then $X_n \Longrightarrow X$. Conversely, if $X_n \Longrightarrow c$ where c is a constant, then $X_n \longrightarrow c$ in probability.*

Theorem 3.9 (slutsky)**.** *If $X_n \Longrightarrow X$ and $Y_n \Longrightarrow c$, where c is a constant, then:*
(i) $X_n + Y_n \Longrightarrow X + c$
(ii) $X_nY_n \Longrightarrow cX$

3.3 Characteristic Functions

Definition 3.4 (characteristic function)**.** If X is a random variable, we define its characteristic function by $\varphi(t) = Ee^{itX}$.

Theorem 3.10 (properties of ch.f.)**.** *All ch.f.s have the following properties:*
(i) $\varphi(0) = 1$
(ii) $\varphi(-t) = \overline{\varphi(t)}$
(iii) $|\varphi(t)| \le 1$
(iv) $\varphi(t)$ is uniformly continuous on $(-\infty, \infty)$
(v) $Ee^{it(aX+b)} = e^{itb}\varphi(at)$

Theorem 3.11. *If X_1 and X_2 are independent and have ch.f.'s φ_1 and φ_2, then $X_1 + X_2$ has ch.f. $\varphi_1(t)\varphi_2(t)$.*

Lemma 3.2. *If $F_1, \cdots, F_n$ have ch.f. $\varphi_1, \cdots, \varphi_n$ and $\lambda_i \ge 0$ have $\lambda_1 + \cdots + \lambda_n = 1$, then $\sum_{i=1}^n \lambda_i F_i$ has ch.f. $\sum_{i=1}^n \lambda_i \varphi_i$.*

Theorem 3.12 (Continuity theorem)**.** *Let μ_n, $1 \le n \le \infty$ be probability measures with ch.f. φ_n.*
(i) If $\mu_n \Longrightarrow \mu_\infty$, then $\varphi_n(t) \to \varphi_\infty(t)$ for all t.
(ii) If $\varphi_n(t)$ converges pointwise to a limit $\varphi(t)$ that is continuous at 0, then the associated sequence of distributions μ_n is tight and converges weakly to the measure μ with characteristic function φ.

The next result is useful for constructing examples of ch.f.'s.

Example 3.1 (Polya's distribution)**.**

$$\begin{array}{rl} \textit{Density} & \dfrac{1-\cos(x)}{\pi x^2} \\ \textit{Ch.f.} & (1-|t|)^+ \end{array}$$

Theorem 3.13 (Polya's criterion)**.** *Let $\varphi(t)$ be real nonnegative and have $\varphi(0) = 1$, $\varphi(t) = \varphi(-t)$, and φ is decreasing and convex on $(0, \infty)$ with $\lim_{t\downarrow 0} \varphi(t) = 1, \lim_{t\uparrow\infty} \varphi(t) = 0$. Then there is a probability measure ν on $(0, \infty)$, so that*

$$\varphi(t) = \int_0^\infty (1 - \left|\frac{t}{s}\right|^+)\nu(\mathrm{d}s)$$

and hence φ is a characteristic function.

3.4 The Moment Problem

Example 3.2 (Heyde(1963)). *Consider the lognormal density*

$$f_0(x) = \frac{1}{\sqrt{(2\pi)}}\frac{1}{x}\exp^{-\frac{(\log x)^2}{2}} 1_{x\geq 0}$$

and for $-1 \leq a \leq 1$ *let*

$$f_a(x) = f_0(x)(1 + a\sin(2\pi\log x))$$

We claim that f_a *is a density and has the same moment as* f_0

Example 3.3.

A usual sufficient condition for a distribution to be determined by its moments is:

Theorem 3.14. *If* $\limsup_{n\to\infty}\frac{\mu_{2n}^{\frac{1}{2n}}}{2n} = r < \infty$*, then there is at most one d.f.* F *with* $\mu_n = \int x^n \mathrm{d}F(x)$ *for all positive integers* n*.*

Proof. First we explain why the condition only consider $2n$. Let F be any d.f. with the moment μ_n and let $\nu_n = \int |x|^n \mathrm{d}F(x)$. The Cauchy-Schwarz inequality implies $\nu_{2n+1}^2 \leq \mu_{2n}\mu_{2n+2}$, so

$$\limsup_{n\to\infty}\frac{\nu_n^{\frac{1}{n}}}{n} = r < \infty$$

Next, we have

$$\left|e^{i\theta X}(e^{itX} - \sum_{m=0}^{n-1}\frac{(itX)^m}{m!})\right| \leq \frac{|tX|^n}{n!}$$

Taking expected value, we have

$$\left|\varphi(\theta+t) - \varphi(\theta) - t\varphi'(\theta) - \cdots - \frac{t^{n-1}}{(n-1)!}\varphi^{(n-1)}(\theta)\right| \leq \frac{|t|^n}{n!}\nu_n$$

So we see that for any θ,

$$\varphi(\theta+t) = \varphi(\theta) + \sum_{m=1}^{\infty}\frac{t^m}{m!}\varphi^{(m)}(\theta) \quad \forall\,|t| < \frac{1}{er}$$

Let G be another distribution with the given moments and ψ its ch.f.. Since $\psi(0) = \varphi(0) = 1$, it follows from the above equation and induction that $\psi(t) = \varphi(t)$ for $|t| \leq \frac{k}{3r}$ for all k, so the two ch.f. coincide and the distributions are equal. □

Here is an application.

Theorem 3.15 (Semi-Circle Law).

3.5 Central Limit Theorems

Theorem 3.16. *Let* $X_1, X_2, \cdots$ *be i.i.d. with* $EX_i = \mu$*,* $Var(X_i) = \sigma^2 \in (0,\infty)$*. If* $S_n = X_1 + \cdots + X_n$*, then*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \Longrightarrow \mathcal{N}(0,1)$$

Proof. WLOG suppose $\mu = 0$. $\varphi(t) = Ee^{itX_1} = 1 - \frac{\sigma^2t^2}{2} + o(t^2)$, so $Ee^{itS_n/\sigma n^{\frac{1}{2}}} = (1 - \frac{t^2}{2n} + o(\frac{1}{n}))^n$. The last quantity $\to e^{-\frac{t^2}{2}}$ as $n\to\infty$, and the conclusion follows from the continuity theorem. □

Theorem 3.17 (Lindeberg-Feller theorem). *For each* n*, let* $X_{n,m}, 1 \leq m \leq n$ *be independent random variables with* $EX_{n,m} = 0$*. Suppose*
(i) $\sum_{m=1}^n EX_{n,m}^2 \to \sigma^2 > 0$
(ii) $\forall\epsilon > 0$*,* $\lim_{n\to\infty}\sum_{m=1}^n E(|X_{n,m}|^2; |X_{n,m}| > \epsilon) = 0$
Then $S_n = X_{n,1} + \cdots + X_{n,n} \Longrightarrow \mathcal{N}(0,\sigma^2)$*.*

Proof. □

3.6 Local Limit Theorems

Local limit theorems are subtly different from central limit theorems. The story is this:

Example 3.4.

Definition 3.5 (lattice distribution)**.** A random variable has a lattice distribution if there are constant $b, h > 0$ so that $P(X \in b + h\mathbb{Z}) = 1$. The largest h for which the last statement holds is called the span of the distribution.

Theorem 3.18. *Let $\varphi(t) = Ee^{itX}$. Regarding to the relationship between $|\varphi(t)|$ and 1, there are only three possiblities.*
(i) $|\varphi(t)| < 1$ for all $t \neq 0$.
(ii) There is a $\lambda > 0$ so that $|\varphi(\lambda)| = 1$ and $|\varphi(t)| < 1$ for $0 < t < \lambda$. In this case, X has a lattice distribution with span $\frac{2\pi}{\lambda}$.
(iii) $|\varphi(t)|$ for all t. In this case, $X = b$ a.s. for some b.

Proof. □

Theorem 3.19 (LLT for the lattice case)**.** *Let $X_1, X_2, \cdots$ be i.i.d. with $EX_i = 0, EX_i^2 = \sigma^2 \in (0, \infty)$, and having a common lattice distribution with span h. If $S_n = X_1 + \cdots + X_n$ and $P(X_i \in b + h\mathbb{Z}) = 1$. We put*

$$p_n(x) = P(\frac{S_n}{\sqrt{n}} = x) \text{ for } x \in \mathcal{L}_n = \left\{ \frac{nb + hz}{\sqrt{n}} : z \in \mathbb{Z} \right\}$$

and

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

Then as $n \to \infty$,

$$\sup_{x \in \mathcal{L}_n} \left| \frac{\sqrt{n}}{h} p_n(x) - n(x) \right| \to 0$$

Proof. Recall the inversion formula for lattice r.v. Y with $P(Y \in a + \theta\mathbb{Z}) = 1$ and $\psi(t) = Ee^{itY}$:

$$P(Y = x) = \frac{\theta}{2\pi} \int_{-\frac{\pi}{\theta}}^{\frac{\pi}{\theta}} e^{-itx} \psi(t) \mathrm{d}t$$

Use this formula for $\frac{S_n}{\sqrt{n}}$ gives

$$\frac{\sqrt{n}}{h} p_n(x) = \frac{1}{2\pi} \int_{-\frac{\pi\sqrt{n}}{h}}^{\frac{\pi\sqrt{n}}{h}} e^{-itx} \varphi^n(\frac{t}{\sqrt{n}}) \mathrm{d}t$$

and we have

$$n(x) = \frac{1}{2\pi} \int e^{itx} e^{-\frac{\sigma^2 t^2}{2}}$$

Substracting the last two equations and doing some estimation gives

$$\left| \frac{\sqrt{n}}{h} p_n(x) - n(x) \right| \leq \int_{-\frac{\pi\sqrt{n}}{h}}^{\frac{\pi\sqrt{n}}{h}} \left| \varphi^n(\frac{t}{\sqrt{n}}) - e^{-\frac{\sigma^2 t^2}{2}} \right| \mathrm{d}t + \int_{\frac{\pi\sqrt{n}}{h}}^{\infty} e^{-\frac{\sigma^2 t^2}{2}} \mathrm{d}t$$

So we are left to estimate the integrals. □

3.7 Poisson Convergence

Theorem 3.20. *For each n let $X_{n,m}$, $1 \leq m \leq n$ be independent random variables with $P(X_{n,m} = 1) = p_{n,m}, P(X_{n,m} = 0) = 1 - p_{n,m}$. Suppose*
(i) $\sum_{m=1}^{n} p_{n,m} \to \lambda \in (0, \infty)$.
(ii) $\max_{1 \leq m \leq n} \to 0$.
If $S_n = X_{n,1} + \cdots + X_{n,n}$, then $S_n \Longrightarrow$ Poisson(λ).

Here is a second proof of this theorem which provides new insight.

Definition 3.6 (total variation distance)**.** The total variation distance between two measures on a countable set S. $\|\mu - \nu\| = \frac{1}{2}\sum_z |\mu(z) - \nu(z)|$.

Lemma 3.3. $\|\mu - \nu\| = \sup_{A\subset S} |\mu(A) - \nu(A)|$

Lemma 3.4. *$d(\mu,\nu) = \|\mu - \nu\|$ defines a metric on probability measures on $\mathbb{Z}$. furthermore*

Lemma 3.5. *Consider measures on $\mathbb{Z}$. Then $\|\mu_1 \times \mu_2 - \nu_1 \times \nu_2\| \le \|\mu_1 - \nu_1\| + \|\mu_2 - \nu_2\|$.*

Lemma 3.6. *Consider measures on $\mathbb{Z}$.Then $\|\mu_1 * \mu_2 - \nu_1 * \nu_2\| \le \|\mu_1 \times \mu_2 - \nu_1 \times \nu_2\|$.
Here $*$ stands for the convolution.*

3.8 Stable Laws

In this section, we will investigate the case $EX_1^2 = \infty$ and give necessary and sufficient conditions for the existence of constants a_n and b_n so that

$$\frac{S_n - b_n}{a_n} \Longrightarrow Y$$

where Y is nondegenerate.

Definition 3.7 (slowly varying)**.** L is said to be slowly varying if $\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1 \ \forall t > 0$.

Theorem 3.21. *Suppose $X_1, X_2, \cdots$ are i.i.d. with a distribution that satisfies:
(i) $\lim_{x\to\infty} \frac{P(X_1>x)}{P(|X_1|>x)} = \theta \in [0,1]$
(ii) $P(|X_1| > x) = x^{-\alpha}L(x)$ where $\alpha < 2$ and L is slowly varying
Let $S_n = X_1 + \cdots + X_n$, $a_n = \inf\left\{x : P(|X_1| > x) \le \frac{1}{n}\right\}$ and $b_n = nE(X_1 1_{(|X_1|\le a_n)})$.*

Definition 3.8. The distributions whose ch.f are given by the following family with parameters κ, α, b, c are called stable laws.

$$\exp\left(itc - b\,|t|^{\alpha}\left(1 + i\kappa \mathrm{sgn}(t) w_\alpha(t)\right)\right)$$

where $\kappa \in [-1, 1]$, $\alpha \in (0, 2)$,

$$w_\alpha(t) = \begin{cases} \tan(\frac{\pi}{2}\alpha) & \alpha \ne 1 \\ \frac{2}{\pi}\log|t| & \alpha = 1 \end{cases}$$

Theorem 3.22. *Y is the limit of $\frac{X_1+\cdots+X_k-b_k}{a_k}$ for some i.i.d. sequence X_i if and only if Y has a stable law.*

3.9 Infinitely Divisible Distributions

Definition 3.9. Z has an infinitely divisible distribution if for each n there is an i.i.d. sequence $Y_{n,1}, \cdots, Y_{n,n}$ so that $Z =_d Y_{n,1} + \cdots + Y_{n,n}$.

Theorem 3.23. *Z is a limit of sums of type $Z = X_{n,1} + \cdots + X_{n,n}$ if and only if Z has an infinitely divisible distribution.*

Proof. ☐

Theorem 3.24 (Levy-Khinchin Theorem)**.** *Z has an infinitely divisible distribution if and only if its characteristic function has*

$$\log\varphi(t) = ict - \frac{\sigma^2 t^2}{2} + \int (e^{itx} - 1 - \frac{itx}{1+x^2})\mu(\mathrm{d}x)$$

where μ is a measure with $\mu(\{0\}) = 0$ and $\int \frac{x^2}{1+x^2}\mu(\mathrm{d}x) < \infty$.

The theory of infinitely divisible distributions is simpler in the case of finite variance. In this case, we have:

Theorem 3.25 (Kolmogorov's Theorem)**.** *Z has an infinitely divisible distribution with mean 0 and finite variance if and only if its ch.f. has the form*

$$\log\varphi(t) = \int \frac{(e^{itx} - 1 - itx)}{x^2}\nu(\mathrm{d}x)$$

ν is called the canonical measure, and $Var(Z) = \nu(\mathbb{R})$.

3.10 Limit Theorems in $\mathbb{R}^d$

Theorem 3.26 (Convergence theorem). *Let $X_n, 1 \leq n \leq \infty$ be random vectors with ch.f. φ_n. A necessary and sufficient condition for $X_n \Longrightarrow X_\infty$ is that $\varphi_n(t) \to \varphi_\infty(t)$.*

Theorem 3.27 (Cramer-Wold Device). *A sufficient condition for $X_n \Longrightarrow X_\infty$ is that $\theta \cdot X_n \Longrightarrow \theta \cdot X_\infty$ for all $\theta \in \mathbb{R}^d$.*

3.11 Stein's method

There is a lack of calculus in probability theory. We have only used Fourier transform, i.e. the characteristic function. Stein invented a exotic way of using calculus to derive the convergence rate of CLT.

Definition 3.10.

Stein's method is related to Slepian's interpolation, which is in turn related to Lindeberg's telescopic interpolation.

4 HDP

The goal of HDP is to quantify the convergence rate of limit theorems such as CLT. So it is of non-asymptotic nature.

4.1 Concentration with Independence

Theorem 4.1 (Hoeffding's inequality). *Let $X_1, \cdots, X_n$ be i.i.d. symmetric Bernoulli random variable. Then*

Definition 4.1 (sub-gaussian random variable).

Definition 4.2 (sub-exponential random variable).

Lemma 4.1 (Hoeffding's lemma). *Assume $a \leq X \leq b$. Then $\phi(t) \leq \frac{t^2(b-a)^2}{8}$.*

Proof. WLOG assume $EX = 0$. Recall that $\phi(t) = \log E(e^{tX})$. Then

$$\phi'(t) = \frac{E(Xe^{tX})}{E(e^{tX})}, \quad \phi''(t) = \frac{E(X^2 e^{tX})}{E(e^{tX})} - (\frac{E(Xe^{tX})}{E(e^{tX})})^2$$

Let Q denote the distribution with $\frac{\mathrm{d}Q}{\mathrm{d}P} = \frac{e^{tX}}{Ee^{tX}}$, where P is the distribution of X. Then $\phi''(t) = \mathrm{Var}_Q(X) \leq \frac{(b-a)^2}{4}$. ☐

Lemma 4.2 (maximal inequality). *Assume that $X_1, \cdots, X_n$ be n sub-gaussian random variables*

$$E \max_{i \in [n]} X_i \leq \sqrt{2 \log n}$$

Proof. LogSumExp Trick. ☐

Remark. *The bound is sharp even though we do not assume independence.*

Example 4.1. *Let $X_1, \cdots, X_n$ be independent $N(0,1)$ random variables. Then*

4.2 Concentration without Independence

The approach to concentration inequality we developed so far relies crucially on independence of random variables. We now pursue some alternative approaches to concentration, which are not based on independence.

4.3 Kernel Trick

Theorem 4.2 (Grothendick's inequality). *Consider an $m \times m$ matrix (a_{ij}) of real numbers. Assume that for numbers $x_i, y_j \in \{0, 1\}$, we have*

$$\left| \sum_{i,j} a_{ij} x_i y_j \right| \leq 1$$

Then, for any Hilbert space H and any vector $u_i, v_j \in H$ satisfying $\|u_i\| = \|v_j\| = 1$, we have

$$\left| \sum_{i,j} a_{ij} x_i y_j \right| \leq K$$

where $K \leq 1.783$ is an absolute constant.

4.4 Decoupling

In the beginning of HDP, we studied independent random variables of the type $\sum_{i=1}^n a_i X_i$. Now we want to study quadratic forms of the type $\sum_{i,j=1}^n a_{ij} X_i X_j$ where X_i's are independent. Such a quadratic form is called a chaos in probability theory. It is harder to establish a concentration of a chaos. The main difficulty is that the terms of the sum are not independent. This difficulty can be overcome by the decoupling technique.

The purpose of decoupling is to replace the quadratic form with the bilinear form $\sum_{i,j=1}^n a_{ij} X_i X'_j$ where X' is an independent copy of X.

Theorem 4.3 (decoupling). *Let A be an $n \times n$ diagonal-free matrix. Let X be a random vector with independent mean zero coordinates X_i. Then for every convex function $F : \mathbb{R} \to \mathbb{R}$,*

$$EF(X^T AX) \leq EF(4X^T AX')$$

where X' is an independent copy of X.

Lemma 4.3. *Let Y and Z be independent random variables s.t. $EZ = 0$. Then for every convex function $F : \mathbb{R} \to \mathbb{R}$,*

$$EF(Y) \leq EF(Y + Z)$$

Proof. First condition on Y and use Jensen's inequality. Then take expecation w.r.t Y. □

Remark. *Intuitively, this lemma tells us, under some conditions, adding a mean zero disturbance increases the value.*

4.5 Symmetrization

In this section we develop the simple and useful technique of symmetrization. It allows one to reduce problems about arbitary distributions to symmetric distributions.

Definition 4.3 (Rademacher random variable).

Throuhout this section, we denote by $\xi_1, \xi_2, \cdots$ a sequence of independent Rademacher random variables. We assume that they are independent not only with each other, but also of any other random variable in question.

Lemma 4.4 (symmetrization). *Let $X_1, \cdots, X_N$ be independent, mean zero random vectors in a normed space. Then*

$$\frac{1}{2} E \left\| \sum_{i=1}^N \xi_i X_i \right\| \leq E \left\| \sum_{i=1}^N X_i \right\| \leq 2E \left\| \sum_{i=1}^N \xi_i X_i \right\|$$

Proof. The proof relies on introducing an independent copy X_i''s of X_i to symmetrize the expression, and noticing that $X =^d \xi X$ if X is symmetric.

$$\begin{aligned} E\left\|\sum_{i=1}^N X_i\right\| &\le E\left\|\sum_{i=1}^N (X_i - X_i')\right\| \\ &= E\left\|\sum_{i=1}^N \xi_i(X_i - X_i')\right\| \\ &= 2E\left\|\sum_{i=1}^N \xi_i X_i\right\| \end{aligned}$$

$$\begin{aligned} E\left\|\sum_{i=1}^N \xi_i X_i\right\| &\le E\left\|\sum_{i=1}^N \xi_i(X_i - X_i')\right\| \\ &= E\left\|\sum_{i=1}^N (X_i - X_i')\right\| \\ &\le 2E\left\|\sum_{i=1}^N X_i\right\| \end{aligned}$$

□

4.6 Chaining

This section should be moved to stochastic analysis notes because it considers a continuum of random variables. Chaining is a multi-scale version of the ϵ-net argument.

Lemma 4.5 (discrete Dudley's inequality)**.** *Let $(X_t)_{t\in T}$ be a mean zero random process on a metric space (T, d) with sub-gaussian increments. Then*

$$E\sup_{t\in T} X_t \le CK \sum_{k\in\mathbb{Z}} 2^{-k}\sqrt{\log \mathcal{N}(T, d, 2^{-k})}$$

In the single-scale ϵ-net argument, we discretize T by choosing an ϵ-net $\mathcal{N}$ of T. Then every point $t \in T$ can be approximated by a closest point from the net $\pi(t) \in \mathcal{N}$ with accuracy ϵ. The increment condition yields $\left\|X_t - X_{\pi(t)}\right\|_{\phi_2} \le K\epsilon$. This gives $E\sup_{t\in T} X_t \le E\sup_{t\in T} X_{\pi(t)} + E\sup_{t\in T}(X_t - X_{\pi(t)})$. The first term can be controlled by a union bound over $|\mathcal{N}| = \mathcal{N}(T, d, \epsilon)$ points $\pi(t)$. But for the second term, it is not clear how to control the supremum over $t \in T$.

5 Martingales

5.1 Conditional Expectation

Definition 5.1 (conditional expectation)**.** Given a probability space $(\Omega, \mathcal{F}_o, P)$, a σ-field $\mathcal{F} \subset \mathcal{F}_o$, and a random varaible $X \in \mathcal{F}_o$ with $E|X| < \infty$. The conditional expectation of X given $\mathcal{F}$ is any random variable Y that satisfies:
(i) $Y \in \mathcal{F}$
(ii) $\forall A \in \mathcal{F}, \int_A X\mathrm{d}P = \int_A \mathrm{d}P$.

Lemma 5.1. *If Y satisfies (i)&(ii), then it is integrable.*

Proof. Let $A = \{Y > 0\} \in \mathcal{F}$. We have $\int_A Y\mathrm{d}P = \int_A X\mathrm{d}P \le \int_A |X|\,\mathrm{d}P$ and $\int_{A^c} -Y\mathrm{d}P = \int_{A^c} -X\mathrm{d}P \le \int_{A^c} |X|\,\mathrm{d}P$, therefore we have $E|Y| \le |X|$. □

Theorem 5.1 (uniqueness of conditional expectation)**.** *The conditional expecation of X given $\mathcal{F}$ is unique, denoted by $E(X|\mathcal{F})$.*

Proof. Suppose Y' also satisfies (i)&(ii). Taking $A = \{Y - Y' \ge \epsilon > 0\}$, we see $0 = \int_A X - X\mathrm{d}P = \int_A Y - Y'\mathrm{d}P \ge \epsilon P(A)$ so $P(A) = 0$. Since this holds for all ϵ, we have $Y \le Y'$ a.s., and switching the role of $Y\&Y'$ gives the desiered result. □

Theorem 5.2 (existence of conditional expectation)**.** *$E(X|\mathcal{F})$ exists.*

Proof. The proof is based on Radon-Nikodym Theorem. Suppose first that $X \geq 0$. Construct a measure $\nu(A) = \int_A X \mathrm{d}P$ for $A \in \mathcal{F}$. Then $\nu \ll P$, so by Radon-Nikodym Theorem, there exists $Y \in \mathcal{F}$ satisfying $\nu(A) = \int_A Y \mathrm{d}P$.

To treat the general case, write $X = X^+ - X^-$, let $Y_1 = E(X^+|\mathcal{F})$ and $Y_2 = E(X^-|\mathcal{F})$, then verify condition (i)&(ii). □

Now we investigate the properties of conditional expectation.

Theorem 5.3.

Theorem 5.4. *If φ is convex and $E|X|, E|\varphi(X)| < \infty$, then*

$$\varphi(E(X|\mathcal{F})) \leq E(\varphi(X)|\mathcal{F})$$

Proof. □

Corollary 5.4.1. *Conditional expectation is a contraction in L^p, $p \geq 1$.*

Theorem 5.5. *If $\mathcal{F}_1 \subset \mathcal{F}_2$, then:*
(i) $E(E(X|\mathcal{F}_1)|\mathcal{F}_2) = E(X|\mathcal{F}_1)$
(ii) $E(E(X|\mathcal{F}_2)|\mathcal{F}_1) = E(X|\mathcal{F}_1)$

Proof. Directly follows from the definition. □

Remark. *This theorem shows that whatever the order of conditioning is, the result is always conditioning on the smallest σ-field.*

Theorem 5.6. *If $X \in \mathcal{F}$ and $E|Y|, E|XY| < \infty$, then*

$$E(XY|\mathcal{F}) = XE(Y|\mathcal{F})$$

Proof. Approximate X by the standard process as in the construction of Lebesgue integral. □

Theorem 5.7 (LSE)**.** *Suppose $EX^2 < \infty$. $E(X|\mathcal{F})$ is the variable $Y \in \mathcal{F}$ that minimizes $E(X-Y)^2$.*

5.2 Martingales

Definition 5.2 (filtration)**.** An increasing sequence of σ-fields is called a filtration.

Definition 5.3 (adapted)**.** A sequence X_n is said to be adapted to $\mathcal{F}_n$ if $X_n \in \mathcal{F}_n$ for all n.

Definition 5.4 (martingale)**.** If X_n is a sequence with:
(i) $E|X_n| < \infty$
(ii) X_n is adapted to $\mathcal{F}_n$
(iii) $E(X_{n+1}|\mathcal{F}_n) = X_n$ for all n
then X is said to be a martingale.

Remark. *If in (iii) is replaced by $\leq$ or $\geq$, then X is said to be a supermartingale or submartingale respectively.*

Theorem 5.8. *If X_n is a supermartingale, then for $n > m$, $E(X_n|\mathcal{F}_m) \leq X_m$.*
If X_n is a submartingale, then for $n > m$, $E(X_n|\mathcal{F}_m) \geq X_m$.
If X_n is a martingale, then for $n > m$, $E(X_n|\mathcal{F}_m) = X_m$.

Proof. By definition and induction. □

Theorem 5.9. *If X_n is a supermartingale w.r.t. $\mathcal{F}_n$ and φ is an increasing concave function with $E|\varphi(X_n)| < \infty$ for all n, then $\varphi(X_n)$ is a supermartingale w.r.t $\mathcal{F}_n$.*

If X_n is a submartingale w.r.t. $\mathcal{F}_n$ and φ is an increasing convex function with $E|\varphi(X_n)| < \infty$ for all n, then $\varphi(X_n)$ is a submartingale w.r.t $\mathcal{F}_n$.

If X_n is a martingale w.r.t. $\mathcal{F}_n$ and φ is a convex function with $E|\varphi(X_n)| < \infty$ for all n, then $\varphi(X_n)$ is a submartingale w.r.t $\mathcal{F}_n$.

Proof. Directly follows from the definition of martingale and Jensen's inequality. □

Corollary 5.9.1. *If X_n is a submartingale, then $(X_n - a)^+$ is a submartingale.*

Corollary 5.9.2. *If X_n is a supermartingale, then $X_n \wedge a$ is a supermartingale.*

Definition 5.5 (predictable)**.** Let $\mathcal{F}_n$, $n \geq 0$ be a filtration. H_n, $n \geq 1$ is said to be a predictable sequence if $H_n \in \mathcal{F}_{n-1}$ for all $n \geq 1$.

Example 5.1.

Theorem 5.10. *Let X_n, $n \geq 0$, be a supermartingale. If $H_n \geq 0$ is predictable and each H_n is bounded, then $(H \cdot X)_n = \sum_{m=1}^{n} H_m(X_m - X_{m-1})$ is a supermartingale.*

The same fact is true for submartingales and for martingales, while in the latter case we can relax the restriction $H_n \geq 0$.

Theorem 5.11 (Doob's decomposition)**.** *Any submartingale $X_n, n \geq 0$, can be written in a unique way as $X_n = M_n + A_n$, where M_n is a martingale and A_n is a predictable increasing sequence with $A_0 = 0$.*

Proof. We want $X_n = M_n + A_n, E(M_n|\mathcal{F}_{n-1}) = M_{n-1}$, and $A_n \in \mathcal{F}_{n-1}$. So we must have

$$\begin{aligned} E(X_n|\mathcal{F}_{n-1}) &= E(M_n|\mathcal{F}_{n-1}) + E(A_n|\mathcal{F}_{n-1}) \\ &= M_{n-1} + A_n \\ &= X_{n-1} - A_{n-1} + A_n \end{aligned}$$

So $A_n - A_{n-1} = E(X_n|\mathcal{F}_{n-1}) - X_{n-1}$. Since $A_0 = 0$, we have

$$A_n = \sum_{m=1}^{n} E(X_n - X_{n-1}|\mathcal{F}_{n-1})$$

The last step is to check what we have constructed above indeed satisfies the desired properties. $\square$

5.3 Stopping Times

Definition 5.6 (stopping time)**.** A random variable N is said to be a stopping time if $\{N = n\} \in \mathcal{F}_n$ for all n.

Corollary 5.11.1. *If N is a stopping time and X_n is a supermartingale, then $X_{N \wedge n}$ is a supermartingale.*

Proof. Let $H_n = 1_{N \geq n}$. Verify that H_n is predictable. It follows from the theorem that $(H \cdot X)_n = X_{N \wedge n} - X_0$ is a supermartingale. Thus $X_{N \wedge n}$ is a supermartingale as a sum of two supermartingale. $\square$

5.4 Almost Sure Convergence

Suppose X_n, $n \geq 0$, is a submartingale. Let $a < b$ and $N_0 = -1$, and for $k \geq 1$ let

$$N_l = \begin{cases} \inf\{m > N_{2k-2} : X_m \leq a\}, & l = 2k-1 \\ \inf\{m > N_{2k-1} : X_m \geq b\}, & l = 2k \end{cases}$$

The N_j are stopping times and $\{N_{2k-1} < m \leq N_{2k}\} = \{N_{2k-1} \leq m-1\} \cap \{N_{2k} \leq m-1\}^c \in \mathcal{F}_{m-1}$, so

$$H_m = \begin{cases} 1 & \text{if } N_{2k-1} < m \leq N_{2k} \text{ for some } k \\ 0 & \text{otehrwise} \end{cases}$$

defines a predictable sequence.

Note that $X_{N_{2k-1}} \leq a$ and $X_{N_{2k}} \geq b$. We can regard H_m as a gambling system taking advantage of these upcrossings. In stock market terms, we buy when $X_m \leq a$ and sell when $X_m \geq b$, so every time an upcrossing is completed, we make a profit of $\geq (b-a)$.

Finally, let $U_n = \sup\{k : N_{2k} \leq n\}$ be the number of upcrossings completed by time n.

Theorem 5.12 (upcrossing inequality)**.** *If X_m, $m \geq 0$, is a submartingale, then*

$$(b-a)EU_n \leq E(X_n - a)^+ - E(X_0 - a)^+$$

Proof. Let we introduce $Y_n = a + (X_n - a)^+$ to fix the final incomplete upcrossing, then Y_n is a submartingale that upcrosses $[a, b]$ the same number of times that X_m does. Each upcross results in a profit $\geq (b-a)$ and a final incomplete upcrossing of Y_n (instead of X_n) results in a nonnegative profit, therefore we have $(b-a)U_n \leq (H \cdot Y)_n$.

Let $K_m = 1 - H_m$, then $Y_n - Y_0 = (H \cdot Y)_n + (K \cdot Y)_n$. $(K \cdot Y)_n$ is a submartingale as well, so $E(K \cdot Y)_n \geq E(K \cdot Y)_0 = 0$. Therefore $E(H \cdot Y)_n \leq E(Y_n - Y_0)$. □

Theorem 5.13 (martingale convergence theorem). *If X_n is a submartingale with $\sup EX_n^+ < \infty$, then as $n \to \infty$, X_n converges a.s. to a limit X with $E|X| < \infty$.*

Proof. Since $(X-a)^+ \leq X^+ + |a|$, upcrossing inequality implies that

$$EU_n \leq \frac{EX_n^+ + |a|}{b-a}$$

As $n \uparrow \infty$, $U_n \uparrow U$, where U is the number of upcrossings of $[a, b]$ by the whole sequence, so if $\sup EX_n^+ < \infty$, then $EU < \infty$ and hence $U \leq \infty$ a.s..

Since the last conclusion holds for all rational a and b,

$$P(\bigcup_{a,b\in\mathbb{Q}} \{\liminf X_n < a < b < \limsup X_n\}) = 0$$

and hence $\lim X_n$ exists a.s..

Fatou's lemma guarantees $EX^+ \leq \liminf EX_n^+ < \infty$. For EX^-, we observe that $EX_n^- = EX_n^+ - EX_n \leq EX_n^+ - EX_0$ since X_n is a submartingale, so another application of Fatou's lemma shows $EX^- \leq \liminf EX_n^- \leq \sup EX_n^+ - EX_0$ and completes the proof. □

5.5 Convergence in L^p

Lemma 5.2 (bounded optional stopping). *If X_n is a submartingale and N is a stopping time with $P(N \leq k) = 1$, then*

$$EX_0 \leq EX_N \leq EX_k$$

Proof. $X_{N\wedge n}$ is a submartingale, so $EX_0 = EX_{N\wedge n} \leq EX_{N\wedge k} = EX_N$.

To prove the other inequality, let $K_n = 1_{\{N<n\}} = 1_{\{N\leq n-1\}}$. K_n is predictable, so $(K \cdot X)_n = X_n - X_{n\wedge N}$ is a submartingale, and it follows that $EX_k - EX_N = E(K \cdot X)_k \geq E(K \cdot X)_0 = 0$. □

Lemma 5.3. *If X_n is a submartingale and $M \leq N$ are stopping times with $P(N \leq k) = 1$, then $EX_M \leq EX_N$.*

Proof. Let $K_n = 1_{\{M<n\leq N\}}$ and modify the above proof. □

Lemma 5.4 (bounded Doob's stopping). *If X_n is a submartingale and $M \leq N$ are stopping times with $P(N \leq k) = 1$, then $X_M \leq E(X_N|\mathcal{F}_M)$.*

Proof. Let $A \in \mathcal{F}_M$. Define a random time $L = M1_A + N1_{A^c}$. Actually, this is a stopping time. So $EX_M \leq EX_L \leq EX_N$ by the above lemma. Thus $E(X_M 1_A) \leq E(X_N 1_A)$ and the result follows. □

Corollary 5.13.1. *An adapted and integrable process X_t is a martingale if and only if*

$$E(X_M) = E(X_N)$$

for every such pair of stopping times.

Proof. Let $A \in \mathcal{F}_{n-1}$ and $L = (n-1)1_A + n1_{A^c}$. By $EX_L = EX_n$, we have $EX_n 1_A = EX_{n-1}1_A$, so $E(X_n|\mathcal{F}_{n-1}) = X_{n-1}$. □

Theorem 5.14 (Doob's inequality). *Let X_m be a submartingale, $\lambda > 0$, and $A = \{\max_{0\leq m\leq n} X_m^+ \geq \lambda\}$, then*

$$\lambda P(A) \leq EX_n 1_A \leq EX_n^+$$

Proof. Let $N = \inf\{m : X_m \geq \lambda\} \wedge n$, then $X_N \geq \lambda$ on A. Therefore $\lambda P(A) \leq EX_N 1_A \leq EX_n 1_A$, where the second inequality follows from the lemma above.

The other inequality is obvious. □

Example 5.2 (Kolmogorov's maximal inequality)**.** *If we let $S_n = \xi_1 + \cdots + \xi_n$, where the ξ_m is independent and have $E\xi_m = 0$, $\sigma_m^2 = E\xi_m^2 < \infty$. S_n is a martingale, so S_n^2 is a submartingale. If we let $\lambda = x^2$ and apply Doob's inequality, we get Kolmogorov's maximal inequality:*

$$P(\max_{1 \le m \le n} |S_m| \ge x) \le x^{-2} var(S_n)$$

Theorem 5.15 (L^p maximum inequality)**.** *If X_n is a submartingale, and $\bar{X}_n = \max_{0 \le m \le n} X_m^+$, then for $1 < p < \infty$,*

$$E(\bar{X}_n^p) \le (\frac{p}{p-1})^p E(X_n^+)^p$$

Proof. The ingredients are Doob's inequality and Hölder's inequality. To avoid dividing infinity, we will work with $\bar{X}_n \wedge M$ rather than $\bar{X}_n$. This does not change the application of Doob's inequality.

$$\begin{aligned} E((\bar{X}_n \wedge M)^p) &= \int_0^\infty p\lambda^{p-1} P(\bar{X}_n \wedge M \ge \lambda) \mathrm{d}\lambda \\ &\le \int_0^\infty p\lambda^{p-1} (\lambda^{-1} \int X_n^+ 1_{\{\bar{X}_n \wedge M \ge \lambda\}} \mathrm{d}P) \mathrm{d}\lambda \\ &= \int X_n^+ \int_0^{\bar{X}_n \wedge M} p\lambda^{p-2} \mathrm{d}\lambda \mathrm{d}P \\ &= \frac{p}{p-1} \int X_n^+ (\bar{X}_n \wedge M)^{p-1} \mathrm{d}P \end{aligned}$$

If we let $q = \frac{p}{p-1}$ be the conjugate to p and apply Hölder's inequality, we see that

$$\le (\frac{p}{p-1})(E\left|X_n^+\right|^p)^{1/p} (E\left|\bar{X}_n \wedge M\right|^p)^{1/q}$$

If we divide both sides of the last inequality by $(E\left|\bar{X}_n \wedge M\right|^p)^{1/q}$, which is finite thanks to $\wedge M$, then take the pth power of each side, and letting $M \to \infty$ and using the monotone convergence theorem gives the desired result. ☐

Theorem 5.16 (L^p convergence theorem)**.** *If X_n is a martingale with $\sup E|X_n|^p < \infty$ where $p > 1$, then $X_n \to X$ a.s. and in L^p.*

Proof. $(EX_n^+)^p \le (E|X_n|)^p \le E|X_n|^p$, so it follows from the martingale convergence theorem that $X_n \to X$ a.s..

Applying L^p maximum inequality to $|X_n|$ implies

$$E(\sup_{0 \le m \le n} |X_n|)^p \le (\frac{p}{p-1})^p E(|X_n|)^p$$

Letting $n \to \infty$ and using the monotone convergence theorem implies $\sup|X_n| \in L^p$. Since $|X_n - X|^p \le (2\sup|X_n|)^p$, it follows from the dominated convergence theorem that $E|X_n - X|^p \to 0$. ☐

5.6 Square Integrable Martingales

In this section, we will suppose

$$X_n \text{ is a martingale with } X_0 = 0 \text{ and } EX_n^2 < \infty \text{ for all } n$$

Thus, X_n^2 is a submartingale. It follows from Doob's decomposition that we can write $X_n^2 = M_n + A_n$, where M_n is a martingale, and

$$A_n = \sum_{m=1}^n E(X_n^2 - X_{n-1}^2 | \mathcal{F}_{n-1}) = \sum_{m=1}^n E((X_n - X_{n-1})^2 | \mathcal{F}_{n-1})$$

A_n is called the increasing process associated with X_n.

Theorem 5.17. *$\lim_{n\to\infty} X_n$ exists and is finite a.s. on $\{A_\infty < \infty\}$.*

Lemma 5.5. $E(\sup_m |X_m|^2) \le 4EA_\infty$

Proof. By L^2 maximum inequality, $E(sup_{0\le m\le n}|X_m|^2) \le 4EX_n^2 = 4EA_n + 4EM_n = 4EA_n + 4EM_0 = 4EA_n + 4EX_0^2 = 4EA_n$. Using the monotone convergence theorem now gives the desired result. □

Proof. Let $a > 0$. Since $A_{n+1} \in \mathcal{F}_n$, $N = \inf\left\{n : A_{n+1} > a^2\right\}$ is a stopping time. Applying the lemma to $X_{N\wedge n}$ and noticing $A_{N\wedge n} \ge a^2$ gives $E(\sup_n |X_{N\wedge n}|^2) \le 4a^2$, so the L^2 convergence theorem implies that $\lim X_{N\wedge n}$ exists and is finite a.s.. Since a is arbitary, the desired result follows. □

Theorem 5.18. *Let $f \ge 1$ be increasing with $\int_0^\infty f(t)^{-2}\mathrm{d}t < \infty$. Then $\frac{X_n}{f(A_n)} \to 0$ a.s. on $\{A_\infty = \infty\}$.*

5.7 Convergence in L^1

Now we seek the necessary and sufficient conditions for a martingale to converge in L^1. This leads to the definition of uniformly integrability.

Definition 5.7 (uniformly integrablity)**.** A collection of random variables $X_i, i \in I$ is said to be uniformly integrable if

$$\lim_{M\to\infty}(\sup_{i\in I} E(|X_i|\,;|X_i| > M)) = 0$$

Example 5.3. *A collection of random variables that are dominated by an integrable random variable is uniformly integrable.*

Example 5.4. *A collection of bounded random variables is uniformly integrable.*

Below we give an interesting example of a uniformly integrable family.

Theorem 5.19. *Given a probability space $(\Omega, \mathcal{F}_o.P)$ and an $X \in L^1$, then $\{E(X|\mathcal{F}) : \mathcal{F}$ is a σ-field $\subset \mathcal{F}_o\}$ is uniformly integrable.*

A common way to check uniform integrability is to use:

Lemma 5.6. *Let $\varphi \ge 0$ be any function with $\frac{\varphi(x)}{x} \to \infty$ as $x \to \infty$. If $E\varphi(|X_i|) \le C$ for all $i \in I$, then $X_i, i \in I$ is uniformly integrable.*

Proof. Write $E(|X_i|\,;|X_i| > M) = E(\frac{|X_i|}{\varphi(|X_i|)}\varphi(|X_i|); |X_i| > M)$ and notice that $\frac{|X_i|}{\varphi(|X_i|)} \to 0$. □

Lemma 5.7. *If integrable random variables $X_n \to X$ in L^1, then $E(X_n; A) \to E(X; A)$.*

Proof. The difference is smaller than $E|X_n - X|$. □

Lemma 5.8. *If a martingale $X_n \to X$ in L^1, then $X_n = E(X|\mathcal{F}_n)$.*

Proof. $E(X_m|\mathcal{F}_n)$ for $m > n$, so if $A \in \mathcal{F}_n$, $E(X_m; A) = E(X_n; A)$. By the lemma above, we have $E(X_n; A) = E(X; A)$ for all $A \in \mathcal{F}_n$. By the definition of condition expectation, $X_n = E(X|\mathcal{F}_n)$. □

Theorem 5.20. *For a martingale, TFAE:*
(i) It is uniformly integrable.
(ii) It converges a.s. and in L^1.
(iii) It converges in L^1.
(iv) There is an integrable random variable X s.t. $X_n = E(X|\mathcal{F}_n)$.

5.8 Backwards Martingales

Definition 5.8 (backwards martingale)**.** A backwards martingale is a martingale indexed by the negative integers, i.e., $X_n, n \le 0$, adapted to an increasing sequence of σ-fields $\mathcal{F}_n$ with $E(X_{n+1}|\mathcal{F}_n) = X_n$ for $n \le -1$.

Theorem 5.21. *$X_{-\infty} = \lim_{n\to-\infty} X_n$ exists a.s. and in L^1.*

Proof. Let U_n be the number of upcrossings of $[a, b]$ by $X_{-n}, \cdots, X_0$. The upcrossing inequality implies $(b - a)EU_n \le E(X_0 - a)^+$. Letting $n \to \infty$ and using the monotone convergence theorem, we have $EU_\infty < \infty$, so $X_{-\infty} = \lim_{n\to-\infty} X_n$ exists a.s..
The martingale property implies $X_n = E(X_0|\mathcal{F}_n)$, so X_n is uniformly integrable and the convergence occurs in L_1.

□

Now we identify the limit.

Theorem 5.22. *If $X_{-\infty} = \lim_{n\to-\infty} X_n$ and $\mathcal{F}_{-\infty} = \bigcap_n \mathcal{F}_n$, then $X_{-\infty} = E(X_0|\mathcal{F}_{-\infty})$.*

Proof. Clearly $X_{-\infty} \in \mathcal{F}_{-\infty}$. $X_n = E(X_0|\mathcal{F}_n)$, so if $A \in \mathcal{F}_{-\infty} \subset \mathcal{F}_n$, then $\int_A X_n \mathrm{d}P = \int_A X_0 \mathrm{d}P$, so $\int_A X_{-\infty}\mathrm{d}P = \int_A X_0 \mathrm{d}P$ for all $A \in \mathcal{F}_{-\infty}$, proving the desired conclusion. □

Theorem 5.23. *If $\mathcal{F}_n \downarrow \mathcal{F}_{-\infty}$ as $n \downarrow -\infty$, then $E(Y|\mathcal{F}_n) \to E(Y|\mathcal{F}_{-\infty})$ a.s. and in L^1.*

Proof. $X_n = E(Y|\mathcal{F}_n)$ is a backwards martingale, so $X_n \to X_{-\infty}$ a.s. and in L^1, where $X_{-\infty} = E(X_0|\mathcal{F}_{-\infty}) = E(Y|\mathcal{F}_{-\infty})$. □

5.9 Optional Stopping Theorems

Recall that in Lemma 5.2, we have already established optional stopping theorem for bounded stopping times, so in this section we mainly focus on unbounded stopping time.

Theorem 5.24. *If X_n is a uniformly integrable submartingale, them for any stopping time N, $X_{n\wedge N}$ is uniformly integrable.*

Proof. As X_n^+ is a submartingale, so by bounded optional stopping $EX^+_{N\wedge n} \le EX_n^+$. Since X_n^+ is uniformly integrable, $\sup_n X^+_{N\wedge n} \le \sup_n EX_n^+ < \infty$. So by martingale convergence, $X_{N\wedge n} \to X_N$ a.s. and $E|X_N| < \infty$. Now

$$E(|X_{N\wedge n}|; |X_{N\wedge n}| > K) = E(|X_{N\wedge n}|; |X_{N\wedge n}| > K, N \le n) + E(|X_{N\wedge n}|; |X_{N\wedge n}| > K, N > n)$$

Since $E|X_N| < \infty$ and X_n is uniformly integrable, if K is large, then each term is controlled.

□

Remark. *From the last computation above, we actually get that if $E|X_N| < \infty$ and $X_n 1_{\{N>n\}}$ is uniformly integrable, then $X_{N\wedge n}$ is uniformly integrable. And this is the requirement of the optional sampling theorem we usually see.*

Theorem 5.25. *If X_n is a uniformly integrable submartingale, then for any stopping time $N \le \infty$, we have $EX_0 \le EX_N \le EX_\infty$.*

Proof. By bounded optional stopping, $EX_0 \le EX_{N\wedge n} \le EX_n$. Let $n \to \infty$ and use the $L1$ convergence of uniformly integrable submartingale. □

Theorem 5.26. *If X_n is a uniformly integrable submartingale and $L \le M$ are stopping times and $Y_{M\wedge n}$ is uniformly integrable submartingale, then we have $Y_L \le E(Y_M|\mathcal{F}_L)$.*

For a nonegative supermartingale, we do not require uniform integrability.

Theorem 5.27. *If X_n is a nonnegative supermartingale and $N \le \infty$ is a stopping time, then $EX_0 \ge EX_N$*

Proof. □

Theorem 5.28. *Suppose X_n is a submartingale and $E(|X_{n+1} - X_n||\mathcal{F}_n) \le B$ a.s.. If N is a stopping time with $EN < \infty$, then $X_{N\wedge n}$ is uniformly integrable and hence $EX_N \ge EX_0$.*

Proof. We begin by observing that

$$|X_{N\wedge n}| \le |X_0| + \sum_{m=0}^{\infty} |X_{m+1} - X_m| 1_{(N>m)}$$

To prove uniform integrability, it suffices to show that the right-hand side has finite expectation for then $|X_{N\wedge n}|$ is dominated by an integrable r.v..

$$E(|X_{m+1} - X_m| 1_{\{N>m\}}) = E(E(|X_{m+1} - X_m||\mathcal{F}_m) 1_{\{N>m\}}) \le BP(N > m)$$

and hence the expectation of RHS$\le E|X_0| + BEN$. □

Now we come to Doob's stopping theorem.

Theorem 5.29. *If X_n is a uniformly integrable submartingale, then*

5.9.1 Applications

Example 5.5.

Example 5.6. *Let S_n be symmetric random walk with $S_0 = 0$ and let $T_1 = \min\{n : S_n = 1\}$. Find $P(T_1 = 2n - 1)$.*

Proof. First $P(T_1 < \infty) = 1$ Use the exponential martingale $X_n = \frac{\exp \theta S_n}{E \exp \theta S_n}$. $E \exp \theta S_n = \frac{e^\theta + e^{-\theta}}{2}$. □

Example 5.7. *Let S_n be a symmetric random walk starting at 0, and let $T = \inf\{n : S_n \notin (-a, a)\}$, where a is an integer. Compute ET^2.*

Proof. □

Example 5.8. *Consider a favorable game in which the payoffs are -1,1,2 with probability $\frac{1}{3}$ each. Compute the probability we ever go broke when we start with $i > 0$.*

Lemma 5.9. *Let $S_n = \xi_1 + \cdots + \xi_n$ be a random walk. Suppose*

Proof. □

Proof. The original problem is the case where $\theta_0 = \ln(\sqrt{2} - 1)$. So the probability is $(\sqrt{2} - 1)^i$. □

6 The Probabilistic Method

6.1 The Method

6.2 123 Theorem

As a sort of 'inverse' of the probabilistic method, combinatorial techniques can also apply to probabilistic statement.

6.3 The Local Lemma

Definition 6.1 (dependency digraph)**.** Let $A_i (i \in [n])$ be n events. A directed graph $D = (V, E)$ on the set of vertices $V = \{1, \cdots, n\}$ is called a dependency digraph for the events if for each i the event is mutually independent of all the events $\{A_j : (i, j) \notin E\}$.

Theorem 6.1 (the local lemma)**.** *Let $A_i (i \in [n])$ be n events. Suppose $D = (V, E)$ is a dependency digraph for the above events and suppose there are real numbers $x_1, \cdots, x_n$ s.t. $1 \le x_i < 0$ and $P(A_i) \le x_i \prod_{(i,j)\in E}(1 - x_j)$ for all $1 \le i \le n$. Then*

$$P(\bigcap_{i=1}^{n} \overline{A_i}) \ge \prod_{i=1}^{n}(1 - x_i)$$

Remark. *In particular, with positive probability, no event A_i holds.*

Corollary 6.1.1 (the local lemma: symmetric case)**.**

Remark. *The constant e is the best possible constant.*

6.4 Correlation Inequalities

Theorem 6.2 (the four functions theorem)**.**

Theorem 6.3 (FKG inequality)**.** *Let L be a finite distributive lattice, and let $\mu : L \to \mathbb{R}^+$ be a log-supermodular function. Then for any two increasing functions f, g, we have*

$$E(fg) \ge (Ef)(Eg)$$

where the expectation is taken w.r.t.

Remark. *If both f and g are decreasing, the the result still holds. If one is increasing and one is decreasing, then the inequality is reversed.*

Lemma 6.1 (Kleitman's lemma)**.** *Let*

Theorem 6.4.

Definition 6.2 (linear extension)**.**

Theorem 6.5 (XYZ theorem)**.** *Let*

7 Random Matrices

The goal of this section is to extend previous results for random variables to random matrices.