Preliminaries to Stochastic Analysis
Autumn 2011 version

Xue-Mei Li
The University of Warwick

Typset: October 21, 2011

Chapter 1

Preliminaries

For completion and easy reference we include some basic concepts from Measure Theory. We assume that you are already familiar with most of the text here.

1.1 Sets

Let E be a set. We summarise below operations on subsets of E. If A and B are two sets, $A \cup B$, $A \cap B$ stand for the union and the intersections of the two sets; and $A^c \equiv E - A$ the complement of A in E and $A \Delta B = (A - b) \cup (B - a)$ the symmetric difference of B and A. The two binary operations, the set union and the set intersection, satisfies the commutative laws, associative laws and distributive laws for mixed operations. The distributive law reads:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \qquad A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

Proposition 1.1.1 *DeMorgan's law. For any index set I*

$$(\cap_{\alpha \in I} E_\alpha)^c = \cup_{\alpha \in I} E_\alpha^c, \qquad (\cup_{\alpha \in I} E_\alpha)^c = \cup_{\alpha \in I} E_\alpha^c.$$

There are concepts of limit superior and limit inferior. If A_n is an increasing sequence of sets, $A_1 \subset A_2 \subset A_3 \ldots$, write

$$\lim_{n \to \infty} A_n := \cup_{n=1}^\infty A_n = \{x \in E : x \in A_n \text{for some } n\}.$$

If A_n is a decreasing sequence of sets, $A_1 \supset A_2 \supset A_3 \ldots$, write

$$\lim_{n \to \infty} A_n := \cap_{n=1}^\infty A_n = \{x \in E : x \in A_n \text{ for all } n\}.$$

If $\{A_n\}$ is a sequence of sets define

$$\limsup_{n\to\infty} A_n = \cap_{m=1}^{\infty} \cup_{k=m}^{\infty} A_k.$$

$$\liminf_{n\to\infty} A_n = \cup_{m=1}^{\infty} \cap_{k=m}^{\infty} A_k.$$

The set $\limsup_{n\to\infty} A_n$ contains precisely all x such that there is an infinite subsequence of A_{n_k} with $x \in \cap_{k=1}^{\infty} A_{n_k}$. The set $\liminf_{n\to\infty} A_n$ contains precisely all x such that there is a N such that $x \in \cap_{k=1}^{\infty} A_{N+k}$.

1.2 Measurable Spaces

Let Ω be a set.

Definition 1.2.1 *An algebra $\mathcal{A}$ of a set Ω is a collection of subsets such that*

1. $\emptyset \in \mathcal{A}$,
2. *If $A \in \mathcal{A}$, the complement $A^c \equiv \Omega/A$ belongs to $\mathcal{A}$.*
3. *If $A_1, A_2 \in \mathcal{A}$, then $A_1 \cup A_2 \in \mathcal{A}$*

Note that Ω, as compliment of the empty set belongs to $\mathcal{A}$. Also finite unions and intersections of sets from $\mathcal{A}$ belongs to $\mathcal{A}$. A σ-algebra $\mathcal{F}$ requires that countable unions of sets from $\mathcal{F}$ belongs to $\mathcal{F}$. Intersection of algebras is an algebra.

Definition 1.2.2 *A σ-algebra $\mathcal{F}$ on Ω is a collection of subsets of Ω such that*

(1) $\emptyset \in \mathcal{F}$

(2) $\mathcal{F}$ is closed under complement: if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$.

(3) $\mathcal{F}$ is closed under countable unions: if $A_n \in F$, $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$.

The pair $(\Omega, \mathcal{F})$ is a Measurable Space. Elements of $\mathcal{F}$ are called Measurable Sets. They are also called Events.

It follows that $\mathcal{F}$ is closed under countable intersections. Note also that non-emptyness+(2) is the same as (1)+(2). A collection of sets is an algebra if is it closed under finite unions and taking complements.

Example 1.2.3 • The trivial σ-algebra $\{\emptyset, \Omega\}$ is the smallest σ-algebra on Ω.

- For any set Ω, we may consider the set of all subsets of Ω. This is the largest σ-algebra on Ω.
- For a finite set $\Omega = \{1, 2, \ldots, n\}$ we typically take $\mathcal{F}$ to be the set of all subsets of Ω.
- Take $\Omega = \{1, 2, 3, 4\}$. Define $\mathcal{F} = \{\{1, 2, 3\}, \{4\}, \emptyset, \omega\}$.

The intersection of any number of σ-algebras $\{\mathcal{F}_I, I \in \Lambda\}$ is a σ-algebra trivially. If $A \in \cap_{I\in\Lambda}\mathcal{F}_I$ then $A^c \in \mathcal{F}_I$ for all I. If $A_n \in \cap_{I\in\Lambda}\mathcal{F}_I$ clearly $\cup_{n=1}^{\infty} A_n \in F_I$ for all I.

Definition 1.2.4 *If G is a collection of subsets of Ω, the σ- algebra generated by G is the smallest σ-algebra that contains all elements of G, and is denoted by $\sigma(G)$. It is the intersection of all σ-algebras of Ω that contains G.*

Example 1.2.5 Let $\Omega = \{1, 2, 3, 4\}$ and $G = \{\{1, 2, 4\}, \{4\}\}$, show that the smallest σ-algebra generated by G consists of the following sets: $\emptyset$, Ω, $\{1, 2\}$, $\{3\}$, $\{4\}$, $\{1, 2, 4\}$, $\{3, 4\}$ and $\{1, 2, 3\}$.

Definition 1.2.6 *If (X, d) is a metric space with metric d, the Borel σ-algebra is the σ-algebra generated by open balls $\{x : d(x, x_0) < a\}$, $x_0 \in X, a \in \mathbf{R}$. It is denoted by $\mathcal{B}(X)$, and elements of $\mathcal{B}(X)$ are called Borel sets of X.*

If not otherwise mentioned, this is the σ-algebra we will use. Single points and therefore a countable subset of a metric is always measurable since $\{x_0\} = \cap_n\{x : |x - x_0| < \frac{1}{n}\}$. For a topological space, the Borel σ-algebra is the σ-algebra generated by open sets.

Example 1.2.7
- The Borel σ-algebra of $\mathbf{R}$ is the σ algebra generated by the open intervals (a, b).
- Let $W_0 = \{\sigma : [0, 1] \to \mathbf{R}$ continuous $, \sigma(0) = 0\}$ be the set of all continuous functions from $[0, 1]$ to $\mathbf{R}$. Define the distance function to be:

  $$d(\sigma_1, \sigma_2) = \sup_{t\in[0,1]} |\sigma_1(t) - \sigma_2(t)|.$$

  Since the functions are continuous if Q is countable dense subset of $[0, 1]$,

  $$d(\sigma_1, \sigma_2) = \sup_{t\in Q} |\sigma_1(t) - \sigma_2(t)|.$$

  The collection of sets $\{\sigma : |\sigma(t_i) - \sigma_0(t_i)| \le a, 0 \le t_1 < \cdots < t_n = 1\}$, $a > 0$, $0 \le t_1 < \cdots < t_n = 1$ and partition, σ_0 any curve in Ω gives the topology on Ω, which is a π-system, and hence determine the Borel σ-algebra.

1.3 Measures

We are interested in functions that gives a value to each set in a given σ algebra. They should have set additive property $F(A \cup B) = F(A) + F(B)$ if $A \cap B = \emptyset$. For it to represent measurements of sets, we assume further conditions:

Definition 1.3.1 *A measure on a measurable space* $(\Omega, \mathcal{F})$ *is a* $[0, \infty]$ *valued function on* $\mathcal{F}$*, which assigns to every set from* $\mathcal{F}$ *a value from* $[0, \infty]$ *and has the following property*

1. $P(\emptyset) = 0$
2. $P(\cup_{i=1}^{\infty}(A_i) = \sum_{i=1}^{\infty} P(A_i)$*, for all sequences* $\{A_j\}$ *with* $A_j \in \mathcal{F}$ *and* $A_i \cap A_j = \emptyset$ *whenever* $i \neq j$

The triple $(\Omega, \mathcal{F}, P)$ *is called a measure space.*

If $A \subset B$ then $P(B) = P(A \cup (B \backslash A)) = P(A) + P(B \backslash A) \geq P(A)$. If $\{A_i\}$ are not pairwise disjoin then $P(\cup_{i=1}^{\infty}(A_i) \leq \sum_{i=1}^{\infty} P(A_i)$, when P is a positive finite measure: $0 < P(\Omega) < \infty$. In fact let $C_1 = A_1$, $C_2 = A_2 - A_1$, then $C_1 \cup C_2 = A_1 \cup A_1$. Set $C_3 = A_3 - C_2 = A_3 - (A_1 \cup A_2)$, $C_n = A_n - C_{n-1}$. Then $\cup_{i=1}^{n} C_i = \cup_{i=1}^{n} A_i$ and $\{C_n\}$ are disjoint, $P(C_n) \leq P(A_n)$ and by property (2), $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(C_i) \leq \sum_{i=1}^{\infty} P(A_i)$.

Property (2) is called the σ-additive property, which implies that if A_n is a non-decreasing sequence of measurable sets $A_1 \subset A_2 \subset \dots$,

$$\lim_{n \to \infty} P(A_n) = P(\cup_{n=1}^{\infty} A_n). \tag{1.1}$$

We also write $\lim_{n \to \infty} A_n = \cup A_n$. To see this define $C_1 = A_1, C_n = A_n - C_{n-1} = A_n - A_{n-1}$. Similarly if A_n is a non-increasing sequence of measurable sets $A_1 \supset A_2 \supset \dots$,

$$\lim_{n \to \infty} P(A_n) = P(\cap_{n=1}^{\infty} A_n).$$

Definition 1.3.2 *A measure on* Ω *is a probability (measure) if the total mass is one:* $\mu(\Omega) = 1$. *It is called a* σ**-finite** *measure if* Ω *is the countable union of measurable sets of finite measure. Furthermore if* Ω *is a topological space with the Borel* σ*-field* $\mathcal{B}(\Omega)$*, the measure is called a* **Borel measure***.*

A Borel measure is **locally finite** if for every point ω in Ω there is an open neighbourhood U of ω with $\mu(U)$ finite. It is **tight** (or inner regular) if $\mu(A) = \text{Sup}\{\mu(K) : K \subset A \text{ and } K \text{ is compact}\}$ for all $A \in \mathcal{B}$. Equivalently for any $\epsilon > 0$ there is K compact such that $\mu(K) > 1 - \epsilon$, and it is **Radon** if it is locally finite and tight. If

Ω is a **Polish space**, i.e. separable and admits a complete metric, then every locally finite measure μ is tight and hence Radon. Following from this Borel measures are constructed on $\mathbf{R}^n$ be defining their values on open balls.

Example 1.3.3 Let Ω consisting of a countable number of elements ω_i and let $\mathcal{F}$ be a σ-algebra. Define a function on $m : \Omega \to \mathbf{R}$. Set $m_i = m(\omega_i) \in \mathbf{R}$. Then

$$P(A) = \sum_{i:\omega_i \in A} m_i$$

defines a σ-finite measure. The total mass is $\sum_{i=1}^{\infty} m_i$.

Example 1.3.4 A standard way to construct measure on $\mathbf{R}^n$ is by introducing a density function $p : \mathbf{R}^n \to \mathbf{R}$:

$$\mu(A) = \int_{x \in A} p(x) dx.$$

Definition 1.3.5 *Consider the measure space $(\Omega, \mathcal{F}, P)$. We say that a measure or a σ-algebra is complete if, whenever A is a set in $\mathcal{F}$ with zero measure, any subset of A is in $\mathcal{F}$. If $P(\Omega) = 1$, the triple $(\Omega, \mathcal{F}, P)$ is called a probability space.*

We often work on the completion of $\mathcal{F}$ which is the possibly larger σ-algebra containing $\mathcal{F}$ and all subsets of zero measure sets.

Standard Assumption. We assume that the measure is a probability measure and that the σ-algebra is complete. In this case if $f = g$ almost surely, that is they agree on a set of full measure, then they are both measurable or both not measurable.

Definition 1.3.6 *We say that an event A happens* **almost surely**, *abbreviated as a.s., if $P(A) = 1$.*

Remark 1.3.7 A measure on $\mathcal{F}$ is determined by its value on a generating set of $\mathcal{F}$ that is a π system.

Let μ_1 and μ_2 be the two measures. Set

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

Then C contains a π system by definition. It contains the empty set. If $A_1 \subset A_2$, $A_1, A_2 \in C$, then $\mu_i(A_2) = \mu_i(A_1 \cup (A_2 - A_1)) = \mu_i(A_1) + \mu_i(A_2 - A_1)$. Hence

$\mu_1(A_2 - A_1) = \mu_1(A_2) - \mu_1(A_1) = \mu_2(A_2) - \mu_2(A_1) = \mu_2(A_2 - A_1)$. So $A_2 - A_1 \in C$. Next let $\{A_i\}$ be an increasing sequence in C, then

$$\mu_1(\cup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \cup_{i=1}^{n} \mu_1(A_1) = \lim_{n\to\infty} \mu_2(\cup_{i=1}^{n} A_i) = \mu_2((\cup_{n=1}^{\infty} A_n).$$

This shows that C is a σ-additive class and hence contains $\sigma(F)$.

Example 1.3.8 The sets in the completion of the Borel σ-algebra are said to be Lebesgue measurable. On $\mathbf{R}$ define the measure L of an interval to be the length of the interval. If A is a Borel set,

$$L(A) = \inf\{\sum_j (b_j - a_j) : \cup_{j=1}^{\infty}(a_j, b_j) \supset A\}.$$

This determines the Lebesgue Measure. In general let F be a right continuous increasing function, there is a unique Borel measure on $\mathbf{R}$, called the Lebesgue-Stieltjes measure associated to F, such that $\mu_F((a, b]) = F(b) - F(a)$. Furthermore

$$\mu_F(A) = \inf\{\sum_j (F(b_j) - F(a_j)) : \cup_{j=1}^{\infty}(a_j, b_j) \supset A\}.$$

It is customary to denote $\int h\, dF$ for the intergal $\int h\, d\mu_F$. If F is the difference of two right increasing functions, a signed measure can be defined. Integration with respect to the quadratic variation process of a martingale, c.f. Theorem **??**, Theorem **??** belongs to this category, see Kunita-Watanabe inequality **??**.

If μ is a probability measure on the real line, there is the assocaited distribution function $F(x) = \mu((-\infty, x])$. Then $\mu((a, b]) = F(b) - F(a)$. Then F is increasing and right continuous. Also $F(-\infty) = 0$ and $F(\infty) = 1$, To see right continuity, take $x_n > x$ converging to x,

$$\lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} \mu((-\infty, x_n]) = \mu(\cap_{n=1}^{\infty}(-\infty, x_n]) = F(x)$$

by the σ additive property of μ. I If $X : \Omega \to \mathbf{R}$ is a random variable, its distribution function is that of its probability distribution: $F(x) = P(\omega : X(\omega) \le x)$.

Similarly $F : [a, b] \to \mathbf{R}$ right continuous and increasing defines a measure on $\mathcal{B}([a, b])$. If F is absolutely continuous on $[a, b]$ in which case $F(x) = f(a) + \int_a^x g(t)dt$, with g integrable w.r.t. the Lebesgue measure and f differentiable at almost surely all points with $F' = g$ a.s. with respect to the Lebesgue measure. In this case

$$\int_a^b h(x)\mu_F(dx) = \int_a^b h(x)F'(x)dx.$$

A right continuous increasing function F on $[0,\infty)$ has finite total variation, see Definition 1.12.1. We may assume that $F(0)=0$. The measure μ_F is absolutely continuous with respect to the Lebesque measure if and only if F is absolutely continuous.

Proposition 1.3.9 (The Borel-Canteli Lemma) *Let $\{A_i\}$ be a sequence of measurable sets.*

1. $$\sum_{n=1}^{\infty} P(A_n) < \infty \qquad \textit{implies that} \qquad P(\cap_{N=1}^{\infty} \cup_{m=N}^{\infty} A_m) = 0.$$

2. *If the sets from $\{A_i\}$ are independent, then*

$$\sum_{n=1}^{\infty} P(A_n) = \infty \qquad \textit{implies that} \qquad P(\cap_{N=1}^{\infty} \cup_{m=N}^{\infty} A_m) = 1.$$

1.4 The Monotone Class Theorem

Definition 1.4.1 *Let Ω be a set. A non-empty family of subsets $\mathcal{G}$ of Ω is called a monotone class if limits of monotone sequences in $\mathcal{G}$ belong to $\mathcal{G}$.*

1. *$\cup_{n=1}^{\infty} A_n \in \mathcal{G}$ if A_n is an increasing sequence in $\mathcal{G}$.*
2. *$\cap_{n=1}^{\infty} A_n \in \mathcal{G}$ if A_n is an decreasing sequence in $\mathcal{G}$*

Finite additive property together with monotone class property is equivalent to the σ-additive property. Suppose that $\mathcal{A}$ is both an algebra and monotone class, it is an σ algebra. In fact if $A_i \in \mathcal{A}$, let $B_n = \cup_{i=1}^{n} A_i$ then $\cup B_{i=1}^{n} = \cup A_{i=1}^{n}$ and B_n is increasing. Hence $\cup_{n=1}^{\infty} B_n \in \mathcal{A}$. Since $\cup_{n=1}^{\infty} A_n = \cup_{n=1}^{\infty} B_n$, σ-additive property follows.

A σ-algebra is a monotone class. Intersection of monotone classes containing a common element is a monotone classes.

Theorem 1.4.2 (Monotone Class Theorem) *If $\mathcal{A}$ is a algebra, then the smallest monotone class $m(\mathcal{A})$ of sets containing $\mathcal{A}$ is the σ algebra $\sigma(A)$ generated by $\mathcal{A}$.*

Proof It is sufficient to show that $m(\mathcal{A})$ is an algebra, in which case it is a σ-algebra and hence contains $\sigma(\mathcal{A})$. First we show that it is closed under taking complements. Consider the set

$$C = \{A \in m(\mathcal{A}) \;:\; A^c \in m(\mathcal{A})\}.$$

Then $\mathcal{A} \subset C \subset m(\mathcal{A})$. We show that C is also monotone and hence C must contain the smallest monotone class that contains $\mathcal{A}$. If A_n increases and belongs to C, clearly $\cup A_n$ belong to $m(\mathcal{A})$ and $(\cup A_n)^c = \cap A_n^c \in m(\mathcal{A})$. Similarly if A_n belongs to C and decreases then $\cap A_{n=1}^\infty \in \mathcal{C}$. We proved that C is a monotone class and hence $C = m(\mathcal{A})$.

Now we prove that intersection of two set in $m(\mathcal{A})$ is in $m(\mathcal{A})$. Let $A \in m(\mathcal{A})$. Define

$$C^A = \{B \in m(\mathcal{A}) : A \cap B \in m(\mathcal{A})\}.$$

It is easy to check that C^A is a monotone class. We now prove the assertion in two steps. If $A \in \mathcal{A}$ then $\mathcal{A} \subset C^A \subset m(\mathcal{A})$ and hence $C^A = m(\mathcal{A})$ which means that for $A \in \mathcal{A}$ and $B \in m(\mathcal{A})$, $A \cap B \in m(\mathcal{A})$. Hence for $B \in m(A)$ then C^B contains $\mathcal{A}$: $\mathcal{A} \subset C^B \subset m(\mathcal{A})$. Since C^B is monotone, $C^B = m(\mathcal{A})$. Hence for all $A, B \in m(\mathcal{A})$, $A \cap B \in m(\mathcal{A})$.

We have proved that $m(\mathcal{A})$ is an algebra, and hence a σ-algebra. So $m(\mathcal{A}) \supset \sigma(\mathcal{A})$. On the other hand $\sigma(\mathcal{A})$ is a monotone class trivially and so contains $m(\mathcal{A})$. We proved the assertion that $\sigma(\mathcal{A}) = m(\mathcal{A})$. □

Definition 1.4.3 *A collection of subset $\mathcal{A}$ is a π system if $A, B \in \mathcal{A}$ implies that $A \cap B \in \mathcal{A}$.*

The intersection of π-systems is a π-system. Note that some people may insist that the empty set $\emptyset$ is in π system.

Definition 1.4.4 *A family $\mathcal{B}$ of subsets of a set Ω is a σ-additive class also called a Dynkin system if*

1. $\Omega \in \mathcal{B}$.
2. *If $B_1 \subset B_2$, $B_1, B_2 \in \mathcal{B}$ then $B_1 \backslash B_2 \in \mathcal{B}$.*
3. *If A_n is an increasing sequence in $\mathcal{B}$ then $\cup_{n=1}^\infty A_n \in \mathcal{B}$.*

Note that property 1 and 2 imply that σ-additive class is closed under taking compliment. A σ-additive class is a monotone class. The intersection of σ-additive class is a σ-additive class. Under condition 1 and 2, if $\mathcal{B}$ is a π-system it is closed under disjoint union and the following are equivalent.

a) $\cup_{n=1}^\infty B_n \in \mathcal{B}$ whenever $B_n \in \mathcal{B}$ and $\{B_n\}$ are pairwise disjoint.

b) $B_1 \cup B_2 \in B$ whenever B_1 and B_2 are disjoint and A_n is an increasing sequence in $\mathcal{B}$ then $\cup_{n=1}^\infty A_n \in \mathcal{B}$.

If $A_n \in \mathcal{B}$, $A_1 \subset A_2 \subset \dots$, let $B_1 = A_1$, $B_2 = A_2 - A_1$, $B_3 = A_3 - B_2$, and $B_{n+1} = A_{n+1} - \cup_{i=1}^n B_i$. It is clear that $B_n \in \mathcal{B}$ and $\{B_n\}$ are pairwise disjoint and $\cup_{i=1}^n A_n = \cup_{i=1}^n B_i$. We proved that a) implies 3.

Assume b). If B_n are pairwise disjoint sets from $\mathcal{B}$, letting $A_n = \cup_{i=1}^n B_i$. Then $A_n \in \mathcal{B}$ and A_n is an increasing sequence and $\cup_{n=1}^\infty A_n = \cup_{n=1}^\infty B_n$. So b) implies a).

Remark: If $\mathcal{A}$ is both a σ-additive class and a π system, then it ia an algebra. If $A, B \in \mathcal{A}$, $A \cup B = (A \backslash A \cap B) \cup (B \backslash A \cap B) \in \mathcal{A}$, hence we have a σ algebra. Clearly the condition $\Omega \in \mathcal{B}$ can be replaced that $\mathcal{A}$ is not empty. On the other hand σ-algebra is both a π system and a σ-additive class.

Theorem 1.4.5 (Dynkin's Lemma) *If $\mathcal{A}$ is a π system then the σ-additive class (Dynkin system) generated by $\mathcal{A}$ is the σ-algebra generated by $\mathcal{A}$.*

Proof First $\sigma(\mathcal{A}) \subset \mathcal{B}$ as it is a Dynkin system. Denote by $\mathcal{B}$ the σ-additive class generated by the π-system $\mathcal{A}$. We show that $\mathcal{B}$ is a π system. Let $A \subset B$, and

$$C^A = \{B \subset \mathcal{B} : A \cap B \in \mathcal{B}\}.$$

Then C^A is a Dynkin system and for $A \in \mathcal{A}$, $C^A = \mathcal{B}$. Hence $C^A \supset \mathcal{A}$ for $A \in \mathcal{B}$ and $C^A = \mathcal{B}$. Hence $\mathcal{B} \supset \sigma(\mathcal{A})$. $\square$

The moral of the section is: If we wish to prove a property holds for all sets in a σ algebra we only need to show that the collection of sets with the desired property contains an algebra and forms a monotone class. Or it should contain a π system and is a Dynkin system.

1.5 Measurable functions

Let $f : X \to Y$ be a mapping from one set X to a set Y. For $D \subset X$ we define the image of D by f,

$$f(D) = \{f(x) : x \in D\} \subset Y.$$

For $A \subset Y$ we define the pre-image of A by f,

$$f^{-1}(A) = \{x \in X : f(x) \in A\} \subset X.$$

A continuous function is a function such that the pre-image of open sets are open. Measurable functions are functions such that the pre-image of measurable sets are measurable.

Definition 1.5.1 *Given measurable spaces $(\Omega, \mathcal{F})$ and $(S, \mathcal{B})$, a function $f : \Omega \to S$ is said to be measurable if $f^{-1}(B) = \{\omega : f(\omega) \in B\}$ belongs to $\mathcal{F}$ for each $B \in \mathcal{B}$. Measurable functions are known as random variables, especially if $S = \mathbf{R}$.*

If f is identically 1 on a set A and zero everywhere else, it is called the indicator function of A and denoted by $\mathbf{1}_A$:

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

Example 1.5.2 Consider a function $f : E \to \mathbf{R}$ which only takes a finite number of values. They have the form,

$$f(x) = \sum_{j=1}^{n} a_j \mathbf{1}_{A_j}(x),$$

where a_j are values f take and $A_j = \{x : f(x) = a_j\}$. Then f is measurable if and only if A_j are measurable sets.

Definition 1.5.3 *A simple function is of the form*

$$f(x) = \sum_{j=1}^{n} a_j \mathbf{1}_{A_j}(x),$$

where $a_j \in \mathbf{R}$ and A_j are measurable sets.

Remark 1.5.4 To show that $f : (\Omega, \mathcal{F}) \to (E, \mathcal{B})$ is measurable, we only need to prove that $f^{-1}(A) \in \mathcal{F}$ for any set A from a π system G that generates $\mathcal{B}$. In fact note that

$$C = \{A \in \mathcal{B} : f^{-1}(A) \in \mathcal{F}\}$$

is a σ-algebra and $\sigma(G) \subset C \subset \mathcal{B}$.

Example 1.5.5 Let $W = C([0, T], \mathbf{R}^n)$ be the space of continuous functions from $[0, 1]$ to $\mathbf{R}^n$. It is a Banach space with the supremum norm:

$$\|g\|_W := \sup_{t \in [0,1]} \|g(t)\|.$$

A sample continuous real valued stochastic process $(X_t, 0 \le t \le 1)$ on a probability space $(\Omega, \mathcal{F}, P)$ can be considered as a function, denoted by X, from $(\Omega, \mathcal{F})$ to W.

$$X(\omega)(t) = X_t(\omega).$$

For each time t, and $a \in \mathbf{R}^n$, $\{\omega : |X_t(\omega) - a| < \epsilon\}$ belongs to $\mathcal{F}$ as X_t is measurable.

We show that the function $X : \Omega, \mathcal{F} \to W$ is Borel measurable. Take $f \in W$ and the open ball in the Banach space W centred at f with radius $\varepsilon > 0$. Its pre-image by X is:

$$\begin{aligned}\{\omega : \sup_{0\leq t\leq 1} |X_t(\omega) - f_{t_i}| < \epsilon\} &= \{\omega : \sup_{0\leq t_i\leq 1, t_i\in Q} |X_{t_i}(\omega) - f_{t_i}| < \epsilon\} \\ &= \cap_{0\leq t_i\leq 1, t_i\in Q}\{\omega : |X_{t_i}(\omega) - f_{t_i}| < \epsilon\}.\end{aligned}$$

Since for each i, $\{\omega : |X_{t_i}(\omega) - f_{t_i}| < \epsilon\} \in \mathcal{F}$, the intersection of the countable number of such sets belongs to $\mathcal{F}$.

1.5.1 Properties

If $\{A_\alpha, \alpha \in I\}$ is a collection of subsets of Y then

$$\begin{aligned} f^{-1}(A^c) &= (f^{-1}(A))^c \\ f^{-1}(\cap_{\alpha\in I} E_\alpha) &= \cap_{\alpha\in I} f^{-1}(E_\alpha) \\ f^{-1}(\cup_{\alpha\in I} E_\alpha) &= \cup_{\alpha\in I} f^{-1}(E_\alpha).\end{aligned}$$

Proposition 1.5.6

- *If $f_1 : (\Omega, \mathcal{F}) \to (E_1, \mathcal{A}_1)$ and $f_2 : (\Omega, \mathcal{F}) \to (E_2, \mathcal{A}_2)$ are measurable functions. Define the product σ-algebra*

  $$\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma\{A_1 \times A_2 : A_\in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}.$$

  Then $h = (f_1, f_2) : (\Omega, \mathcal{F}) \to (E_1\times, \mathcal{A}_1 \otimes \mathcal{A}_2)$ is measurable.

- *If $f : (X_1, \mathcal{B}_1) \to (X_2, \mathcal{B}_2)$ and $g : (X_2, \mathcal{B}_2) \to (X_3, \mathcal{B}_3)$ are measurable functions then so $f \circ g : (X_1, \mathcal{B}_1) \to (X_3, \mathcal{B}_3)$ is measurable.*

- *If $f, g, f_n : (X, \mathcal{B}) \to (\mathbf{R}, \mathcal{B}(\mathbf{R})$ are measurable functions, then the following functions are measurable: $f + g$, fg, $\max(f, g) = \frac{(f+g)+|f-g|}{2}$, $f^+ = \max(f, 0)$, f^-, $\limsup f_n$, $\liminf f_n$, $\sup_{n\geq 1} f_n$, $\inf_{n\geq 1} f_n$, $\lim_{n\to\infty} f_n$.*

For 1), define

$$C = \{A \in \mathcal{A}_1 \otimes \mathcal{A}_2 : h^{-1}(A) \in \mathcal{F}\}$$

Then σ-algebra C contains sets of the form $A_1 \times A_2$, $A_i \in \mathcal{B}_i$. The collection of sets of the form $\{A_1 \times A_2\}$ generates $\mathcal{A}_1 \otimes A_2$. For 2) we observe that $(f \circ g)^{-1}(A) = g^{-1}(f^{-1}(A))$.

For 3), note that $\otimes^n \mathcal{B}(\mathbf{R}) = \mathcal{B}(\mathbf{R}^n)$ and $f + g$ is the composition of $(a, b) \in \mathbf{R}^2 \mapsto (a + b) \in \mathbf{R}$, a continuous function with a measurable function $(f, g) : \Omega \to \mathbf{R}^2$. The other algebraic operations are proved similarly.

The measurability of $\sup_n f$ and $\inf f_n$ are seen from $\{x : \inf_{n\geq 1} f_n(x) < a\} = \cup_n \{x : f_n(x) < a\} \in \mathcal{F}$ and $\{x : \sup_{n\geq 1} f_n(x) < a\} = \cap_n \{x : f_n(x) < a\} \in F$. It follows that $\limsup_{n\to\infty} f_n = \inf_n \sup_{k\geq n} f_k$ is measurable. Now

$$U := \{x : \lim_{n\to\infty} f_n(x) \text{ exists}\} = \{x : \limsup_{n\to\infty} f_n(x) = \liminf_{n\to\infty} f_n(x)\}$$

is measurable. Consider the inverse of the diagonal Δ on $\mathbf{R}^2$ by $Z := (\liminf f_n, \limsup f_n)$. Then $U = Z^{-1}(\Delta)$ is measurable. Define $f(x) = 0$ on U. Then for $a < 0$, $\{x : \lim_{n\to\infty} f_n(x) < a\} = \{x : \limsup_{n\to\infty} f_n(x) < a\}$ and for $a \geq 0$, $\{x : \lim_{n\to\infty} f_n(x) < a\} = \{x : \limsup_n f_n(x) < a\} \cup U$.

Proposition 1.5.7 *If $f : E \to [0, \infty]$ is a positive measurable function, there is a sequence $\{f_n\}$ of simple functions such that*

$$0 \leq f_1 \leq f_2 \leq \dots f_n \leq \dots f,$$

and $f_n \to f$ pointwise. Furthermore $f_n \to f$ uniformly on any sets on which f is bounded. Here both E and $[0, \infty]$ are equipped with the relevant Borel σ-algebras.

Proof We consider $[0, 2^n]$ and $(2^n, \infty)$. Let

$$A^n_{2^n} = \{x : f(x) > 2^n\}.$$

Let us now consider dyadic partitions of the interval $[0, 2^n]$. The advantage of dyadic partitions are that as n gets bigger, the partition contains all partition points from the previous partitions. Let us cut each length 1 interval into 2^n pieces and there are $(2^n)^2 - 1$ subintervals of length 2^{-n}. Let

$$A^n_j = \left\{x : \frac{j}{2^n} < f(x) \leq \frac{j+1}{2^n}\right\}, \qquad 0 \leq j \leq 2^{2n} - 1.$$

We approximate f with value $\frac{j}{2^n}$ on A^n_j and define

$$f_n(t) = \sum_{j=0}^{2^{2n}} \frac{j}{2^n} \mathbf{1}_{A^n_j}(t).$$

Then $f_n \leq f_{n+1}$ and

$$|f(x) - f_n(x)| \leq \frac{1}{2^n}, \qquad \text{if } x \notin A^n_{2^n}.$$

Thus for each x, for all $n > f(x)$, $|f(x) - f_n(x)| \leq \frac{1}{2^n} \to 0$. $\square$

If $f : E \to \mathbf{R}$ is a measurable function, not necessarily positive, let

$$f = f^+ - f^-$$

where $f^+ = \max\{f(x), 0\}$ and $f^-(x) = \max\{-f(x), 0\}$. Approximate f^- and f^+ separately to give an approximation for f.

Theorem 1.5.8 *For every measurable function f there is a sequence of simple functions f_n such that for every x,*

$$\lim_{n\to\infty} f_n(x) = f(x).$$

If f is bounded, f_n can be chosen to be uniformly bounded.

Proposition 1.5.9 *The class of functions measurable with respect to a given σ algebra $\mathcal{F}$ is the smallest class of functions that contain all simple functions and is closed under pointwise limits.*

Proposition 1.5.10 *Let X, Y be a metric spaces and let $(X, \mathcal{B}(X))$ and $(Y, \mathcal{B}(Y))$ be the measurable spaces with Borel σ-algebras. If $f : X \to Y$ is continuous then it is Borel measurable .*

Proof A mapping $f : X \to Y$ is continuous if the pre-image of an open set is an open set. Since $\mathcal{B}(Y)$ is generated by open sets which forms a π-system

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

contains all open sets, $C = \mathcal{B}(Y)$. $\square$

Proposition 1.5.11 *Let $(\Omega, \mathcal{F})$ be a measure space. Let E be a π system that generates $\mathcal{F}$. Let $\mathcal{C}$ be a class of functions with the following property:*

1. *If $A \in E$, then $\mathbf{1}_A \in \mathcal{C}$;*
2. $1 \in \mathcal{C}$
3. *If $f, g \in \mathcal{C}$ and k a real number then $kf \in \mathcal{C}$, $f + g \in \mathcal{C}$.*
4. *(Monotone class property) if f is bounded and there is $f_n \in \mathcal{A}$ with $0 \le f_1 \le \cdots \le f_n \le \cdots \le K$ and $f = \lim_{n\to\infty} f_n$ pointwise then $f \in \mathcal{C}$.*

Then $\mathcal{C}$ contains the set of all bounded $\mathcal{F}$-measurable functions.

Proof We first show that $\mathbf{1}_B \in \mathcal{C}$ for all $B \in \mathcal{F}$.

$$J = \{B \in \mathcal{F} | \mathbf{1}_B \in \mathcal{C}\}$$

Then $J \supset E \cup \{\Omega, \emptyset\}$ and we show that J is a Dynkin system and hence the σ-algebra $\mathcal{F}$. The first two requirement for the Dynkin system following from that $\mathcal{C}$ is a linear space and $\mathbf{1}_B - \mathbf{1}_A = \mathbf{1}_{B-A}$. The disjoint union property follows from the monotone class property.

If f is positive and bounded measurable there is an sequence of positive uniformly bounded functions f_n converging to f and which are finite linear combinations of $\mathcal{F}$-measurable set. Finally if f is not positive let $f = f^+ - f^-$ $\square$

Definition 1.5.12 *Let $f : (\Omega, \mathcal{F}) \to (S, \mathcal{G})$ be a measure function. It pulls back the σ-algebra $\mathcal{G}$ to a σ-algebra on Ω. This is denoted by $\sigma(f)$ and is called the σ-algebra generated by f . It is*

$$\sigma(f) = \left\{ f^{-1}(A) : A \in \mathcal{G} \right\},$$

It is clear that $\left\{ f^{-1}(A) : A \in \mathcal{G} \right\}$ is a σ-algebra.

Example 1.5.13 If $f : (X, \mathcal{B}) \to (\mathbf{R}, \mathcal{B}(\mathbf{R})$ is of the form

$$f(x) = \sum_{j=1}^{n} a_j \mathbf{1}_{A_j}(x),$$

where A_j are subsets of X. Then $\sigma(f)$ is generated by the finite collection of sets $\{A_1, \dots, A_n\}$.

Proposition 1.5.14 *Let $(\Omega, \mathcal{F})$ be a measurable space and $X : \Omega \to \mathbf{R}$ a measurable function. Let $\sigma(X)$ be the σ-algebra generated by X. If $Z : \Omega \to \mathbf{R}$ is $\sigma(X)$-measurable then there is a Borel function $F : \mathbf{R} \to \mathbf{R}$ with $Z = F(X)$.*

Proof Proof. Consider the family

$U := \{Z : \Omega \to \mathbf{R}$ is $\sigma(X)$-measurable, $Z = f(X)$ for $f : \mathbf{R} \to \mathbf{R}$ Borel measurable $\}$.

It is easy to see that U is a linear space containing 1. We only need to show it is closed under pointwise convergence.

Let $f_n(X) \in U$ be a sequence with $\lim_{n\to\infty} f_n(X(\omega)) = Y(\omega)$. We must show that $Y \in U$. Define B to be the subset of $\mathbf{R}$ where $\lim_{n\to\infty} f_n(x)$ exists. Then $B \in \mathcal{B}(\mathbf{R})$. Define

$$f(x) = \begin{cases} \lim_{n\to\infty} f_n(x), & \text{if } \lim_{n\to\infty} f_n(x) \text{ exists} \\ 0, & \text{if } \lim_{n\to\infty} f_n(x) \text{ does not exist} \end{cases}$$

Then $f_n(X(\omega)) \to f(X(\omega))$ when $\omega \in \{\omega : X(\omega) \in B\}$. Hence $Y(\omega) = f(X(\omega))$.

$\square$

1.6 Integration with respect to a measure

Let $(E, \mathcal{F}, \mu)$ be a measure space. Let

$$\mathcal{E} = \{f(x) = \sum_{j=1}^{n} a_j \mathbf{1}_{A_j}(x) : A_j \in \mathcal{F}, a_j \in \mathbf{R}\}$$

be the set of (measurable) simple functions. We define its integral with respect to μ as below:

$$\int_{x \in E} f \sum_{j=1}^{n} a_j \mathbf{1}_{A_j}(x)\ d\mu(x) = \sum_{j=1}^{n} a_j \mu(A_j).$$

Note that if a_j are distinct, then $A_j = f^{-1}(\{a_j\})$.

Observe that if f, g are two simple functions with $0 \le f \le g$ a.s.

$$\int_E f d\mu \le \int_E g d\mu.$$

In particular if $f = g$ a.s. then $\int f = \int g$.

Definition 1.6.1 *If $f : E \to [0, \infty]$ be a positive measurable function, define*

$$\int f d\mu = \sup_{g \in \mathcal{E} : g \le f} \int g d\mu.$$

Theorem 1.6.2 (Monotone Convergence Theorem) *If $f_n : E \to [0, \infty]$ is an increasing sequence of positive functions,* $\lim_{n\to\infty} f(x)$ *exists almost surely and*

$$\int_E \lim_{n\to\infty} f(x) d\mu(x) = \lim_{n\to\infty} \int_E f_n(x) d\mu(x).$$

If $f : E \to [0, \infty]$ is a positive measurable function, there is a sequence of increasing simple functions f_n converging pointwise to f. Then $\lim_{n\to\infty} \int f_n(x) d\mu(x)$ is an increasing sequence and has a limit if it is bounded. If this limit exists,

$$\int_E f(x) d\mu(x) = \lim_{n\to\infty} \int_E f_n(x) d\mu(x).$$

Proposition 1.6.3 *If f, g are positive measurable, $a, b \geq 0$, then*

1. $\int (af + bg) d\mu = a \int f \, d\mu + b \int g \, d\mu$.
2. *If* $\int f d\mu < \infty$ *then* $f < \infty$ *almost surely.*
3. $\int f d\mu = 0$ *if and only if* $f = 0$ *a.s.*

Definition 1.6.4 *Let $f : E \to \mathbf{R}$ be a general measurable function. Write $f = f^+ - f^-$. If both $\int f^+ d\mu < \infty$ and $\int f^- \, d\mu < \infty$ are finite, we say that f is integrable with respect to μ and the integral is defined to be:*

$$\int_E f d\mu = \int_E f^+ \, d\mu - \int_E f^- \, d\mu.$$

The set of all integrable functions are denoted by $L^1(\mu)$ and called L^1 functions. If A is a Borel measurable set, we write $\int_E \mathbf{1}_A(x) f(x) d\mu(x)$ as following:

$$\int_E \mathbf{1}_A(x) f(x) \, d\mu(x) \equiv \int_A f(x) \, \mu(dx) = \int_A f \, d\mu.$$

Proposition 1.6.5 *Assume that $f, g \in L^1(\mu)$.*

1. *If $f \in L^1(\mu)$, so does $|f| \in L^1(\mu)$, and*
$$\left| \int f d\mu \right| \leq \int |f| \, d\mu.$$
2. *If $f, g \in L^1(\mu)$, $a, b, \in \mathbf{R}$, then $af + bg \in L^1(\mu)$ and*
$$\int (af + bg) \, d\mu = a \int f \, d\mu + b \int g \, d\mu.$$
3. *If $f \leq g$,*
$$\int f d\mu \leq \int g d\mu.$$
4. *If A is a measurable set, define $\int_A f d\mu = \int \mathbf{1}_A f d\mu$. Then*
$$\int_A f \, d\mu = \int_A g \, d\mu$$
for all measurable set A implies that $f = g$ almost surely.
5. *If $f \geq 0$ and $\int f d\mu = 0$ then $f = 0$ a.e.*

1.6.1 Integration with respect to the Pushed Forward Measures

Suppose we are given a measure space $(X, \mathcal{A}, \mu)$ and a function f from X to a measurable space $(Y, \mathcal{B})$. Then there is the **pushed forward (induced) measure** $f_*(\mu)$ on $(Y, \mathcal{B})$ defined as follows: $f_*(\mu)(B) = \mu\{x : f(x) \in B\}$.

Lemma 1.6.6 *If α is a real-valued function on Y, then*

$$\int_X \alpha f \, d\mu = \int_Y \alpha \, df_*(\mu). \tag{1.2}$$

Let $\phi : S \to \mathbf{R}$ and $X : \Omega \to \mathbf{R}$ be measurable functions, define

$$\mathbf{E}[\phi(X)] := \int_\Omega \phi(X(\omega)) \, dP(\omega) = \int_S \phi(y) \; X_*(\mu)(dy).$$

Definition 1.6.7 *If f is a measurable map from X to X, we say μ is* **invariant** *by f if $f_*(\mu) = \mu$.*

Example 1.6.8 Let δ_x be the **Dirac measure** on $\mathbf{R}^n$:

$$\delta_x(A) = \begin{cases} 0, & \text{if } x \in A \\ 1, & \text{if } x \not\in A \end{cases}$$

Then for any transformation $T : \mathbf{R}^n \to \mathbf{R}^n$, $T_*(\delta_x) = \delta_{Tx}$. The δ-measure is not invariant under rotations, unless $x = 0$, or by translation.

1.6.2 Lebesgue Integrals

Integrals developed in the last section, in the case of μ being a Lebesgue measure, are called Lebesgue integrals.

Let a, b be real numbers. Consider the measure space $([a, b], \mathcal{B}([a, b])$. Let μ be the the Lebesgue measure on $[a, b]$. It is determined by its value on set of the form $A = \cup_{i=1}^n (c_i, d_i)$ (union of disjoint intervals),

$$\mu(A) = \sum_i^n (d_i - c_i).$$

Consider $a = t_0 < t_1 < \cdots < t_n = b$. Let

$$f(x) = f_0 \mathbf{1}_0(x) + \sum_{j=0}^{n-1} a_j \mathbf{1}_{(t_j, t_{j+1}]}(x).$$

Since $\mu(\{0\}) = 0$,

$$\int_a^b f(x)dx = \sum_{j=0}^{n-1} a_j(t_{j+1} - t_j).$$

1.7 Riemann Integrals and Riemann-Stieltjes Integrals

If a function $f : [a, b] \to \mathbf{R}$ is Riemann integrable, it is Lebesque integrable and the two integrals agree. In this and the next section we review Riemann integrals and Riemann-Stieltjes integrals.

Let $f : [a, b] \to \mathbf{R}$ be a bounded function. Let $a = t_0 < t_1 < \cdots < t_n = b$ be a partition of $[a, b]$. Let $m_i = \max\{f(x), x \in [t_i, t_{i+1}]\}$, $M_i = \max\{f(x), x \in [t_i, t_{i+1}]\}$. If for all partitions Δ

$$\inf_{\Delta} M_i(t_{i+1} - t_i) = \sup_{\Delta} \sum_i m_i(t_{i+1} - t_i),$$

the common value it called the Riemann integral and f is Riemann integrable. If so, let $\Delta_n : a = t_0^n < t_1^n < \cdots < t_n^n = b$ be a sequence of partitions of $[a, b]$ with mess converging to zero, then

$$\int_a^b f(x)dx = \lim_{n\to\infty} \sum_j f(t_j^n)(t_{j+1}^n - t_j^n).$$

Proposition 1.7.1 *A bounded function $f : [a, b] \to \mathbf{R}$ is Riemann integrable if and only if the set of discontinuity of f has Lebesque measure zero.*

Proposition 1.7.2 *If a bounded function $f : [a, b] \to \mathbf{R}$ is Riemann integrable, then it is Lebesgue measurable and Lebesgue integrable. Furthermore, the integrals have the common value:*

$$\int_{[a,b]} f(x)d\mu = \int_a^b f(x)dx.$$

Let Δ be the dyadic partition of $[a, b]$: $t_j = \frac{(b-a)}{2^n}$. Consider

$$f_n(t) = \sum_j \sup\{f(x) : x \in [t_i, t_{i+1}]\}\mathbf{1}_{(t_i,t_{i+1}]}(t).$$

Then f_n is measurable and converges to f a.e. Check that the integrals agree on functions of the form:

$$\sum_{i=1}^{n} \alpha_i \mathbf{1}_{(c_i,d_i)}$$

where (c_i, d_i) are disjoint intervals.

Definition 1.7.3 *Let $f, g : [a, b] \to \mathbf{R}$ be bounded functions. We say f is* **Riemann-Stieltjes integrable** *with respect to g if there is a number l such that for all $\epsilon > 0$, there is δ such that for all partitions*

$$\Delta : a = t_0 < x_1 < \cdots < t_n = b$$

with mesh $|\Delta| := \max_{1\le i\le n}(t_i - t_{i-1}) < \delta$,

$$\left| \sum_{j=0}^{n} f(t_j^*) \left[g(t_{j+1}) - g(t_j) \right] - l \right| < \epsilon.$$

Here t_j^* is any point in $[t_{j-1}, t_j]$. If so we write $\lim_{|\Delta|\to 0} \sum_j f(t_j) \left(g(t_{j+1}) - g(t_j) \right) = l$ and define

$$\int_a^b f(t)\, dg(t) = \lim_{|\Delta|\to 0} \sum_j f(t_j) \left(g(t_{j+1}) - g(t_j) \right)$$

to be **Riemann-Stieltjes integral** *of f with respect to g.*

Take $g(x) = x$ to obtain Riemann integrals.

The Riemann-Stieltjes integral has the remarkable symmetry property:

Proposition 1.7.4 *Let $f, g : [a, b] \to \mathbf{R}$ be bounded functions. Then f is Riemann-Stieltjes integrable with respect to g if and only if g is Riemann-Stieltjes integrable with respect to f. If the integrals exist we have*

$$\int_0^t f dg + \int_a^b g df = f(b)g(b) - f(a)g(a).$$

The most commonly used integrators are functions of bounded variations. Functions of bounded variations are differences of increasing functions. See section 1.12.1. Riemann-Stieljes integrals do not exists if the integrator and the integrand have the same point of continuity. That $\int_a^v f dF$, $\int_b^c f dF$ exist in Riemann-Stieljes sense do not necessarily imply that $\int_a^c f dF$ is Riemann-Stieljes integrable. The Lebesque-Stieljes integrals, see Example 1.3.8, with respect to a bounded variation function ϕ has a larger class of integrands.

Proposition 1.7.5 *If $g; [a,b] \to \mathbf{R}$ is continuous and $F : [a,b] \to \mathbf{R}$ has bounded variation then the Riemann-Stieljes integral $\int_a^b g dF$ and $\int_a^b F dg$ exist. If F is furthermore absolutely continuous and the integral equals to the corresponding Lebesgue integral:*

$$\int_a^b g dF = \int_a^b g F' dx.$$

1.8 Random variables

Given a probability measure μ on $\mathbf{R}$, we define its mean to be

$$m_\mu = \int_{-\infty}^{\infty} x \; \mu(dx)$$

and its variance by

$$\sigma_\mu^2 = \int_{-\infty}^{\infty} (x - m_\mu)^2 \mu(dx).$$

1.8.1 Law of a Random Variable

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $f : (\Omega, \mathcal{F}) \to (\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$ a measurable function. Then f induces the pushed forward measure on $\mathbf{R}^n$, which is denoted by $f_*(P)$.

Definition 1.8.1 *Let $X : \Omega \to \mathbf{R}$ be a random variable. The law or the probability distribution of X is a measure on $\mathbf{R}^n$ is the pushed forward measure and denoted by P_X:*

$$P_X(A) = P(\omega : X(\omega) \in A).$$

The distribution function for X is defined to be that for its law:

$$F_X(x) = P(\omega : X(\omega) \le x).$$

If $\phi : \mathbf{R}^n \to \mathbf{R}$ is a Borel measurable function, $\phi(X)$ is a measurable function. Its expectation is defined by

$$\mathbf{E}\phi(X) := \int_\Omega \phi(X) dP = \int_{\mathbf{R}} \phi(y) P_X(dy).$$

If $F_X(x)$ takes the following special form:

$$F_X(x) = \int_{-\infty}^{x} p(y) dy$$

we say the P_X has a density p with respect to the Lebesque measure, in which case

$$\mathbf{E}[\phi(X)] = \int_{-\infty}^{\infty} \phi(y)\, p(y) dy.$$

We define the expectation or mean of X to be that of its distribution.

$$\mathbf{E}X = \int X dP = \int y dP_X(dy).$$

The variance of the random variable is:

$$\mathbf{var}(X) = \mathbf{E}(X - \mathbf{E}X)^2 = \int (y - \mathbf{E}X)^2 dP_X(y);$$

The covariance of two random variables is $\mathbf{cov}(X, Y) = \mathbf{E}(X - \mathbf{E}X)(Y - \mathbf{E}Y)$. Let $P_{X,Y}$ be the distribution of (X, Y), a measure on $\mathbf{R}^2$, then

$$\mathbf{cov}(X, Y) = \int (x - \mathbf{E}X)(y - \mathbf{E}Y) dP_{X,Y}(x, y).$$

1.8.2 Multi-variate Random Variables

Let X be an $\mathbf{R}^n$-valued random variable with entries $X_1, \dots, X_n$. Its average is $\mathbf{E}X = (\mathbf{E}X_1, \dots, \mathbf{E}X_n)$. Given $Y = (Y_1, \dots, Y_n)^T$ the covariance matrix between X and Y is a $n \times n$ matrix who entries are: $\mathbf{cov}(X_i, Y_j)$,

$$\mathbf{cov}(X, Y) = \mathbf{E}(X - \mathbf{E}X)(Y - \mathbf{E}Y)^T$$

where T means transpose of the matrix. The variance of X is defined as:

$$\mathbf{var}(X, X) = \mathbf{E}(X - \mathbf{E}X)(X - \mathbf{E}X)^T$$

Note that $u^T X$ is a real valued random variable for $u \in \mathbf{R}^n$ and so

$$\begin{aligned}
\mathbf{var}(u^T X) &= \mathbf{E}(u^T X - \mathbf{E}u^T X)^2 = \mathbf{E}(u^T X - \mathbf{E}u^T X)(u^T X - \mathbf{E}u^T X)^T \\
&= \mathbf{E}u^T (X - \mathbf{E}X)(X^T - \mathbf{E}X^T) u = u^T \mathbf{var}(X) u
\end{aligned}$$

The symmetric matrix $\mathbf{var}(X)$ can be considered as a linear operator, called the covariance operator:

$$\mathbf{cov}(X)(u, v) = u^T \mathbf{var}(X) v.$$

A random variable $X = (X^1, \dots, X^n)$ determines a probability distribution, that is a measure on $\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n)$: for A_i Borel subsets of $\mathbf{R}^n$,

$$P_X(\Pi_{i=1}^n A_i) := P(\{\omega : X_1(\omega) \in A_1, \dots, X_n(\omega) \in A_n\})$$

If X is one dimensional, define $F_X(x) = P(X \le x)$, the distribution function of X. Then F_X determines the Borel measure $\{P_X(A), A \in \mathcal{B}\}$.

1.8.3 Characteristic Functions

The Fourier transform of an L^1 function on $\mathbf{R}^n$ is $\hat{f}(\lambda) = \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbf{R}^n} e^{-i\langle \lambda, x\rangle} f(x)dx$, where $\lambda \in \mathbf{R}^n$ and inverse transform is $\breve{f}(\lambda) = \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbf{R}^n} e^{i\langle \lambda, x\rangle} f(x)dx$. The choice of the constants $\frac{1}{(2\pi)^{\frac{n}{2}}}$ and the sign in the exponential are by convention.

The characteristic function of a Borel probability measure on $\mathbf{R}^n$ is is the complex valued function on $\mathbf{R}^n$:

$$\hat{\mu}(\lambda) = \int_{\mathbf{R}^n} e^{i\langle \lambda, x\rangle} \mu(dx).$$

This is the Fourier transform of the measure consider to be a distribution, in the dual of continuous functions vanishing at infinity. From Fourier analysis it is easy to see Bochner's theorem:

Theorem 1.8.2 *If two probability measures on $\mathbf{R}^n$ have the same Fourier transform they coincide.*

The characteristic function of random variable X with values in $\mathbf{R}^n$ is that of its distribution.

$$\phi_X(\lambda) = \mathbf{E} e^{i\langle \lambda, X\rangle} = \mathbf{E} e^{i\lambda_i X^i}.$$

Theorem 1.8.3 *A family of real random variables $(Y_1, \dots, Y_k)$ are independent if and only if their joint characteristic function is the product of their marginal characteristic functions:*

$$\phi_{(Y_1,\dots,Y_k)}(\lambda_1, \dots, \lambda_k) = \Pi_{j=1}^k \phi_{Y_j}(\lambda_k).$$

1.8.4 Characteristic Functions of Gaussian measures

Definition 1.8.4 *A Borel measure on $\mathbf{R}$ is determined by*

$$\mu(A) = \frac{1}{\sqrt{2\pi}\sigma} \int_A e^{\frac{-(x-a)^2}{2\sigma^2}} dx.$$

This measure is called the normal distribution or Gaussian measure on $\mathbf{R}$ with mean m and variance σ^2 and denoted by $N(a, \sigma^2)$.

The function $e^{\frac{-(x-a)^2}{2\sigma^2}}$ is the density for the Gaussian measure $N(a, \sigma^2)$. Note that

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{\frac{-(x-a)^2}{2\sigma^2}} dx = 1,$$

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x e^{\frac{-(x-a)^2}{2\sigma^2}} dx = a, \qquad \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x^2 e^{\frac{-(x-a)^2}{2\sigma^2}} dx = \sigma^2.$$

Definition 1.8.5 *A measure on $\mathbf{R}^n$ is Gaussian if for each linear functional l on $\mathbf{R}^n$, the pushed forward measure $l_*(\mu)$ is Gaussian on $\mathbf{R}$.*

Let $x, a \in \mathbf{R}^n$ and $\langle , \rangle$ denotes the Euclidean inner product. A non-degenerate Gaussian measure on $\mathbf{R}^n$ is absolutely continuous with respect to the Lebesque measure and has density

$$\frac{1}{\sqrt{(2\pi)^n \det C}} e^{-\frac{1}{2}\langle C^{-1}(x-a),(x-a)\rangle}.$$

where C is a positive definite symmetric $n \times n$ matrix. Let $x = (x_1, \dots, x_n)$, $dx = dx_1 \dots dx_n$, then

$$\int_{\mathbf{R}^n} f \, d\mu = \frac{1}{\sqrt{(2\pi)^n \det C}} \overbrace{\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty}}^{n} f(x) e^{-\frac{1}{2}\langle C^{-1}(x-a),(x-a)\rangle} dx.$$

Definition 1.8.6 *A random variable is a Gaussian random variable if its distribution is a Gaussian measure.*

Proposition 1.8.7 *Suppose that (X, Y) is Gaussian random variable. Then they are independent iff they are uncorrelated, that is $\mathbf{E}(X - \mathbf{E}X)(Y - \mathbf{E}Y) = 0$.*

This is not so if they are only individually Gaussian without being jointly Gaussian.

Example 1.8.8 Take a Gaussian random variable X with mean zero. Let Z be a Bernolli random variable taking values $\{1, -1\}$ and define $Y = ZX$. Then

$$\mathbf{E}e^{izY} = \mathbf{E}e^{izX}\mathbf{1}_{\{Z=1\}} + \mathbf{E}e^{-izX}\mathbf{1}_{\{Z=1\}} = e^{\frac{-z^2}{2}},$$

so Y is a Gaussian random variable. $\mathbf{E}Y = \mathbf{E}(XZ\mathbf{1}_{\{Z=1\}}) + \mathbf{E}(XZ\mathbf{1}_{\{Z=-1\}}) = \mathbf{E}(X\mathbf{1}_{\{Z=1\}}) - \mathbf{E}(X\mathbf{1}_{\{Z=-1\}}) = \mathbf{E}XP(Z=1) - \mathbf{E}XP(Z=-1) = 0$.

$$\mathbf{cov}(X, Y) = \mathbf{E}(X^2 Z) = \mathbf{E}(X^2)P(Z=1) - \mathbf{E}X^2 P(Z=-1) = 0.$$

But X and Y are far from independent. This means the pair (X, Y) is not jointly Gaussian.

Lemma 1.8.9 *For α a complex number,*

$$(2\pi)^{-\frac{n}{2}} \frac{1}{\sqrt{\det C}} \int_{\mathbf{R}^n} e^{-\frac{\langle C^{-1}(x-a),(x-a)\rangle}{2}} e^{\alpha\langle x,\lambda\rangle} dx = e^{\alpha\langle a,\lambda\rangle + \frac{\alpha^2}{2}\langle C\lambda,\lambda\rangle}.$$

In particular, taking $\alpha = i$, the characteristic function of a Gaussian measure centred at a is

$$\phi(\lambda) =: \mathbf{E}e^{i\langle\lambda,X\rangle} = e^{i\langle a,\lambda\rangle - \frac{1}{2}\langle C\lambda,\lambda\rangle}.$$

Proof First assume $\alpha \in \mathbf{R}$, then the left hand side (L.H.S) of the required equality equals

$$\begin{aligned} & \frac{1}{\sqrt{(2\pi)^n \det C}} \int_{\mathbf{R}^n} \exp^{-\frac{1}{2}\langle C^{-1}(x-\alpha Cy),\, x-\alpha Cy\rangle} e^{\frac{\alpha^2}{2}\langle Cy,y\rangle} e^{\alpha\langle a,y\rangle} dx \\ = & \frac{1}{\sqrt{(2\pi)^n \det C}} e^{\alpha\langle a,y\rangle + \frac{\alpha^2}{2}\langle Cy,y\rangle} \int_{\mathbf{R}^n} e^{-\frac{1}{2}\langle C^{-1}x,x\rangle} dx \\ = & e^{\alpha\langle a,y\rangle + \frac{\alpha^2}{2}\langle Cy,y\rangle} \end{aligned}$$

by translation invariance of the Lebesgue measures. The result holds for $\alpha \in \mathcal{C}$ since both sides of the equality are analytic functions of α and they agree on $\mathbf{R}$.

$\square$

Note also that if the $\mathbf{R}^2$ valued (X, Y) is a Gaussian random variable then X and Y are independent if and only if $\mathbf{cov}(X, Y) = 0$.

The characteristic function of $N(a, \sigma^2)$ on $\mathbf{R}^1$ is $e^{i\langle a,\lambda\rangle - \frac{1}{2}\sigma^2\lambda^2}$, $\lambda \in \mathbf{R}$.

1.9 L^p spaces

Let $(E, \mathcal{B}, \mu)$ be a measure space. Two functions $f, g : \Omega \to \mathbf{R}$ are considered to be in the same equivalent class if $f = g$ almost surely. For $1 \le p < \infty$, denote by $L^p(E, \mathcal{B}, \mu)$ the L^p space:

$$L^p(E, \mathcal{B}, \mu) = \{f : E \to \mathbf{R} : \text{ measurable, } \int_E |f(x)|^p \, d\mu(x) < \infty\}.$$

Note that $\int |f|^p d\mu = 0$ if and only if $|f| = 0$ almost surely. That L^p is a vector space is clear from the elementary inequality $|a + b|^p \le C_p|a|^p + C_p|b|^p$. Define $\|f\|_{L_p} = \left(\int_E |f(x)|^p d\mu(x)\right)^{\frac{1}{p}}$. Let L^∞ be the set of bounded measurable functions with $|f|_{L_\infty} = \inf\{a : \mu(|f(x)| > a) > 0\}$. Then L^p spaces are Banach space. If there is only one measure involved, the notation $L^p(E, B)$ maybe used. Similarly the notations $L^p(E)$ and L^p are also used for simplicity.

Hölder's inequality (Cauchy-Schwartz if $p = q = 2$).

$$\int |fg| d\mu \le \left(\int |f|^p d\mu\right)^{\frac{1}{p}} \left(\int |g|^q d\mu\right)^{\frac{1}{q}}, \qquad \frac{1}{p} + \frac{1}{q} = 1.$$

If μ is a finite measure, by Hölder inequality, if $p < q$ we have $L^p \subset L^q$. There is also the Minkowski Inequality, $p \geq 1$,

$$\|f+g\|_{L_p} \leq \|f\|_{L_p} + \|g\|_{L_p}.$$

A function $\phi : \mathbf{R} \to \mathbf{R}$ is convex if $\phi(\sum p_i x_i) \leq \sum_i p_i \phi(x_i)$. If ϕ is twice differentiable f is convex if and only if $\phi^{''} \geq 0$. Example of convex functions are: $\phi(x) = x^p$ for $p > 1$, $\phi(x) = e^x$.

Theorem 1.9.1 (Jensen's Inequality) *If ϕ is a convex function then for all integrable functions f and μ a probability measure*

$$\phi\left(\int f d\mu\right) \leq \int \phi(f) d\mu.$$

Proposition 1.9.2

$$|\int f d\mu| \leq \int |f| d\mu.$$

and

Proposition 1.9.3 (Chebychevs inequality) *If X is an L^p random variable then for $a \geq 0$,*

$$P(|X| \geq a) \leq \frac{1}{a}\mathbf{E}|X|^p.$$

Theorem 1.9.4 *If $f \in L^1$. For any $\epsilon > 0$ there is a simple function ϕ such that $\int |f - \phi| < \epsilon$.*

We say that simple functions are dense in L^1 or an L^1 function can be approximated by simple functions in the L^1 norm.

Theorem 1.9.5 (The monotone convergence theorem) *If f_n are non-decreasing sequence converging to f almost surely, and $f_n^- \leq g$ and g is integrable then*

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

Proof This is shown when $f_n \geq 0$. Otherwise take $g_n = f_n + g$. Then $g_n \geq 0$ and $\{g_n\}$ is a non-decreasing sequence. $\square$

Theorem 1.9.6 (Dominated Convergence Theorem) *If $|f_n| \leq g$ where g is an integrable function, and $f_n \to f$ a.e. then f is integrable*

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

Theorem 1.9.7 (Fatou's lemma) *If f_n is bounded below by an integrable function then*

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

1.10 Convergence of Functions

1.10.1 Notions of convergence of Functions

Definition 1.10.1 *Let $(E, \mathcal{B}, \mu)$ be a measure space. Let f_n be a sequence of measurable functions.*

- *f_n converges to f almost surely if there is a set Ω_0 with $\mu(\Omega_0) = 0$*

$$\lim_{n\to\infty} f_n(\omega) = f(\omega), \qquad \forall \omega \notin \Omega_0.$$

- *The sequence f_n is said to converge to f in measure if for any $\delta > 0$*

$$\lim_{n\to\infty} P(|f_n - f| > \delta) = 0.$$

- *L^p convergence*

$$\int |f_n - f|^p dP \to 0.$$

Example 1.10.2 1. (Example of a sequence which converges in probability but not almost surely) Let A_n be measurable sets. The indicator functions $\mathbf{1}_{A_n} \to 0$ in probability if and only if $\mu(A_n) \to 0$. Just note that for any $\delta > 0$, $\{\omega : \mathbf{1}_{A_n}(\omega) > \delta\} = A_n$.

Define a sequence of sets $A_{j,n} = [\frac{j}{n}, \frac{j+1}{n}]$ for all n and all j. Define the sequence of sets $A_{0,1}, A_{0,2}, A_{1,2}, \dots, A_{0,n}, \dots, A_{1,n} \dots A_{n-1,n}, \dots$. The Lebesque measure of the set goes to zero. But at any point, it had values 0 and 1 infinitely often, and the sequence of functions given by the indicator functions is nowhere convergent.

2. A convergent sequence of functions which does not converge in L^1. Take $f_n = \mathbf{1}_{[n,n+1]}$. Then $f_n \to 0$. But $\int_{\mathbf{R}} |f_n - f| dx = 1$.

   Take $g_n = n\mathbf{1}_{\{[0,\frac{1}{n}]\}}$. Then $g_n \to 0$ and $\int_0^1 |g_n - g| dx = 1$.

Heuristically if $f_n \to f$ and the mass of f_n escapes to 'infinity', then f_n may not converge to f in L^1, or f becomes unbounded, which also amounts to an escape of mas

Proposition 1.10.3 *Let A_n be measurable sets. The indicator functions $\mathbf{1}_{A_n} \to 0$ in probability if and only if $\mu(A_n) \to 0$.*

Just note that for any $\delta > 0$, $\{\omega : \mathbf{1}_{A_n}(\omega) > \delta\} = A_n$.

1.11 Convergence Theorems

Theorem 1.11.1 (Egoroff's Theorem) *Suppose that $\mu(E) < \infty$. Let f_n be a sequence of functions converging to f almost surely. Then for any $\epsilon > 0$ there is a set $E_\epsilon \subset E$ with $\mu(E_\epsilon^c) \le \epsilon$ such that f_n converges to f uniformly on E_ϵ. In particular f_n converges to f in measure.*

Proof Let

$$U_n^m = \cup_{k \ge m}\{x : |f_k(x) - f(x)| \ge \frac{1}{n}\}.$$

Then $x \notin U_n^m$ means that $|f_k(x) - f(x)| < \frac{1}{n}$ when $k \ge m$.

$$\lim_{m\to\infty} U_n^m = \cap_{m=1}^\infty \cup_{k \ge m} \{x : |f_k(x) - f(x)| \ge \frac{1}{n}\},$$

has measure zero from the almost sure convergence of f_k. Hence $\lim_{m\to\infty} \mu(U_n^m) = 0$. Choose $m(n,\epsilon)$ so that $\mu(U_n^{m(n,\epsilon)}) \le \epsilon \frac{1}{2^n}$. Let $U_\epsilon = \cup_{n=1}^\infty U_n^{m(n,\epsilon)}$ whose mass is smaller than ϵ. The compliment of the set is $\cap_{n\ge 1} \cap_{k \ge m(n,\epsilon)} \{x : |f_k(x) - f(x)| < \frac{1}{n}\}$. On this set, $|f_k(x) - f(x)| \ge \frac{1}{n}$ when $k \ge m(n,\epsilon)$. $\square$

Proposition 1.11.2

1. *If f_n converges to f in L^p, $p \ge 1$, it converges to f in measure.*
2. *If f_n converges to f in measure there is a subsequence of f_n which converges almost surely.*
3. *f_n converges to f in measure if and only if every subsequence has a further subsequence that converges a.s.*

1.11.1 Uniform Integrability

For $p \ge 1$, let $L^p \equiv L^p(\Omega, \mathcal{F}, \mu)$ be the equivalence class of real valued functions with $\mathbf{E}|f|^p < \infty$.

Lemma 1.11.3 *Assume that $\mu(\Omega) < \infty$. Take $f \in L^1$. The following holds.*

1. *Given any $\epsilon > 0$ there is $\delta > 0$ such that if $\mu(A) < \delta$ then*

$$\int_A |f| d\mu < \epsilon.$$

2. *Given any $\epsilon > 0$ there is $C > 0$ such that*

$$\int_{|f|>C} |f| d\mu < \epsilon.$$

Proof Suppose that there is $\epsilon_0 > 0$ such that there is A_k with $\mu(A_k) < \frac{1}{2^k}$ and $\int_{A_k} |f| d\mu > \epsilon_0$. Then

$$\epsilon_0 < \liminf_{k \to \infty} \int_{A_k} |f| d\mu \le \int \limsup \mathbf{1}_{A_k} |f| d\mu.$$

But $\mu(\limsup \mathbf{1}_{A_k}) = 0$ by the Borel Cantelli lemma and the right hand side equals zero.

Note that $\mu(\{|f| > C\}) \le \frac{1}{C} \int |f| d\mu$ is a small set and hence part 2 follows. $\square$

Definition 1.11.4 *A family of real-valued functions $(f_\alpha, \alpha \in I)$, where I is an index set, is uniformly integrable (u.i.) if for any $\epsilon > 0$ there is $C > 0$ such that for all α*

$$\int_{\{|f_\alpha| \ge C\}} |f_\alpha| d\mu < \epsilon$$

Note that if $\sup_\alpha \mathbf{E}|f_\alpha|^p < \infty$ for some $p > 1$, the family is uniformly integrable.

Proposition 1.11.5 *Let μ be a finite measure. A family of real-valued functions $(f_\alpha, \alpha \in I)$ is uniformly integrable (u.i.) if and only if*

1. $\sup_{\alpha \in I} \int |f_\alpha| d\mu < \infty$

2. *The family f_α is uniformly absolutely continuous: Given $\epsilon > 0$ there is a $\delta > 0$ such that if a measurable set A has $\mu(A) < \delta$ then for all $\alpha \in I$*

$$\int_A |f_\alpha| d\mu < \epsilon.$$

Proof Suppose that u.i. holds. For any $0 < \epsilon < 1$, take C such that $\int_{|f_\alpha|\ge C} |f_\alpha| d\mu \le \epsilon/2$,

$$\sup_\alpha \int |f_\alpha| d\mu \le \int_{|f_\alpha|\ge C} |f_\alpha| d\mu + \sup_\alpha \int_{|f_\alpha|<C} |f_\alpha| \le \frac{\epsilon}{2} + C\mu(\Omega) < \infty,$$

since $\mu(\Omega) < \infty$. Also

$$\sup_\alpha \int_A |f_\alpha| d\mu \le \sup_\alpha \int_{A\cap\{|f_\alpha|\ge C\}} |f_\alpha|\, d\mu + \sup_\alpha \int_{A\cap\{|f_\alpha|<C\}} |f_\alpha| \le \frac{\epsilon}{2} + C\mu(A) \to 0$$

as $\mu(A) \to 0$.

Assume that X_t is uniformly absolutely continuous and is L^1 bounded, $\sup_\alpha \mu(|f_\alpha| \ge C) \le \frac{1}{C} \sup_\alpha \int |f_\alpha| d\mu$ is mall when C is large. By choosing C large we see that $\int_{|f_\alpha|\ge C} |f_\alpha| d\mu$ is as small as we want. □

Proposition 1.11.6 *Let $f_n, f \in L^1$.*

a) Suppose that $f_n \to f$ in measure and f_n is uniformly integrable. Then $\int |f_n - f| d\mu \to 0$.

b) Suppose that $\int |f_n - f| d\mu \to 0$ and $\mu(\Omega) < \infty$. Then $f_n \to f$ in measure and f_n is uniformly integrable.

Proof Proof of a). For any $\epsilon > 0$ choose $C > 1$ so that

$$\sup_n \int_{\{|f_n|\ge C\}} |f_n| d\mu + \int_{\{|f|\ge C\}} |f| d\mu < \epsilon.$$

Choose N such that if $n > N$, $\mu(|f_n - f| > \epsilon) < \frac{\epsilon}{2C}$

$$\begin{aligned}
\int |f_n - f| d\mu &\le \sup_n \int_{|f_n - f|\le\epsilon} |f_n - f| d\mu\; + \left(\int_{\{|f_n-f|>\epsilon\}\cap\{|f_n|\ge C\}} + \int_{\{|f_n-f|>\epsilon\}\cap\{|f_n|\le C\}} \right) |f_n| d\mu \\
&+ \left(\int_{\{|f_n-f|>\epsilon\}\cap\{|f|\ge C\}} + \int_{\{|f_n-f|>\epsilon\}\cap\{|f|\le C\}} \right) |f| d\mu \\
&\le \epsilon + \epsilon + 2\sup_n \int_{\{|f_n - f|>\epsilon\}} C d\mu \\
&\le 2\epsilon + 2C\mu\{|f_n - f| > \epsilon\}) \le 3\epsilon.
\end{aligned}$$

Hence $\int |f_n - f| d\mu \to 0$. Finally $|\int f_n d\mu - \int f d\mu| \le \int |f_n - f| d\mu \to 0$.

Proof of b). Note that

$$\int_A |f_n| d\mu \leq \int_\Omega |f_n - f| d\mu + \int_A |f| d\mu.$$

Since μ is finite, by Lemma 1.11.3, $\{f_n\}$ is uniformly absolutely continuous. Also f_n is L^1 bounded. The uniform integrability of $\{f_n\}$ follows.

□

1.12 Functions

1.12.1 Functions of Bounded Variation

Let $\Delta : a = t_0 < x_1 < \cdots < t_n = b$ be a partition of $[a, b]$. Define $\Delta = \max_{1 \leq i \leq n}(t_i - t_{i-1})$, called the mesh of Δ. Denote by $P([a, b])$ the set of all partitions on $[a, b]$.

For $g : [a, b] \to \mathbf{R}$, define $g^\Delta = \sum_{j=1}^n \{|g(t_j) - g(t_{j-1})|\}$.

Definition 1.12.1 *The total variation of a function g on $[a, b]$ is defined by*

$$g_{TV}([a, b]) = \sup_{\Delta \in P([a,b])} g^\Delta.$$

Definition 1.12.2 *1. If $g_{TV}([a, b])$ is finite we say g is of bounded variation on $[a, b]$.*

2. Define $g_{TV}(t) := g_{TV}([a, t])$, the total variation function associated to g.

3. Consider a function $g : [a, \infty] \to \mathbf{R}$. We say g is a function of bounded variation on $[0, \infty)$ if $\lim_{t \to \infty} g_{TV}(t)$ exists and is finite.

Since stochastic processes are in general defined for positive times, we are only interested in functions defined on the positive half line.

The total variation function associated to g is right continuous if g is. Note that f could have finite total variation on each $[a, b]$ without having a finite total variation on $\mathbf{R}$ (e.g. $f(x) = \sin x$). The function $x \sin(\frac{1}{x})$ does not have bounded total variation over any interval containing 0.

Exercise 1.12.1 1. Let $c = (a, b)$ and $\Delta : a = t_0 < x_1 < \cdots < t_n = b$ a partition. Construct a new partition by adding the point c in: $\Delta' = \Delta \cup \{c\}$. Show that $g^{\Delta'} \geq g^\Delta$. Show that for $c \in (a, b)$,

$$g_{TV}([a, b]) = g_{TV}([a, c]) + g_{TV}([c, b]).$$

2. Show that $g_{TV}(t)$ is an increasing function.

3. If $g : [a, b] \to \mathbf{R}$ has finite total variation. Define $\tilde{g} : [a, \infty) \to \mathbf{R}$ by

$$\tilde{g}(x) = \begin{cases} g(x) & \text{if } x \in [a, b] \\ g(b) & \text{if } x > b \end{cases}$$

Show that $\lim_{t\to\infty} \tilde{g}_{TV}(t)$ is finite.

Exercise 1.12.2 1. If $f : [0, \infty) \to \mathbf{R}$ is an increasing function,

$$f_{TV}(x) = f(x) - f(0).$$

2. If f is locally Lipschitz continuous, then f is of bounded variation on any finite time interval.

3. Let g be a Lebesgue integrable function on $[a, b]$ define $f(x) = \int_0^x g(t)dt$. Then f is of bounded variation.

4. The set of bounded variation functions on $[a, b]$ forms a vector space.

Proposition 1.12.3 *A function $g : [a, b) \to \mathbf{R}$ of bounded variation is the difference of two bounded increasing functions.*

This is called the Jordan Decomposition of g:

$$g(x) = \frac{1}{2}[g_{TV}(x) + g(x)] - \frac{1}{2}[g_{TV}(x) - g(x)].$$

Exercise 1.12.3 Show that $g_{TV}(x) + g(x)$ is an increasing function.

1.12.2 Monotone Functions and Regularity

Proposition 1.12.4 *If $f : \mathbf{R} \to \mathbf{R}$ is an increasing function, it has left and right limits at every point. It has a countable number of points of discontinuity. It has derivative at almost every point.*

We usually take a right continuous version of an increasing function:

$$\tilde{f}(x) = \lim_{y\to x+} f(x).$$

Left and right limits follow from the Monotone convergence Theorem. Th countability comes from the following consideration. We only need to consider f an interval $[a, b]$ of length 1. Let x_i be points of discontinuity in $[a, b]$, then

$$\sum_i f(x_i+) - f(x_i-) \le f(b) - f(a).$$

There are only a finite number of x_i with jumps $f(x_i+) - f(x_i-)$ of size greater or equal to $\frac{1}{m}$. The union of all such points with $m\ inN$ contains all points of discontinuity.

1.12.3 Regularity of BV Functions

Proposition 1.12.5 *Suppose that $g : [a, b] \to \mathbf{R}$ is of bounded variation. Then*

- *g has left and right limits at every point;*
- *g has a countable number of points of discontinuity;*
- *g is differentiable almost surely everywhere;*
- *In the case of g is defined on $[0, \infty)$, $\lim_{x\to+\infty} g_{TV}(x)$ exist.*
- *In the case that $g : \mathbf{R} \to \mathbf{R}$, $\lim_{x\to-\infty} g_{TV}((-\infty, x))$ exists.*

Hence we often choose a right continuous version of such a function which is also differentiable almost surely everywhere.

1.12.4 Measures Induced by and Integration w.r.t. BV Functions

A right continuous increasing function $g : [0, \infty) \to \mathbf{R}$ defines a Borel measure on $[0, \infty)$, which we denote by μ_g. It is determined by, for $g \geq a \geq 0$, $\mu_g((a, b]) = g(b) - g(a)$. This measure μ_g is called the Lebesgue-Stieltjes measure corresponding to g, see Example 1.3.8. We may assume that $g(0) = 0$. Similarly if $g : [a, b] \to \mathbf{R}$ is of bounded variation and right continuous it induces a measure on $[a, b]$. It is customary to denote the integral with respect to this measure by

$$\int_0^t f dg := \int_0^t f d\mu_g.$$

If g is a right continuous functions of bounded variation instead of an increasing function, a signed measure is defined via the Jordan decomposition of BV functions as the difference of two increasing functions. For relation between integrals with respect to Lebesgue-Stieltjes measures and Riemann-Stieltjes integrals see section 1.7.

Exercise 1.12.4 Let $g : [0, 1]$ be of bounded variation. Suppose that g is not right continuous but left continuous. Could one define a measure μ? If $A_1 \subset A_2 \subset \dots$, does it hold that $\lim_{n\to\infty} \mu(A_n) = \mu(\cup_{n=1}^{\infty} A_n)$?
Hint: Consider how to give a reasonable value to $\mu([c, b])$.

Proposition 1.12.6 *The measure μ_g is singular with respect to the Lebesgue measure iff $g' = 0$ almost surely. The measure μ_g is absolutely continuous with respect to the Lebesque measure iff $g(x) = \int_0^x g'(t)dt$.*

In the absolutely continuous case $\mu_g(A) = \int_A g'(t)dt$. If μ is the Lebesgue measure on the line, $\frac{d\mu_g}{d\mu}(t) = g'(t)$. Hence $\int f dg = \int f(t)g'(t)dt$.

Example 1.12.7 Take $C(x)$ to be the Cantor function (see Appendix **??**). It is an increasing function and constant on each interval of the complement of the cantor set. And $C' = 0$ a.e.. It defines a measure that is singular with respect to the Lebesgue measure.

Exercise 1.12.5 Show that if $f : [a, b] \to \mathbf{R}$ is positive Borel measurable and g an increasing function then $\int f_s dg_s \geq 0$ is of bounded variation.

1.12.5 Absolutely Continuous and BV Functions

Definition 1.12.8 *A function $f : \mathbf{R} \to \mathbf{R}$ is absolutely continuous if for any $\epsilon > 0$ there is $\delta > 0$ such that for any finite number of disjoint intervals $(a_1, b_1), \ldots, (a_n, b_n)$ with $\sum_{j=1}^n (b_j - a_j) < \delta$ we have*

$$\sum_{j=1}^n |f(b_j) - f(a_j)| < \epsilon.$$

Absolutely continuous functions are uniformly continuous (by taking $n = 1$). They are also of bounded variation on any finite interval $[a, b]$. Take $\epsilon = 1$, there is $\delta > 0$ such that

$$\sum_{j=1}^n |f(b_j) - f(a_j)| < 1,$$

if $\sum_{j=1}^n (b_j - a_j) < \delta$. Now we divide let $k = 2\frac{b-a}{\delta}$ be an integer. Then for each partition with Δ sufficiently small, we have

$$\sum_{j=1}^n \left|f(t_{j+1}) - f(a_{t_j})\right| \leq k.$$

Proposition 1.12.9

- *If $g : [0, \infty) \to \mathbf{R}$ is Lebesgue integrable define $f(x) = \int_0^x g(t)dt$. Then f is of bounded variation and is absolutely continuous. Furthermore $f' = g$ almost surely.*
- *If $g : [0, \infty) \to \mathbf{R}$ is of bounded variation, $g(0) = 0$ and g is absolutely continuous then g' is Lebesque integrable and $g(x) = \int_0^x g'(t)dt$.*