Probability Theory Notes

Carsen Grote

December 2024

# 1 Introduction

This document serves as a place to list important theorems and concepts that I learned in my graduate probability theory class (Math 733) taught by Timo Seppäläinen at UW-Madison in Fall 2024. The main text for the class is Durrett's Probability: Theory and Examples.

# 2 Measure Theory

## 2.1 Sigma Algebras

**Definition 1** (Sigma Algebra)**.** *Given a space $X$, a $\sigma$-algebra on $X$ is a collection $A$ of subsets of $X$ that satisfy*

- $\emptyset \in A$
- $B \in A$*, then* $B^c \in A$
- $\{A_k\}_{k=1}^{\infty} \subset A$*, then* $\cup_{k=1}^{\infty} A_k \in A$

We see that $X$ contains the empty set, is closed under compliment and countable unions.

**Definition 2** (Measure)**.** *A measure is a function on a sigma algebra,* $\mu : \mathcal{A} \to [0, \infty]$*, such that*

- $\mu(\emptyset) = 0$
- $\mu(\cup_k A_k) = \sum_k \mu(A_k)$ *for a disjoint sequence* $\{A_k\} \subset \mathcal{A}$

A pair $(X, \mathcal{A})$ is a measurable space while the triplet $(X, \mathcal{A}, \mu)$ is a measure space. Now we can define a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ as a measure space where $\mathbb{P}(\Omega) = 1$.

**Definition 3** (Measurable Function)**.** *Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces and $f : X \to Y$ a function. Then, $f$ is a measurable function if for every measurable set $B \in \mathcal{B}$,*

$$f^{-1}(B) = \{x \in X : f(x) \in B\} \in \mathcal{A} \tag{1}$$

*is measurable.*

We can now list a theorem concerning measurable functions.

**Theorem 4.** *Given measurable spaces $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$, and function $X$, if $\{\omega : X(\omega) \in A\} \in \mathcal{F}$ for all $A \in \mathcal{A}$ and $\mathcal{A}$ generates $\mathcal{S}$, then $X$ is measurable.*

Writing $\{X \in B\}$ as shorthand for $\{\omega : X(\omega) \in B\}$, it follows from above that the set $\{\{X \in B\} : B \in \mathcal{S}\}$ is a $\sigma$-field. Further, it is the smallest $\sigma$-field on $\Omega$ that makes $X$ a measurable map and is called the $\sigma$-field generated by $X$.

**Definition 5** (Push Forward)**.** *Given $(X, \mathcal{A}, \mu)$, $(Y, \mathcal{B})$, and measurable function $f : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B})$ we define the push forward of $\mu$ as the measure $\nu$ on $(Y, \mathcal{B})$*

$$\nu(B) = \mu(f^{-1}(B))$$

**Definition 6** (Absolute Continuity)**.** *Let $\mu, \lambda$ be measures on on $(X, \mathcal{A})$. Then $\mu$ is absolutely continuous with respect to $\lambda$ if $\lambda(A) = 0 \implies \mu(A) = 0$ for all $A \in \mathcal{A}$. This is abbreviated by $\mu \ll \lambda$.*

If $\mu \ll \lambda$, then there exists a measurable function $f : X \to \mathbb{R}_{\geq 0}$ such that

$$\mu(A) = \int_A f d\lambda \quad \forall A \in \mathcal{A}. \tag{2}$$

Here $f$ is the Radon-Nikodym derivative and derivative given by

$$f(x) = \frac{d\mu(x)}{d\lambda} \tag{3}$$

# 3 Random Variables

A random variable is a measurable function,

$$X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}}), \tag{4}$$

where $\mathcal{B}_{\mathbb{R}}$ denotes the Borel subsets of $\mathbb{R}$.

**Definition 7** (Distribution)**.** *The distribution of $X$ is the push forward $\mu = \mathbb{P} \circ (X^{-1})$, where*

$$\mu(B) = \mathbb{P}\{\omega \in \Omega : X(\omega) \in B\}$$

**Definition 8** (CDF)**.** *The CDF of $X$ is the function $F$ on $\mathbb{R}$ defined by*

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

**Definition 9** (Expectation)**.** *The expectation of a random variable $X$ is defined as*

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

For functions of random variables, we can define their expectation as

$$\mathbb{E}[f(X)] := \int_\Omega f(X) d\mathbb{P} = \int_{\mathbb{R}} f(x) d\mu. \tag{5}$$

Here $\mu$ is the distribution of $X$, $\mu = \mathbb{P} \circ X^{-1}$, and $f : \mathbb{R} \to \mathbb{R}$ is a Borel function. We are also using the notation $f(x)d\mu = f(x)\mu(dx)$.

**Lemma 1.** *Given that $X > 0$ and $p > 0$, then*

$$\mathbb{E}[X^p] = \int_0^\infty p X^{p-1} \mathbb{P}(X > s) ds. \tag{6}$$

When $p = 1$ this provides that

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > s) ds. \tag{7}$$

## 3.1 Independence

Generally we know that events $A$ and $B$ are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Now we rigorously extend this idea to sigma algebras and random variables.

**Definition 10** (Independent $\sigma$-algebras). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\mathcal{F}_1, ..., \mathcal{F}_n$ be $\sigma$-algebras contained in $\mathcal{F}$. Then $\mathcal{F}_1, ..., \mathcal{F}_n$ are independent if for all $A_1 \in \mathcal{F}_1, ..., A_n \in \mathcal{F}_n$,*

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \mathbb{P}(A_i).$$

We can apply a similar definition for the independence of random variables

**Definition 11** (Independent Random Variables). *The random variables $X_1, ..., X_n$ are independent if the $\sigma$-algebras $\sigma(X_1), \sigma(X_2), ..., \sigma(X_n)$ are independent. Equivalently, for all measurable sets in the range spaces,*

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \in B_i\}\right) = \prod_{i=1}^{n} \mathbb{P}\{X_i \in B_i\}.$$

From this definition it follows that if $X_1, ..., X_n$ are independent, then so are the functions $f(X_1), ..., f(X_n)$.

**Theorem 12.** *Let $\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_n$ be collection of subsets of $\Omega$. Assume that each $\mathcal{A}_i$ is closed under intersection and $\Omega \in \mathcal{A}_i$. Further assume that $\mathbb{P}(\cap_{i=1}^{n} A_i) = \prod_{i=1}^{n} \mathbb{P}(A_i)$ for all $A_i \in \mathcal{A}_i$. Then the $\sigma$-algebras $\sigma(\mathcal{A}_1), ..., \sigma(\mathcal{A}_n)$ are independent.*

The tool for proving the theorem is Dynkin's $\lambda - \pi$ theorem.

**Definition 13** ($\lambda - \pi$ Systems). *Let $\mathcal{A}$ be a collection of subsets of $\Omega$.*

1. *$\mathcal{A}$ is a $\pi$-system if it is closed under intersections.*
2. *$\mathcal{A}$ is a $\lambda - \pi$ system if it satisfies:*
   (a) *$\Omega \in \mathcal{A}$*
   (b) *$A, B \in \mathcal{A}$ and $A \subset B$, then $B \setminus A \in \mathcal{A}$*
   (c) *If $A_1 \subset A_2 \subset \cdots \subset A_n \subset \cdots$ and each each $A_i \in \mathcal{A}$ then $\cup_i^{\infty} A_i \in \mathcal{A}$*

With the definitions of $\pi$ and $\lambda - \pi$ systems we can now state the $\lambda - \pi$ theorem.

**Theorem 14** (Dynkin's $\lambda - \pi$ theorem). *If $\mathcal{P}$ is a $\pi$-system, $\mathcal{L}$ is a $\lambda$-system, and $\mathcal{P} \subset \mathcal{L}$, then $\sigma(\mathcal{P}) \subset \mathcal{L}$.*

Lastly, a corollary to the above theorems on independence of random variables is that $X_1, ..., X_n$ are independent if and only if $\mathbb{P}(\cap_{i=1}^{n} \{X_i \leq s_i\}) = \prod_{i=1}^{n} \mathbb{P}\{X_i \leq s_i\}$ for all $s_i \in (-\infty, \infty]$.

# 4 Product Measures

**Definition 15** (Product Measure Space). *Suppose $(X_1, \mathcal{A}_1, \mu_1), ..., (X_n, \mathcal{A}_n, \mu_n)$ are $\sigma$-finite measure spaces. A product measure space $(X, \mathcal{A}, \mu)$ is then defined as*

$$X = \prod_{i=1}^{n} X_i,$$

*where $\mathcal{A}$ is the product $\sigma$-algebra,*

$$\mathcal{A} = \bigotimes_{i=1}^{n} \mathcal{A}_i = \sigma\{A_1 \times \cdots \times A_n : A_i \in \mathcal{A}_i\},$$

*and $\mu$ is the product measure $\mu = \bigotimes_{i=1}^n \mu_i$ which is by definition the unique measure $\mu$ on $\mathcal{A}$ such that*

$$\mu(A_1 \times \cdots \times A_n) = \prod_{i=1}^n \mu_i(A_i) \quad \forall A_i \in \mathcal{A}_i.$$

For the case in which $(X, \mathcal{A}, \mu)$ is the product of two measure spaces, $(X_1, \mathcal{A}_1, \mu_1)$ and $(X_2, \mathcal{A}_2, \mu_2)$, we have the following theorem of integration.

**Theorem 16** (Fubini's Theorem). *If $f \geq 0$ or $\int |f| < \infty$, then*

$$\int_{X_1} \int_{X_2} f(x,y)\mu_2(dy)\mu_1(dx) = \int_{X_1 \times X_2} f d\mu$$

**Theorem 17.** *Let $X_1, ..., X_n$ be random variables on the same probability space with $\mu_i(B) = \mathbb{P}(X_i \in B)$, $B \in \mathcal{B}_\mathbb{R}$. Take $X = (X_1, ..., X_n)$ as the random vector with probability measure $\mu$ on $\mathbb{R}^n$. Then, $X_1, ..., X_n$ are independent if and only if $\mu = \bigotimes_{i=1}^n \mu_i$.*

A corollary to the above is that if $X = (X_1, ..., X_n)$, where $f_i$ are the associated PDFs of the $X_i$s and $f$ is the PDF of $X$, then $X_1, ..., X_n$ are independent if and only if $f(x_1, ..., x_n) = \prod_i^n f_i(x_i)$ for all $(x_1, ..., x_n) \in \mathbb{R}^n$.

**Theorem 18.** *Let $X$ and $Y$ be independent RVs with distributions $\mu$ and $\nu$, respectively. Then the random variable $X + Y$ has distribution $\mu + \nu$.*

## 4.1 Construction of Independent Random Variables

Given a finite number of distribution functions, $F_i$, we can construct independent random variables $X_1, X_2, ...X_n$with $\mathbb{P}(X_i \leq x_i) = F_i(x_i)$. We take $\Omega = \mathbb{R}^n$, $\mathcal{F} = \mathcal{R}^n$, $X_i(\omega_1, ..., \omega_n) = \omega_i$, and $\mathbb{P}$ the measure on $\mathcal{R}^n$ such that

$$\mathbb{P}((a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_n, b_n]) = ((F_1(b_1) - F_1(a_1)) \cdots (F_n(b_n) - F_n(a_n)). \tag{8}$$

**Theorem 19** (Kolmogorov's extension theorem). *Given consistent probability measures $\nu_i$ on $(\mathbb{R}^i, \mathcal{B}_{\mathbb{R}^i})$. Let $\Omega = \mathbb{R}^{z \geq 0} = \{\omega = (x_i)_{i=1}^\infty : x_i \in \mathbb{R}\}$ and $\mathcal{F} = \sigma\{A_1 \times \cdots \times A_n, A_1, \ldots A_n \in \mathcal{B}_\mathbb{R}\}$. This is the smallest sigma algebra on $\Omega$ under which each $X_i : \Omega \to \mathbb{R}$ is measurable. Then, there exists a unique probability measure $\mathbb{P}$ on $\Omega$ such that*

$$\mathbb{P}\{\omega \in \Omega : (X_1(\omega), \ldots, X_n(\omega)) \in B\} = \nu_n(B), \tag{9}$$

*for all integers $n > 0$ and $B \in \mathcal{B}_\mathbb{R}$.*

## 4.2 Borel-Cantelli Lemmas

**Definition 20** (Infinitely often). *Let $\{A_n\}$ be a sequence of events in $(\Omega, \mathcal{F}, \mathbb{P})$. The events that appear in $A_n$ infinitely often are given as*

$$\{A_n \ i.o.\} = \{\omega : \omega \in A_n \text{ for infinitely many } n\}, \tag{10}$$

$$= \bigcap_{m \geq 1} \bigcup_{n \geq m} A_n. \tag{11}$$

We note that the definition of the set $\{A_n \ i.o.\}$ is equivalent to the lim sup of the sequence $A_n$, denoted $\limsup_{n \to \infty} A_n$. We have a similar definition for the lim inf of a sequnce of sets $A_n$,

**Definition 21.** *Let $\{A_n\}$ be a sequence of events in $(\Omega, \mathcal{F}, \mathbb{P})$. The events that appear in alll but finitely many sets in $A_n$ are given as*

$$\liminf_{n \to \infty} A_n = \{\omega : \omega \notin A_n \text{ for finitely many } n\}, \tag{12}$$

$$= \bigcup_{m \geq 1} \bigcap_{n \geq m} A_n. \tag{13}$$

Now we can state the first and second Borel-Cantelli Lemmas.

**Theorem 22** (1st Borel-Cantelli Lemma). *Given the sequence of events $\{A_n\}$,*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty \implies \mathbb{P}(A_n \ i.o.) = 0. \tag{14}$$

The lemma simply states that if the sum of the probabilities of all the events converges, then the number of events that occur is guarenteed to be finite.

**Theorem 23** (2nd borel-Cantelli Lemma). *If $\{A_n\}$ are independent, then*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty \implies \mathbb{P}(A_n \ i.o.) = 1. \tag{15}$$

Here we have that if the sum of the probabilities of events diverges, then they will occur infinitely often.

### 4.3 Convergence

**Definition 24** (Almost surely convergence). *Given a sequence of random variables $X_n$, we say that $X_n$ converges to $X$ almost surely if*

$$\mathbb{P}\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\} = 1 \tag{16}$$

**Definition 25** (Convergence in probability). *Given a sequence of random variables $X_n$, we say that $X_n$ converges to $X$ in probability if*

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0 \quad \forall \varepsilon > 0. \tag{17}$$

Almost sure convergence is the stronger statement of the two, and implies convergence in probability.

**Theorem 26.** *$X_n$ converges to $X$ in probability if and only if every sub-sequence of $X_n$ has a further sub-sequence that converges almost surely to $X$.*

## 5 Laws of Large Numbers

**Theorem 27** (Weak law of large numbers). *Let $X_i$ be IID random variables with $EX_i = \mu$ and finite variance. If $S_n = X_1 + \cdots + X_n$, then $S_n/n$ converges to $\mu$ in probability.*

**Theorem 28** (Strong law of large numbers). *Let $X_i$ be IID random variables with $EX_i = \mu$ and finite variance. If $S_n = X_1 + \cdots + X_n$, then $S_n/n$ converges to $\mu$ almost surely as $n \to \infty$.*

We see that the weak law provides convergence in probability while the strong law gives almost surely convergence.

**Theorem 29** (Chebyshev's inequality). *If $X$ is a random variable with finite variance then,*

$$\mathbb{P}(|X - EX| \geq a) \leq \frac{var(X)}{a^2} \tag{18}$$

**Lemma 2.** *To apply the SLLN, it is enough to prove that $T_n/n$ converges almost surely to $EX_1$, where*

$$T_n = \sum_{k=1}^{n} Y_k \tag{19}$$

$$Y_k = X_k \cdot 1_{|X_k| \leq k} \tag{20}$$

## 5.1 Empirical Distributions

**Definition 30.** *We define the empirical distribution function as*

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} 1_{X_k \leq x}. \tag{21}$$

*As $n \to \infty$, by the SLLN $F_n(x)$ converges almost surely to $\mathbb{P}(X \leq x) = F(x)$.*

## 5.2 Tail $\sigma$-Algebras

**Definition 31.** *Let $\{X_k\}$ be RVs on $(\Omega, \mathcal{F}, \mathbb{P})$, $S_n = \sum_{k=1}^{n} X_k$ and define the $\sigma$-algebras $\mathcal{F}_n = \sigma\{X_1, \ldots, X_n\}$ and $\mathcal{F}'_n = \sigma\{X_n, X_{n+1}, \ldots\}$. Then the tail sigma algebra $\mathcal{T}$ is defined as*

$$\mathcal{T} = \bigcap_{n \geq 1}^{\infty} \mathcal{F}'_n \tag{22}$$

Intuitively, $A \in \mathcal{T}$ if and only if changing a finite number of values does not affect the occurence of the event. We note as an example that $\limsup X_n$ is $\mathcal{T}$ measurable.

**Theorem 32** (Kolmogorov's 0-1 law)**.** *Let $\{X_k\}$ be independent, then every tail event $A \in \mathcal{T}$ satisfies*

$$\mathbb{P}(A) \in \{0, 1\} \tag{23}$$

## 5.3 Weak convergence

**Definition 33.** *Let $\{\mu_n\}$ and $\mu$ be Borel probability measures on $S$. Then $\mu_n \to \mu$ weakly if*

$$\int_S f d\mu_n \to \int_S f d\mu \quad \forall f \in C_b(S). \tag{24}$$

We note that converging in distribution is the same as converging weakly.

**Theorem 34** (Portmateau theorem)**.** *Let $\{\mu_n\}$, $\mu$ be probability measures on $(S, \rho)$. Then the following are equivalent.*

1. $\mu_n \implies \mu$
2. $\int f d\mu_n \to \int f d\mu \quad \forall f \in C_b(S)$
3. $\int f d\mu_n \to \int f d\mu \quad \forall$ *bounded lipschitz functions* $f : S \to \mathbb{R}$.
4. *For closed sets $F \subset S$, $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$*
5. *For open sets $G \subset S$, $\liminf_{n\to\infty} \mu_n(G) \geq \mu(G)$*
6. *If $A \in \mathcal{B}_S$ and $\mu(bdd\ A) = 0$, then $\mu_n(A) \to \mu(A)$*
7. *If $f : S \to \mathbb{R}$ is a bounded Borel function and $\mu(D_f) = 0$ where $D_f = \{x \in S : f$ is discontinuous at $x$ $\}$, then $\int f d\mu_n = \int f d\mu$*

Portmateau provides a range of statements that are equal. It also follows that if $\int f d\nu = \int f d\mu$ for all $f \in C_b(S)$ then $\nu = \mu$.

**Definition 35** (Convergence in distribution)**.** *We say that a sequence of RVs converges in distribution to a RV if the corresponding distributions converge in distribution.*

**Lemma 3** (Fatou's lemma)**.** *Let $g > 0$ be continuous. If $X_n \implies X$ then*

$$\liminf_{n\to\infty} \mathbb{E}[g(X_n)] \geq \mathbb{E}[g(X)] \tag{25}$$

An example of convergence in distribution is the Central Limit Theorem.

**Theorem 36** (Central Limit Theorem)**.** *Let $X_i$ be iid with $EX_1 = \mu$ and $\sigma^2 = Var(X_1) < \infty$. Then with the random variable $S_n = X_1 + \cdots + X_n$, we have*

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \implies \mathcal{N}(0,1), \tag{26}$$

*where $N$ is the standard normal distribution with mean zero and variance 1.*

**Theorem 37** (Continuous mapping theorem)**.** *Let $g$ be measurable and $\mathbb{P}(X \in D_g) = 0$. If $X_n \implies X$ in distribution, then $g(X_n) \implies g(X)$.*

From the continuous mapping theorem it follows that $X_n$ converges to $X$ if and only if $F_n(x)$ converges to $F(x)$ for all continuity points of $F$.

**Theorem 38** (Glivenko Cantelli)**.** *Let $X_i$ be iid and define*

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{X_i \leq x}. \tag{27}$$

*Then $F_n(x)$ converges almost surely to $F(x)$.*

**Theorem 39.** *Assume that $X_n \implies X$ in distribution. Then we can find $Y_n$, $Y$ on the same probability space such that $Y_n = X_n$ and $Y = X$ in distribution while $Y_n \to Y$ almost surely.*

# 6 Cumulative distribution functions

**Theorem 40** (Helly's selection theorem)**.** *Let $\{F_n\}$ be a sequence of of CDFs. Then there exists a subsequence $\{F_{n_k}\}$ and a nondecreasing, right continuous function $F$ such that $F_{n_k}(x) \to F(x)$ at each continuity point of $F$.*

**Definition 41.** *A sequence $\mu_n$ of probability measures on $\mathbb{R}$ is tight if for all $\varepsilon > 0$ there exists a $m > 0$ such that*

$$\inf_n \mu_n[-m, m] \geq 1 - \varepsilon. \tag{28}$$

*Equivalently, the CDFs $\{F_n\}$ are tight if for all $\varepsilon > 0$ there exists a $m < \infty$ such that $F_n(m) - F_n(-m) \geq 1 - \varepsilon$ for all $n$.*

**Theorem 42.** *Let $\{F_n\}$ be a sequence of CDFs on $\mathbb{R}$. Then every subsequential limit $F_{n_k}(x) \to F(x)$ at each continuity point of $F$ gives a CDF if and only if $\{F_n\}$ is tight.*

**Corollary 1.** *Suppose $\{\mu_n\}$ is a tight sequence of probability measures $\mathbb{R}$. Then every subsequence $\{\mu_{n_k}\}$ has a further subsequence $\{\mu_{n_{k_j}}\}$ that converges weakly to a probability measure on $\mathbb{R}$.*

Weak convergence of probability measures $\mu_n \to \mu$ follows from showing that

1. $\{\mu_n\}$ is tight, and
2. Every convergent subsequence $\{\mu_{n_k}\}$ has the same limit $\mu$

# 7 Characteristic Function

**Definition 43.** *The characteristic function $\rho(t)$ of a random variable $X$ with distribution $\mu$ is defined as*

$$\rho(t) = \mathbb{E}[e^{itX}] = \int e^{itX} \mu(dx) \tag{29}$$

Properties of the characteristic function include that $\rho(0) = 1$, $|\rho(t)| \leq 1$, $\rho$ is uniformly continuous, and if $X$ and $Y$ are independent then $\rho_{X+Y} = \rho_X(t)\rho_Y(t)$. Given the characteristic function $\rho_X(t)$, there exists a general formula to derive the distribution $\mu$ of $X$ from $\rho_X(t)$. The consequence of this is that a characteristic function uniquely determines the distribution of a random variable.

**Theorem 44** (Continuity Theorem)**.** *Let $\{\mu_n\}_{n\geq 1}$ be a sequence of probability measures on $\mathbb{R}$ where $\mu_n$ has characteristic function $\rho_n(t)$. Then,*

1. *If $\mu_n \implies \mu_\infty$ weakly, then $\rho_n(t) \to \rho_\infty(t)$ as $n \to \infty$ for all $t \in \mathbb{R}$.*
2. *If the limit of $\rho_n(t)$ as $n \to \infty$ exists for all $t \in \mathbb{R}$ and $\rho$ is continuous at $0$, then $\rho$ is the characteristic function of a probability measure $\mu$ and $\mu_n$ converges to $\mu$ weakly.*

# 8 Central Limit Theorems

**Theorem 45** (CLT for iid sequences)**.** *Let $X_1, \ldots$ be iid with $\mathbb{E}X_i = \mu$, $var(X_i) = \sigma^2 \in (0, \infty)$. If $S_n = X_1 + \cdots + X_n$ then*

$$\frac{S_n - n\mu}{\sigma n^{1/2}} \implies \chi, \tag{30}$$

*where $\chi$ has the standard normal distribution and the convergence is in distribution (weakly).*

**Theorem 46** (Lindeberg-Feller theorem)**.** *For each $n$, let $X_{n,m}, 1 \leq m \leq n$ be independent random variables with $E[X_{n,m}] = 0$. Let $S_n = \sum_{m=1}^n X_{n,m}$ and suppose the following hold,*

1. *$\sum_{m=1}^n E[X_{n,m}^2] \to \sigma^2 > 0$*
2. *For all $\varepsilon > 0$, $\lim_{n\to\infty} \sum_{m=1}^n E[X_{n,m}^2 1_{\{|X_{n,m}| \geq \varepsilon\}}] = 0$*

*Then $S_n \implies \sigma\chi$ where $\chi \sim \mathcal{N}(0, 1)$.*

It's important to note here that we are not assuming that $X_{n,m}$ are identically distributed. In words, the theorem says that a sum of a large number of small independent effects has approximately a normal distribution. The proof of Lindeberg-Feller shows that the characteristic function of $S_n$ converges to the characteristic function of $\sigma\mathcal{N}$. It uses the following lemmas.

**Lemma 4.** *For all $n \geq 0, t \in \mathbb{R}$,*

$$\left| e^{it} - \sum_{k=0}^{n} \frac{(it)^k}{k!} \right| \leq \min\left( \frac{2|t|^n}{n!}, \frac{|t|^{n+1}}{(n+1)!} \right) \tag{31}$$

**Lemma 5.** *Let $z_1, \ldots, z_n$ and $w_1, \ldots, w_n$ be complex numbers of modulus $\leq \theta$. Then*

$$\left| \prod_{m=1}^{n} z_m - \prod_{m=1}^{n} w_m \right| \leq \theta^{n-1} \sum_{m=1}^{n} |z_m - w_m| \tag{32}$$

In general we have that if $\{X_k\}$ are iid, $S_n = X_1 + \cdots + X_n$, $a = \mathbb{E}X_i$, and $\sigma^2 = Var(X_i) < \infty$, then the SLLN and CLT provide that

1. SLLN: $S_n/n \to a$ with probability 1.
2. CLT:

$$\mathbb{P}\left(\frac{S_n - na}{\sigma\sqrt{n}} \in A\right) \implies \int_A \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt \tag{33}$$

   for all intervals $A$.

## 8.1 Asymptotics of $S_n$

The asymptotics of $S_n$ fit into four regimes around the expected value:

1. Average behavior
2. Fluctuations $O(n^{1/2})$
3. Moderate deviations $O(n^\gamma)$, where $1/2 < \gamma < 1$
4. Large deviations $O(n)$

The Central Limit Theorem states that in the limit you get a Gaussian on the scale of $n^{1/2}$ about the expected value.

**Theorem 47** (Moderate deviations principle)**.** *As $n \to \infty$,*

$$n^{1-2\gamma} \log \mathbb{P}(S_n > na + xn^\gamma) \to \frac{-x^2}{2\sigma^2}. \tag{34}$$

**Theorem 48** (Large deviations principle)**.** *Let $x > 0$, Assume that there exist a $\delta > 0$, and $\mathbb{E}[e^{\delta|X|}] < \infty$. Then as $n \to \infty$,*

$$\frac{1}{n} \log \mathbb{P}(S_n > na + nx) \to -I(x), \tag{35}$$

*where $I(x)$ is the rate function*

Our last theorem here concerns the speed of convergence in the CLT.

**Theorem 49** (Convergence in CLT)**.** *Let $\{X_k\}$ be iid with mean zero and finite variance. Let $\varphi = \mathbb{E}[|X_k|^3] < \infty$ be the third moment. Let*

$$F_n(x) = \mathbb{P}\left(\frac{S_n}{\sigma\sqrt{n}} \leq x\right), \quad \Phi(x) = \int_{-\infty}^{x} \frac{e^{-t^2/2}}{\sqrt{2n}} dt. \tag{36}$$

*Then,*

$$\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \leq \frac{3\varphi}{\sigma^3\sqrt{n}} \tag{37}$$

# 9 Poisson Convergence

Recall that if $X \sim \text{Poisson}(\alpha)$, then $\mathbb{P}(X = k) = (e^{-\alpha}\alpha^k)/k!$.

**Theorem 50.** *For each $n$, let $\{X_{n,m}\}_{1\leq m\leq n}$ be independent random variables such that*

$$p_{nm} = \mathbb{P}(X_{nm} = 1) = 1 - \mathbb{P}(X_{nm} = 0). \tag{38}$$

*Assume that*

1. *$\sum_{m=1}^{n} p_{nm} \to \lambda \in (0, \infty)$ as $n \to \infty$.*

2. $\max_{1\leq m\leq n} p_{nm} \to 0$ *as* $n\to\infty$.

*Then as* $n\to\infty$ *we have*

$$S_n = \sum_{m=1}^{n} X_{nm} \implies Z \sim Poisson(\lambda). \tag{39}$$

*The convergence here is in distribution.*

**Definition 51** (Homogeneous Poisson Process)**.** *Let* $\lambda \in (0,\infty)$*. A rate* $\lambda$ *homogeneous Poisson process* $\{N_t : t \in \mathbb{R}_{\geq 0}\}$ *is a stochastic process with properties*

1. $N_0 = 0$.
2. *For* $s < t$, $N_t - N_s \sim Poisson(\lambda(t-s))$.
3. *For* $0 = t_0 < \cdots < t_m$, $\{N_{t_i} - N_{t_{i-1}}\}_{1\leq i\leq m}$ *are independent.*

**Theorem 52.** *Let* $\lambda \in (0,\infty)$*. Let* $\{Y_k\}$ *be iid with* $Y_k \sim exp(\lambda)$*. Define* $T_0 = 0$, $T_n = \sum_{k=1}^{n} Y_k$ *and let* $N_t = \max\{n : T_n \leq t\}$*, then* $\{N_t\}_{t\geq 0}$ *is a rate* $\lambda$ *Poisson process.*

# 10 Conditional Expectation

Recall that given events $A$ and $B$, the probability of event $A$ given that event $B$ has occurred is

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}. \tag{40}$$

Similarly we can define the discrete and continuous versions of the conditional expectation of a random variable $X$ as

$$\mathbb{E}[X|A] = \sum_x x\mathbb{P}(X = x|A) \tag{41}$$

$$\mathbb{E}[X|Y = y] = \int_{\mathbb{R}} x f_{X|Y}(x,y)dx, \tag{42}$$

where $Y$ is a random variable and $f_{X|Y}$ is the conditional pdf of $X$ given $Y$, i.e.

$$f_{X|Y} = \frac{f(x,y)}{f_Y(y)}, \tag{43}$$

with $f$ the joint pdf of $X$ and $Y$. We now extend these definitions to their measure theory versions.

## 10.1 Measure Theory Conditional Expectation

**Definition 53.** *Let* $(\Omega,\mathcal{F},\mathbb{P})$ *be a probability space,* $\mathcal{A}\subset\mathcal{F}$ *be a sub-sigma-algebra of* $\mathcal{F}$*, and* $X\in L^1(\mathbb{P})$ *a random variable. The conditional expectation of* $X$ *given* $\mathcal{A}$ *is a random variable* $Y$ *on* $\Omega$ *which satisfies*

1. $Y$ *is* $\mathcal{A}$*-measurable (*$Y\in\mathcal{A}$*)*
2. *for all* $A\in\mathcal{A}$, $\int_A X d\mathbb{P} = \int_A Y d\mathbb{P}$

Without proof we note that the conditional expectation $Y$ exists and is unique. For notation, we normally write $\mathbb{E}[X|\mathcal{A}] = Y$.

**Theorem 54** (Radon-Nikodym theorem)**.** *Le t* $\mu$ *and* $\nu$ *be* $\sigma$*-finite measures on* $(\Omega,\mathcal{F})$*. If* $\nu \ll \mu$*, then there is a function* $f\in\mathcal{F}$ *such that for all* $A\in\mathcal{F}$,

$$\int_A f d\mu = \nu(A). \tag{44}$$

$f$ is usually denoted as the Radon-Nikodym derivative, $d\nu/d\mu$. We note two extreme cases of conditional probability. If $X \in \mathcal{F}$ where $\mathcal{F}$ is a sigma-algebra, then $\mathbb{E}[X|\mathcal{F}] = X$ (i.e. if we know $X$, then our best guess is $X$). This is equivalent to knowing everything we're interested in. On the other hand we can know nothing. If $X$ is independent of $\mathcal{F}$, then $\mathbb{E}[X|\mathcal{F}] = \mathbb{E}[X]$.

## 10.2 Properties

**Property 1.** *Suppose $X$ and $Y$ have joint pdf $f(x,y)$ and let $g(X)$ be such that $\mathbb{E}|g(X)| < \infty$. Then $E(g(X)|Y) = h(Y)$ where*

$$h(y) = \frac{\int g(x)f(x,y)dx}{\int f(x,y)dx} = \int g(x)\mathbb{P}(X = x|Y = y)dx. \tag{45}$$

**Property 2.** *Recall that we have $\mathbb{E}[Z\mathbb{E}[X|\mathcal{A}]] = \mathbb{E}[ZX]$ for all bounded $\mathcal{A}$ measurable $X$ and $Z$. Thus given $\psi(Y)$, we have*

$$\mathbb{E}[X\psi(Y)] = \mathbb{E}[\mathbb{E}[X|Y]\psi(Y)] = \mathbb{E}[g(Y)\psi(Y)] = \int_{\mathbb{R}} g(y)\psi(y)\mu_Y(dy). \tag{46}$$

**Property 3.** *Suppose that $\mathcal{A} \subset \mathcal{D} \subset \mathcal{F}$ are $\sigma$-algebras, then*

$$\mathbb{E}[\mathbb{E}(X|\mathcal{A})|\mathcal{D}] = \mathbb{E}[X|\mathcal{A}] = \mathbb{E}[\mathbb{E}(X|\mathcal{D})|\mathcal{A}], \tag{47}$$

*where the equality is almost surely. Further, we have that*

$$\mathbb{E}[1_A\mathbb{E}(X|\mathcal{D})] = \mathbb{E}[1_A\mathbb{E}(X|\mathcal{A})]. \tag{48}$$

**Property 4.** *If $Z$ is $\mathcal{A}$-measurable, $X$ and $XZ \in L^1$, then*

$$\mathbb{E}[XZ|\mathcal{A}] = Z\mathbb{E}(X|\mathcal{A})a.s. \tag{49}$$

*This shows us that $\mathcal{A}$-measurable random variables behave like constants in $\mathbb{E}[\cdot|\mathcal{A}]$.*

**Property 5.** *Suppose $X$ and $Y$ are independent and $\mathbb{E}|\varphi(X,Y)| < \infty$ for a Borel function $\varphi : \mathbb{R}^2 \to \mathbb{R}$. Let $g(x) = \mathbb{E}[\varphi(x,Y)] = \int_\Omega \varphi(x,Y)\mu_Y(dy)$. Then,*

$$\mathbb{E}[\varphi(X,Y)|X] = g(X(\omega)) = \int_\Omega \varphi(X(\omega),Y(\tilde{\omega}))\mathbb{P}(d\tilde{\omega}). \tag{50}$$

**Property 6** (Jensen's Inequality)**.** *Let $\varphi$ be a convex function, assume $X \in L^1$, and $\varphi(X) \in L^1$. Then,*

$$\varphi(\mathbb{E}(X|\mathcal{A})) \leq \mathbb{E}[\varphi(X)|\mathcal{A}] \;\; a.s. \tag{51}$$

**Theorem 55.** *Let $X \in L^1$, thus $\mathbb{E}(X|\mathcal{A}) \in L^2$. Then for all $Y \in L^2(\Omega,\mathcal{A},\mathbb{P})$ we have*

$$\mathbb{E}[(X-Y)^2] \geq \mathbb{E}[(X - \mathbb{E}(X|\mathcal{A}))^2], \tag{52}$$

*where the equality holds if and only if $Y = \mathbb{E}[X|\mathcal{A}]$ a.s.*

**Definition 56** (Regular conditional distribution)**.** *Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, $X : (\Omega,\mathcal{F}) \to (S,\mathcal{S})$ a measurable map, and $\mathcal{G}$ a $\sigma$-field $\subset \mathcal{F}$. Then $\mu : \Omega \times \mathcal{S} \to [0,1]$ is a regular conditional distribution for $X$ given $\mathcal{G}$ if*

1. *For each $A, \omega \to \mu(\omega,A)$ is a version of $\mathbb{P}(X \in A|\mathcal{G})$*
2. *For a.e. $\omega, A \to \mu(\omega,A)$ is a probability measure on $(S,\mathcal{S})$*

When $S = \Omega$ and $X$ is the identity map then $\mu$ is called a regular conditional probability.

**Theorem 57.** *For any $\psi : S \to \mathbb{R}$ bounded such that $\mathbb{E}[\psi(X)]$ exists and is finite, then we have*

$$\mathbb{E}[\psi(X)|\mathcal{A}](\omega) = \int_S \psi(x)\mu(x, dw) \;\; a.s. \tag{53}$$

**Definition 58** (Stochastic Kernel)**.** *If $(S, \mathcal{A})$ and $(T, \mathcal{B})$ are measurable spaces then a stochastic kernel $\pi$ from $(S, \mathcal{A})$ to $(T, \mathcal{B})$ is a function $\pi : S \times \mathcal{B} \to [0, 1]$ such that*

1. *$x \to \pi(x, \mathcal{B})$ is a measurable function on $(S, \mathcal{A})$*
2. *$B \to \pi(x, \mathcal{B})$ is a probability measure on $(T, \mathcal{B})$ for all $x \in S$.*

# 11 Martingales

**Definition 59** (Filtration)**.** *In a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a filtration is a sequence of $\sigma$-algebras $\{\mathcal{F}_n\}_{n\geq 0}$ that satisfy $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots$ and each $\mathcal{F}_n \subset \mathcal{F}$.*

**Definition 60.** *A stochastic process $\{X_n\}_{n\geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ is adapted to the filtration $\{\mathcal{F}_n\}_{n\geq 0}$ if $X_n$ is $\mathcal{F}_n$ measurable for each $n$.*

**Definition 61** (Martingale)**.** *If $X_n$ is a sequence with*

1. *$E|X_n| < \infty$,*
2. *$X_n$ is adapted to $\mathcal{F}_n$,*
3. *$E(X_{n+1}|\mathcal{F}_n) = X_n$ for all $n$,*

*then $X_n$ is said to be a martingale. If in the last definition $=$ is replaced with $\leq$ or $\geq$, then $X$ is said to be a supermartingale or submartingale, respectively.*

**Theorem 62** (Martingale Convergence Theorem)**.** *Let $\{X_n\}_{n\geq 0}$ be a submartingale and $\sup E(X_n^+) < \infty$. Then there exists a random variable $X$ such that $X_n \to X$ almost surely and $E|X| < \infty$.*

This gives us that if $(X_n)$ is a martingale related to $\mathcal{F}_n$ then it is also a martingale related to $\mathcal{F}_n = \sigma\{X_0, \ldots, X_n\}$.

**Theorem 63.** *Let $X_n$ be adapted to $\mathcal{F}_n$ and $\varphi$ a function such that $\varphi(X) \in L^1$ $\forall n$. Then,*

1. *If $X_n$ is a martingale and $\varphi$ is convex then $\varphi(X_n)$ is a sub-martingale*
2. *If $X_n$ is a submartingale and $\varphi$ is convex, nondecreasing then $\varphi(X_n)$ is a submartingale*

**Theorem 64.** *Suppose $(X_n)_{n\geq 0}$ is a supermartingale with respect to the filtration $(\mathcal{F}_n)_{n\geq 0}$, and suppose $(H_n)_{n\geq 1}$ is a predictable process. Assume that each $H_n \geq 0$ and bounded. Define $H \cdot X$ by $(H \cdot X)_0 = 0$ and*

$$(H \cdot X)_n = \sum_{m=1}^{n} H_m(X_m - X_{m-1}), \tag{54}$$

*for $n \geq 1$. Then $H \cdot X$ is a supermartingale.*

It's also worth noting from the above that if $X_n$ is a submartingale then $H \cdot X$ is a submartingale and similarly if $X_n$ is a martingale.

**Definition 65** (Stopping Time)**.** *A $\mathbb{Z} \cup \{\infty\}$-valued random variable $N$ is a stopping time if for all integers $n \geq 0$ we have $\{\omega : N(\omega) = n\} \in \mathcal{F}_n$ where $\mathcal{F}$ is a filtration.*

This definition means that "$N$ cannot peek into the future."

**Theorem 66.** *If $N$ is a stopping time and $X_n$ is a supermartingale then $X_{N\wedge n}$ is also a supermartingale where*

$$X_{N\wedge n}(\omega) = \begin{cases} X_{N(\omega)},\ n > N(\omega) \\ X_n(\omega),\ n \leq N(\omega) \end{cases} \tag{55}$$

**Definition 67** (Upcrossings)**.** *Let $N_0 = 0$, $k > 1$, and $X_n$ be a submartingale. Then define*

$$\begin{aligned} N_{2k-1} &= \inf\{m > N_{2k-2} : x_m \leq a\} \\ N_{2k} &= \inf\{m > N_{2k-1} : x_m \geq b\} \end{aligned}$$

*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*

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

*defines a predictable sequence. We have that $X(N_{2k-1}) \leq a$ and $X(N_{2k}) \geq b$, so between times $N_{2k-1}$ and $N_{2k}$, $X_m$ crosses from below $a$ to above $b$. Let $U_n = \sup\{k : N_{2k} \leq n\}$ be the number of "upcrossingss" completed by time $n$.*

**Theorem 68** (Upcrossing inequality)**.** *If $X_m$, $m \geq 0$ is a submartingale then*

$$(b-a)\mathbb{E}[U_n] \leq \mathbb{E}[(X_n - a)^+] - \mathbb{E}[(X_0 - a)^+] \tag{56}$$

**Theorem 69** (Martingale convergence theorem)**.** *Let $X_n$ be a submartingale, with $\sup \mathbb{E}[X_n^+] < \infty$. Then there exists a random variable $X$ such that $X_n \to X$ almost surely and $\mathbb{E}[X] < \infty$.*

**Lemma 6.** *Let $X_n \geq 0$ be a supermartingale. Then there exists an almost surely limit $X = \lim_n X_n$ with $0 \leq \mathbb{E}X \leq \mathbb{E}X_0$.*

**Theorem 70.** *If $X_n$ is a submartingale, $k \in \mathbb{Z}_{>0}$, $N$ a stopping time such that $\mathbb{P}(0 \leq N \leq k) = 1$ then*

$$\mathbb{E}[X_0] \leq E[X_N] \leq \mathbb{E}[X_k]. \tag{57}$$

**Theorem 71** (Doob's Inequality)**.** *Let $X_n$ be a submartingale, $\lambda > 0$, and define*

$$A = \max_{0\leq m\leq n} X_m > \lambda. \tag{58}$$

*Then we have that*

$$\mathbb{P}(A) \leq \frac{1}{\lambda}\mathbb{E}[X_n 1_A] \leq \frac{1}{\lambda}\mathbb{E}[X_n^+] \tag{59}$$

**Theorem 72** ($L^p$ Maximum inequality)**.** *Let $X_n$ be a submartingale with $1 < p < \infty$. Define*

$$\bar{X}_n = \max_{0\leq m\leq n} X_m^+. \tag{60}$$

*Then we have that*

$$\mathbb{E}[\bar{X}_n^p] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[(X_n^+)^p]. \tag{61}$$

**Theorem 73** ($L^p$ convergence theorem)**.** *Let $X_n$ be a martingale that satisfies $L^p$ boundedness,*

$$\sup_n \mathbb{E}[|X_n|^p] < \infty. \tag{62}$$

*Then $X_n \to X$ almost surely in $L^p$.*

**Theorem 74** (Kolmogorov's Inequality)**.** *Let* $\{X_k\}_k$ *be independent with* $\mathbb{E}[X_k] = 0$ *and* $Var(X_k) < \infty$. *Then,*

$$\mathbb{P}(\max_{1 \le m \le n} |S_n| \ge \lambda) \le \frac{\mathbb{E}[S_n^2]}{\lambda^2}. \tag{63}$$

**Theorem 75.** *Let* $\{X_k\}_k$ *be independent with* $\mathbb{E}[X_k] = 0$ *and* $Var(X_k) < \infty$. *Then,* $S_n$ *is a martingale and*

$$\sum_{k=1}^{\infty} Var(X_k) < \infty \implies \sum_{k=1}^{\infty} X_k \text{ converges to a finite limit.} \tag{64}$$

*We have that if* $Var(S_n)$ *converges to a finite limit then* $S_n$ *converges to a finite limit as* $n \to \infty$.

## 11.1 Uniform Integrability

**Theorem 76** (Single random variable uniform integrability)**.** *The following are equivalent,*

1. $\mathbb{E}|X| < \infty$ *($L^1$ boundedness)*
2. $\lim_{m \to \infty} \mathbb{E}[|X| 1_{|X| \ge m}] = 0$

**Definition 77** (Uniform integrability of a collection of random variables)**.** *A collection of random variables* $X_i, i \in I$ *is said to be uniformly integrable if*

$$\lim_{M \to \infty} \left( \sup_{i \in I} \mathbb{E}[|Xi|; |X_i| > M] \right) = 0. \tag{65}$$

**Theorem 78.** *Let* $\varphi \ge 0$ *be any function with* $\varphi(x)/x \to \infty$ *as* $x \to \infty$. *If* $\mathbb{E}[\varphi(|X_i|)] \le C$ *for all* $i \in I$, *then* $\{X_i : i \in I\}$ *is uniformly integrable.*

**Theorem 79.** *Let* $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. *Then the family* $\{\mathbb{E}[X|\mathcal{A}] : \mathcal{A} \subset \mathcal{F}, \mathcal{A} \text{ is a sigma-algebra}\}$ *is uniformly integrable.*

**Theorem 80.** *Let* $\{X_n\}$ *be a sequence such that* $X_n \in L^1$ *and* $X_n \to X$ *in probability. Then the following are equivalent,*

1. $\{X_n\}_{n \ge 1}$ *are uniformly integrable*
2. $X \in L^1$ *and* $X_n \to X$ *in* $L^1$
3. $\mathbb{E}|X_n| \to \mathbb{E}|X| < \infty$

**Theorem 81.** *Let* $X_n$ *be a submartingale, then the following are equivalent*

1. $\{X_n\}$ *is uniformly integrable*
2. *There exists a random variable* $X$ *such that* $X_n \to X$ *almost surely and in* $L^1$
3. *There exists a random variable* $X$ *such that* $X_n \to X$ *in* $L^1$

**Lemma 7.** *If* $X_n$ *is a martingale and* $X_n \to X$ *in* $L^1$ *then* $X_n = \mathbb{E}[X|\mathcal{F}_n]\ \forall n$.

**Theorem 82.** *Let* $X_n$ *be a martingale. Then the following are equivalent*

1. $\{X_n\}$ *is uniformly integrable*
2. $X_n \to X$ *almost surely and in* $L^1$
3. $X_n \to X$ *in* $L^1$
4. *There exists an* $X \in L^1$ *such that* $X_n = \mathbb{E}[X|\mathcal{F}_n]\ \forall n$

**Definition 83.** *If* $\{\mathcal{F}_n\}$ *is a filtration then* $\mathcal{F}_n \uparrow \mathcal{F}_\infty$ *means that* $\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n)$.

**Theorem 84.** *Let* $\mathcal{F}_n \uparrow \mathcal{F}_\infty$ *and* $X \in L^1$. *Then* $\mathbb{E}[X|\mathcal{F}_n] \to \mathbb{E}[X|\mathcal{F}_\infty]$ *almost surely and in* $L^1$.

**Theorem 85** (Levy's 0-1 law)**.** *Let* $\mathcal{F}_n \uparrow \mathcal{F}_\infty$. *Then* $A \in \mathcal{F}_\infty$ *implies that* $\mathbb{E}[1_A|\mathcal{F}_n] \to 1_A$ *almost surely in* $L^1$.