Appendix A

Review of Background Material

A.1 Basic Measure Theory

Modern probability theory is spoken in the language of measure theory ... so before we can learn of the advances in probability theory, we should familiarize ourselves with the basic jargon of measure theory. Throughout let Ω be a set that is referred to as the *sample space*. Very roughly speaking, $\omega \in \Omega$ represents an experiment or a "realization" of the randomness. We will also refer to another space S as the state space, which loosely corresponds to the space of all possible values that can be attained in the experiment.

1. Algebras and σ-algebras. $\mathcal{F}$ is a *σ-algebra* (or, equivalently, a *σ-field*) if it is a non-empty collection of subsets of Ω that is closed under countable unions (i.e. if $A_i \in \mathcal{F}, i = 1, \ldots,$ then $\cup_i A_i \in \mathcal{F}$ and $\cap_i A_i \in \mathcal{F}$), as well as closed under complementation (i.e. if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$). It is an algebra if, instead of being closed under countable unions, it is closed under only finite unions.

Comments. The following are easy consequences of the definition.
(1) $\Omega \in \mathcal{F}$ and $\emptyset \in \mathcal{F}$.
(2) $\mathcal{F}$ is closed under countable intersections and under monotone limits.
(3) The largest $\mathcal{F}$ is the power set (the set of all subsets) of Ω and the smallest $\mathcal{F}$ is $\{\emptyset, \Omega\}$.
(4) Intersections of σ-fields are σ-fields.
(5) Given any set $\mathcal{C}$ of subsets of Ω, the $\mathcal{F}$ generated by $\mathcal{C}$ is well-defined and denoted by $\sigma(\mathcal{C})$ (simply take the intersection of all σ-fields containing $\mathcal{C}$.

If S is a topological space, then the σ-algebra generated by the open sets is referred to as the *Borel σ-algebra* of S and denoted $\mathcal{B}(S)$.

Think. Why do we need σ-fields ? What do they represent ?
These are particularly important questions in probability theory where, in constrast to measure theory, measurability (and therefore the concept of a σ-field) tends to be as central an issue as measure.

2. Measurable space.
The collection $(\Omega, \mathcal{F})$ of a space with a σ-algebra is referred to as a *measurable space.* Elements of $\mathcal{F}$ are referred to as *measurable sets.*

3. Measures.
Definitions. A non-negative set function $\mu_0 : \mathcal{F}_0 \to [0, \infty]$ on an algebra $\mathcal{F}_0$ in Ω is said to be *additive* (or *finitely additive*) if $\mu_0(\emptyset) = 0$ and for $F, G \in \mathcal{F}_0$

$$F \cap G = \emptyset \Rightarrow \mu_0(F \cup G) = \mu_0(F) + \mu_0(G).$$

The map μ_0 is said to be *countably additive* on $\mathcal{F}_0$ if, whenever $\{F_n, n \in \mathbb{N}\}$ is a sequence of disjoint sets in $\mathcal{F}_0$ with union $F = \cup_n F_n$ in $\mathcal{F}_0$, then

$$\mu_0(F) = \sum_n \mu_0(F_n).$$

Given a measurable space $(\Omega, \mathcal{F})$, a map $\mu : \mathcal{F} \to [0, \infty]$ is said to be *a measure* if it is countably additive on $\mathcal{F}$.
The triple $(\Omega, \mathcal{F}, \mu)$ is then referred to as *a measure space.* A non-negative measure μ is said to be a probability measure if $\mu(\Omega) = 1$.
A statement C about points $\omega \in \Omega$ is said to hold μ-*almost surely* if the set

$$F \doteq \{\omega : C(\omega) \text{ is false } \} \in \mathcal{F} \quad \text{and} \quad \mu(F) = 0.$$

Given a measure space $(\Omega, \mathcal{F}, \mu)$, for any set $G \subset \Omega$ the *inner* μ-*measure* $\mu_*(G)$ and the *outer* μ-*measure* $\mu^*(G)$ is defined by

$$\mu_*(G) = \sup\{\mu(F) : F \in \mathcal{F}, F \subset G\},$$

and

$$\mu^*(G) = \inf\{\mu(F) : F \in \mathcal{F}, G \subset F\}.$$

respectively. If $\mu_*(G) = \mu^*(G)$ then G is said to be μ-*measurable.* If $\mathcal{F}^\mu$ denotes the collection of μ-measurable sets, then μ can be extended to a measure on $\mathcal{F}^\mu$ by setting

$$\mu(G) = \mu_*(G) = \mu^*(G) \text{ for } G \in \mathcal{F}^\mu.$$

The triple $(\Omega, \mathcal{F}^\mu, \mu)$ is called *the completion* of $(\Omega, \mathcal{F}, \mu)$. The σ-algebra $\mathcal{F}^\mu$ is the smallest σ-algebra that extends $\mathcal{F}$ and contains every set of outer μ-measure 0.

Standard Results.

Theorem A.1.1. Carathéodory's Extension Theorem. *Given a set Ω, an algebra $\mathcal{F}_0$ on Ω and a finitely additive set function μ_0 on $\mathcal{F}_0$, there exists a unique measure μ on $(\Omega, \sigma(\mathcal{F}_0))$ such that $\mu = \mu_0$ on $\mathcal{F}_0$ if and only if μ_0 is countably additive on $\mathcal{F}_0$.*

Lemma A.1.2. *Consider a probability space $(\Omega, \mathcal{F}, \mu)$ and $G \subset \Omega$ with $\mu^*(G) = 1$. Then $\mu^*(G \cap F) = \mu(F)$ for every $F \in \mathcal{F}$. Moreover, $(G, \mathcal{G}, \mu^*)$ is a measure space, where $\mathcal{G}$ is the σ-algebra comprising subsets of G of the form $G \cap F$, for some $F \in \mathcal{F}$.*

3. Measurable mappings, random variables and distributions.
Given two measurable spaces $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$, a mapping $\xi : \Omega \to S$ is said to be $(\mathcal{F}, \mathcal{S})$-measurable if for every $A \in \mathcal{S}$, the set $\xi^{-1}(A) \doteq \{w : \xi(w) \in A\} \in \mathcal{F}$. (When the σ-fields $\mathcal{F}$ and $\mathcal{S}$ are clear from the context, then one sometimes simply says that ξ is measurable.) If S is a topological space and $\mathcal{S}$ is the *Borel σ-field*, then the mapping ξ is said to be *Borel-measurable.* If $S = \mathbb{R}^d$ then the mapping ξ is often referred to as a *random variable* or *random vector* (when $d > 1$) and, more generally, it is referred to as a *random element* or an S-valued random variable. Given a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$, the *distribution* μ of ξ is the *image measure* on $(S, \mathcal{S})$ induced by ξ:

$$\mu(A) = \mathbb{P}\left(\{\omega \in \Omega : \xi(\omega) \in A\}\right), \quad A \in \mathcal{S}.$$

4. L^p spaces, Modes of Convergence and Uniform Integrability

Definition A.1.3. *(L^p spaces) Given a measure space $(\Omega, \mathcal{F}, \mu)$, two measurable functions X and Y are said to be in the same equivalence class if $\mu(X \neq Y) = 0$. If X and Y belong to the same equivalence class, one sometimes says X is a version of Y. For $p > 0$,* the space *$L^p(\Omega, \mathcal{F}, \mu)$ is the collection of equivalence classes of measurable functions that satisfy*

$$\int_\Omega |X|^p \, d\mu < \infty.$$

When the probability space, filtration and/or the probability measure is clear from the context, we will sometimes just write L^p or $L^p(\mu)$ or $L^p(\mathcal{F}, \mu)$ for $L^p(\Omega, \mathcal{F}, \mu)$.

Definition A.1.4. *(Expectation) Given a random variable X defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the* expectation *$\mathbb{E}[X]$ of the random variable is defined to be*

$$\mathbb{E}[X] = \int_\Omega X(\omega) dP(\omega).$$

If $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, then the expectation of X is well-defined and finite.

One is often interested in characterizing the convergence of random variables or random elements. For example, the strong law of large numbers, weak law of large numbers and the central limit theorem all identify the limits of appropriately scaled sequences of random variables/vectors. We review these basic modes of convergence.

Definition A.1.5. (Modes of Convergence) *Given a random element X and a sequence of random elements $\{X_n\}_{n\in\mathbb{N}}$ defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values on a common topological space $(S, \mathcal{B}(S))$,*

1. *X_n is said to converge to X* almost surely *iff*
$$\mathbb{P}\left(\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right) = 1.$$

2. *X_n is said to converge to X* in L^p *(denoted $X_n \xrightarrow{L^p} X$) iff X and each X^n lie in L^p and*
$$\lim_{n\to\infty} \int_\Omega |X_n(\omega) - X(\omega)|^p dP(\omega) = 0.$$

3. *If S is a separable metric space with distance ρ, X_n is said to converge to X* in probability *(denoted $X_n \xrightarrow{\mathbb{P}} X$) iff for every $\varepsilon > 0$,*
$$\lim_{n\to\infty} \mathbb{P}(\omega \in \Omega : \rho(X_n(\omega), X(\omega)) \geq \varepsilon) = 0.$$

In the definition of convergence in probability above, separability of the state space is required to ensure that $\rho(X_n, X)$ is a random variable for every n. It is easy to see that a.s. convergence and L^p convergence imply convergence in probability, and it is possible to show (using the Borel-Cantelli lemma) that convergence in probability implies that every subsequence has a further subsequence, along which we have the a.s. convergence of X_n to X. However, one is often interested in when a.s. convergence implies L^1 convergence, i.e., when does $X_n \to X$ $\mathbb{P}$ a.s. imply
$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$
In other words, when can we interchange limits and expectation?

We start by recalling some classical theorems from measure theory that address this question.

Theorem A.1.6. (Convergence Theorems) *$X_n \to X$ a.s. implies $X_n \xrightarrow{L^1} X$, i.e. $\lim_{n\to\infty} \mathbb{E}[|X_n - X|] \to 0$ in the following cases:*

1. **(Dominated Convergence Theorem - DCT)** *If $X_n(\omega) \leq Y(\omega)$ for every $n \in \mathbb{N}, \omega \in \Omega$ and $\mathbb{E}[|Y|] < \infty$.*

2. **(Bounded Convergence Theorem - BCT)** *There exists $C < \infty$ such that $\sup_n |X_n(\omega)| \leq C$ for every $\omega \in \Omega$.*

3. *(***Monotone Convergence Theorem - MCT***) If $\{X_n\}$ is a non-negative non-decreasing sequence of random variables.*

Moreover, **Fatou's lemma** *(abbreviated, Fatou) states that if $\{X_n\}$ is a non-negative non-decreasing sequence of random variables, then*
$$\mathbb{E}\left[\liminf_n X_n\right] \leq \liminf_n \mathbb{E}\left[X_n\right].$$

The DCT, MCT and Fatou's lemma are all valid for more general (i.e. not necessarily probability or finite) measure spaces. Note that the BCT is essentially a consequence of the DCT, but we state it separately for the purpose of applications. While all these conditions are very useful, the most general condition under which the above question is answered in the affirmative is "uniform integrability".

Definition A.1.7. (Uniform Integrability) A family of random variables $\{X_t\}_{t\in T}$, is said to be *uniformly integrable* if

$$\lim_{r\to\infty}\sup_{t\in T}\mathbb{E}[|X_t|\mathbb{I}_{\{|X_t|>r\}}]=0.$$

The importance of uniform integrability derives from the following theorem.

Theorem A.1.8. (A necessary and sufficient condition for L^1 convergence) *Let $\{X_n\}$ be a sequence in L^1 and let $X\in L^1$. Then $X_n \xrightarrow{L^1} X$ iff the following two conditions are satisfied:*

1. *$X_n \to X$ in probability;*
2. *the sequence $\{X_n\}$ is uniformly integrable.*

It is the "if" part of the last theorem that is particularly useful. We now provide some useful sufficient conditions for uniform integrablility.

Lemma A.1.9. (Characterizations of Uniform Integrability)

1. *The class of random variables $\{X_t\}_{t\in T}$ is uniformly integrable if and only if $\sup_t \mathbb{E}[|X_t|]<\infty$ (i.e., $\{X_t\}_{t\in T}$ is L^1 bounded) and*

$$\lim_{\mathbb{P}(A)\to 0}\sup_{t\in T}\mathbb{E}[|X_t|\mathbb{I}_A]=0. \tag{A.1}$$

2. *The class of random variables $\{X_t\}_{t\in T}$ is uniformly integrable if and only if there exists a nonnegative increasing function G on $[0,\infty)$ such that*

$$\lim_{t\to\infty}\frac{G(t)}{t}=\infty$$

and

$$\sup_t \mathbb{E}[G(X_t)]<\infty.$$

In particular, if $\{X_t\}_{t\in T}$, is L^p bounded for some $p>1$, i.e., $\sup_{t\in T}\mathbb{E}[|X_t|^p]<\infty$, then $\{X_t\}_{t\in T}$ is uniformly integrable.

Proof. Suppose $\{X_t\}_{t\in T}$ is uniformly integrable. Then for $r>0$,

$$\mathbb{E}[|X_t|\mathbb{I}_A]\le r\mathbb{P}(A)+\mathbb{E}[|X_t|\mathbb{I}_{\{X_t>r\}}].$$

Taking $\mathbb{P}(A)\to 0$ and then $r\to\infty$, (A.1) follows, while taking $A=\Omega$ and r large enough, the L^1-boundedness follows.

Conversely, suppose $\{X_t\}_{t\in T}$ is L^1-bounded and satisfies (A.1). Then, by Markov's inequality, it follows that

$$\lim_{r\to\infty}\sup_t \mathbb{P}(|X_t|>r) \le \lim_{r\to\infty}\sup_t \frac{\mathbb{E}[|X_t|]}{r} = 0.$$

Thus, applying (A.1) with $A = \{|X_t| > r\}$, it follows that the family $\{X_t\}_{t\in T}$ is uniformly integrable, and the first assertion of the theorem follows.

We omit the proof of the second assertion, but observe that the third assertion follows from the second by choosing $G(t) = t^p$ for $p > 1$. □

From Theorem A.1.8, it follows that uniform integrability, along with convergence in probability, implies L^1 convergence. In the absence of convergence in probability, uniform integrability implies weak L^1 convergence. It is well-known that any L^2-bounded sequence has a subsequence that converges weakly in L^2. Uniform integrability provides a corresponding sufficient condition for (relative) weak compactness in L^1. This fact is used, for example, in the proof of the Doob-Meyer decomposition in continuous time.

Definition A.1.10. *Assume* $X, X_1, X_2, \ldots \in L^p$ *for some* $p \in [1,\infty)$. *Then* $X_n \to X$ weakly *in* L^p *iff* $\mathbb{E}[X_n\eta] \to \mathbb{E}[X\eta]$ *for every* $\eta \in L^q$, *where* $1/p + 1/q = 1$.

Lemma A.1.11. (weak L^1-compactness, Dunford) *Every uniformly integrable sequence of random variables has a subsequence that converges weakly in* L^1.

Proof. Let $\{X_n\}$ be uniformly integrable. Define $X_n^k \doteq X_n \mathbb{I}_{\{|X_n|\le k\}}$, and note that $\{X_n^k\}$ is L^2-bounded in n for each k. By L^2-compactness and a diagonalization argument, there exists a subsequence $N' \subset \mathbb{N}$ and some random variables $\eta_1, \eta_2 \ldots$, such that for every k, $X_n^k \to \eta_k$ holds weakly in L^2, and thus also in L^1, as $n \to \infty$ along N'.

Now, by Fatou's lemma, $||\eta_k - \eta_l||_1 \le \liminf_n ||X_n^k - X_n^l||_1$, and by uniform integrability the right-hand side tends to zero as $k, l \to \infty$. Thus, the sequence $\{\eta_k\}_{k\in\mathbb{N}}$ is Cauchy in L^1 and so, it converges in L^1 towards some ξ. By approximation, it follows easily that $X_n \to X$ weakly in L^1 along N'. Indeed, for any bounded, $\mathcal{F}$-measurable random variable γ, note that for any $k, n \in \mathbb{N}$,

$$\mathbb{E}[\gamma(X_n - X)] \le ||\gamma||_\infty \sup_n ||X_n - X_n^k||_1 + \mathbb{E}[\gamma(X_n^k - \eta_k)] + ||\gamma||_\infty ||\eta_k - X||_1.$$

Taking limsup of both sides as $n \to \infty$, the second term on the r.h.s. vanishes. Then taking $k \to \infty$, the first term vanishes by the uniform integrability of $\{X_n\}$, while the last term vanishes by the definition of X. □

In addition to the modes of convergence described in Definition A.1.5, in the next section we will also consider weaker forms of convergence (for sequences of random variables not necessarily defined all on the same probability space) such as convergence in distribution.

A.2 Limit Theorems

In what follows, $\{X_i\}_{i=1}^{\infty}$ is a sequence of i.i.d. random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and, for $n \in \mathbb{N}$,

$$S_n \doteq X_1 + X_2 + \cdots + X_n.$$

Theorem A.2.1. *(Khintchine's Weak Law of Large Numbers) If* $\mathbb{E}[|X_1|] < \infty$*, then as* $n \to \infty$*,*

$$\frac{S_n}{n} \to \mathbb{E}[X_1] \qquad \text{in } \mathcal{L}^1(\mathbb{P}) \text{ and hence in probability.}$$

Theorem A.2.2. *(Kolmogorov's Strong Law of Large Numbers) If* $\mathbb{E}[|X_1|] < \infty$*, then as* $n \to \infty$*,*

$$\lim_{n\to\infty} \frac{S_n}{n} = \mathbb{E}[X_1] \qquad \mathbb{P} \text{ a.s.} \tag{A.2}$$

Conversely, if $\limsup_{n\to\infty} |S_n/n| < \infty$ *with positive probability, then* $\mathbb{E}[|X_1|] < \infty$ *and (A.2) holds.*

We now recall the celebrated central limit theorem. First, recall the notion of convergence in distribution of random variables or probability measures on $\mathbb{R}^d$.

Definition A.2.3. *If* μ*,* $\mu_i, i \in \mathbb{N}$ *are probability distributions on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *with respective distribution functions* F*,* $F_i, i \in \mathbb{N}$*. Then* μ_n *converges in distribution to* μ *(denoted* $\mu_n \Rightarrow \mu$*) as* $n \to \infty$ *if and only if*

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

for all $x \in \mathbb{R}^d$ *at which* F *is continuous. Moreover, given* $\mathbb{R}^d$*-valued random variables* X*,* $X_i, i \in \mathbb{N}$*,* X_n *is said to converge in distribution to* X *as* $n \to \infty$ *if and only if their distributions converge weakly or, equivalently, if and only if*

$$\mathbb{P}(X_n \leq x) \Rightarrow \mathbb{P}(X \leq x)$$

for all x *such that* $\mathbb{P}(X = x) = 0$*, where here the inequalities are to be interpreted coordinatewise.*

We now provide a useful criteria for convergence in distribution on $\mathbb{R}^d$.

Definition A.2.4. *Given a finite measure* μ *on* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$*, its Fourier transform* $\widehat{\mu}$ *is given by*

$$\widehat{\mu}(t) = \int_{\mathbb{R}} e^{i\langle t,x\rangle} \mu(dx), \qquad t \in \mathbb{R}^d,$$

where $i = \sqrt{-1}$ *and* $\langle t, x\rangle = \sum_{i=1}^d t_i x_i$ *is the usual Euclidean inner product. If* X *is an* $\mathbb{R}^d$*-valued random variable/vector with distribution* μ*,* $\hat{\mu}$ *is often referred to as the characteristic function of* X*.*

Lemma A.2.5. *(Cramér-Wold Device) Given $\mathbb{R}^d$-valued random variables $X, X_n, n \in \mathbb{N}$ with respective distributions $\mu, \mu_n, n \in \mathbb{N}$, $X_n \Rightarrow X$ if and only if $\langle t, X_n\rangle \Rightarrow \langle t, X\rangle$ for all $t \in \mathbb{R}^d$. Moreover, the latter convergence holds if $\widehat{\mu_n}(t) \to \widehat{\mu}(t)$ for all $t \in \mathbb{R}^d$.*

Theorem A.2.6. *(1-d Central Limit Theorem) If $\mathbb{E}[|X_1|] < \infty$ and $Var(X_1) \in (0, \infty)$, then*

$$\frac{S_n - n}{\sqrt{n}} \Rightarrow \mathcal{N}(0, Var(X_1)).$$

Theorem A.2.7. *(Multidimensional Central Limit Theorem) If $\{X^n\}_{n\in\mathbb{N}}$ are i.i.d. random vectors in $\mathbb{R}^d$ with $\mathbb{E}[X_i^1] = \mu_i$ and $Cov(X_i^1 X_j^1) = Q_{i,j}$ for an invertible $(d \times d)$ matrix Q. Then $(S_n - n\mu)/\sqrt{n}$ converges in distribution to a multi-dimensional Gaussian with mean zero and covariance matrix Q, i.e., for all d-dimensional hypercubes $G = (a_1, b_1] \times \cdots \times (a_d, b_d]$,*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - n\mu}{\sqrt{n}} \in G\right) = \int_G \frac{e^{-\frac{1}{2}y'Q^{-1}y}}{(2\pi)^{d/2}\sqrt{\det(Q)}}\, dy.$$

A.3 Conditional Expectations.

Definition A.3.1. *Let $(\Omega, \mathcal{F}, P)$ be a measure space and let X be an $\mathcal{F}$-measurable random variable with $\mathbb{E}[|X|] < \infty$. Let $\mathcal{G}$ be a sub σ-algebra of $\mathcal{F}$. Then Y is* a version of the conditional expectation $\mathbb{E}[X|\mathcal{G}]$ *if and only if Y is a random variable that satisfies*

(i) Y is $\mathcal{G}$-measurable.

(ii) $\mathbb{E}[|Y|] < \infty$.

(iii) For every set $G \in \mathcal{G}$

$$\int_G Y dP = \int_G X dP.$$

Given another random variable Z on $(\Omega, \mathcal{F}, P)$, a version of the conditional expectation $\mathbb{E}[X|Z]$ *is defined to be* $\mathbb{E}[X|\sigma(Z)]$.

When $X \in L^1(P)$ (that, is, when $\mathbb{E}[|X|] < \infty$), existence of conditional expectations follow from the Radon-Nikodým theorem. When $X \in L^2(P)$ the alternative characterization of the conditional expectation $E[X|\mathcal{G}]$ as the orthogonal projection of the $\mathcal{F}$-measurable function X onto the sub-space of $\mathcal{G}$-measurable functions in the Hilbert space $\mathcal{L}^2(P)$ is also useful for an intuitive interpretation of $E[X|\mathcal{G}]$ as the "best guess" or "best prediction" of X given the information contained in $\mathcal{G}$.

Properties of Conditional Expectations

We now provide a list of basic properties satisfied by conditional expectations. Here, we always assume that X is a random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that satisfies $\mathbb{E}[|X|] < \infty$, and that $\mathcal{G}$ and $\mathcal{H}$ denote sub-σ-algebras of $\mathcal{F}$. (The use of 'c' to denote 'conditional' in (cMON), etc., should be apparent.)

(a) If Y is any version of $\mathbb{E}[X|\mathcal{G}]$ then $\mathbb{E}[Y] = \mathbb{E}[X]$.

(b) If X is $\mathcal{G}$ measurable, then $\mathbb{E}[X|\mathcal{G}] = X$, a.s.

(c) (**Linearity**) $\mathbb{E}[a_1X_1 + a_2X_2|\mathcal{G}] = a_1\mathbb{E}[X_1|\mathcal{G}] + a_2\mathbb{E}[X_2|\mathcal{G}]$, a.s.

Clarification: if Y_1 is a version of $\mathbb{E}[X_1|\mathcal{G}]$ and Y_2 is a version of $\mathbb{E}[X_2|\mathcal{G}]$, then $a_1Y_1 + a_2Y_2$ is a version of $\mathbb{E}[a_1X_1 + a_2X_2|\mathcal{G}]$.

(d) (**Positivity**) If $X \geq 0$, then $\mathbb{E}[X|\mathcal{G}] \geq 0$, a.s.

(e) (**cMON**) If $0 \leq X_n \uparrow X$, then $\mathbb{E}[X_n|\mathcal{G}] \uparrow \mathbb{E}[X|\mathcal{G}]$, a.s.

(f) (**cFATOU**) If $X_n \geq 0$, then $\mathbb{E}\left[\liminf X_n|\mathcal{G}\right] \leq \liminf \mathbb{E}\left[X_n|\mathcal{G}\right]$, a.s.

(g) (**cDOM**) If $|X_n(\omega)| \leq V(\omega), \forall n$ and $\mathbb{E}[|V|] < \infty$ and $X_n \to X$, a.s., then

$$\mathbb{E}[X_n|\mathcal{G}] \to \mathbb{E}[X|\mathcal{G}], \text{ a.s.}$$

(h) (**cJENSEN**) If $c : \mathbb{R} \to \mathbb{R}$ is convex, and $\mathbb{E}\left[c(X)\right] < \infty$, then

$$\mathbb{E}\left[c(X)|\mathcal{G}\right] \geq c\left(\mathbb{E}\left[X|\mathcal{G}\right]\right) \geq c\left(\mathbb{E}\left[X|\mathcal{G}\right]\right), \quad \text{a.s.}$$

Important corollary: $\|\mathbb{E}[X|\mathcal{G}]\|_p \leq \|X\|_p$ for $p \geq 1$, where $||Y||_p$ denotes $(\mathbb{E}[|Y|^p])^{1/p}$ for any r.v. Y.

(i) (**Tower Property**) If $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$, then

$$\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{G}\right]|\mathcal{H}\right] = \mathbb{E}\left[X|\mathcal{H}\right], \quad \text{a.s.}$$

(j) (**'Taking out what is known'**) If Z is $\mathcal{G}$-measurable and bounded, then

$$(*) \qquad \mathbb{E}\left[ZX|\mathcal{G}\right] = Z\mathbb{E}\left[X|\mathcal{G}\right], \quad \text{a.s.}$$

If $p > 1,\ p^{-1} + q^{-1} = 1,\ X \in L^p\left(\Omega, \mathcal{F}, \mathbb{P}\right)$ and $Z \in L^q(\Omega, \mathcal{G}, \mathbb{P})$, then (*) again holds. If $X \in (m\mathcal{F})^+,\ Z \in (m\mathcal{G})^+,\ \mathbb{E}[X] < \infty$ and $\mathbb{E}[ZX] < \infty$, then (*) holds.

(k) (**Rôle of independence**) If $\mathcal{H}$ is independent of $\sigma(\sigma(X), \mathcal{G})$, then

$$\mathbb{E}\left[X|\sigma(\mathcal{G},\ \mathcal{H})\right] = \mathbb{E}[X|\mathcal{G}], \quad \text{a.s.}$$

In particular, if X is independent of $\mathcal{H}$, then $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[X]$, a.s.

A.4 Lebesgue-Stieltjes Integration

In this section we provide a brief summary of Lebesgue-Stieltjes integration. Given a real-valued, right-continuous function f defined on $[0,\infty)$, the (first) variation $V_t(f) = V_t^{(1)}(f)$ of the function f on the interval $[0,t]$ is defined to be

$$V_t(f) \doteq \sup_{\pi} \sum_{i=1}^{k_\pi} |f_{t_{i+1}} - f_{t_i}|, \tag{A.3}$$

where the supremum is over partitions $\pi = \{0 = t_0 < t_1 < \ldots < t_{k_\pi} = t\}$ of $[0,t]$. Since by the triangle inequality, for $0 < s < r < t$,

$$|f_t - f_s| \leq |f_t - f_r| + |f_r - f_s|,$$

it follows that

$$V_t(f) = \lim_{||\pi|| \downarrow 0} \sum_{i=1}^{k_\pi} |f_{t_{i+1}} - f_{t_i}|,$$

where the (mesh) size $||\pi||_t$ of the partition is equal to $\max_i |t_{i+1} - t_i|$. In particular we can write

$$V_t(f) \doteq \sup_{n \geq 1} \sum_{k=1}^{2^n} \left| f_{\frac{tk}{2^n}} - f_{\frac{t(k-1)}{2^n}} \right|.$$

A function f is said to be of *finite variation* if $V_t(f) < \infty$ for every $t \in [0,\infty)$ and then the function $V(f) : t \to V_t(f)$ is referred to as the *total variation* of f. The total variation of f is clearly positive and non-decreasing. If $\lim_{t\to\infty} V_t(f) < \infty$ then the function f is said to be of *bounded variation.*

It is easy to see that C^1 (continuously differentiable) functions are of finite variation. Monotone functions are also of finite variation, and conversely any real-valued function f of finite variation is the difference of two non-decreasing functions. Indeed, note that

$$f_t = f_0 + f_t^+ - f_t^-$$

with $f_0^+ = f_0^- = 0$, where

$$f_t^+ \doteq [V_t(f) + f_t - f_0]/2 \quad \text{and } f_t^- \doteq [V_t(f) - f_t + f_0]/2$$

can be easily shown to be non-negative, non-decreasing functions. This decomposition is minimal in the sense that any other pair of positive non-decreasing functions $\tilde{f}^-$ and $\tilde{f}^+$ with the property that $f - f_0 = \tilde{f}^+ - \tilde{f}^-$, must satisfy $f^+ \leq \tilde{f}^+$ and $f^- \leq \tilde{f}^-$. As a consequence f has left limits at any $t \in (0,\infty)$ and we write f_{t-} or $f(t-)$ for $\lim_{s\uparrow t} f_s$ and set $f_{0-} = f_0$. Moreover, we use

$$\Delta f \doteq f_t - f_{t-}$$

to denote the *jump* of f at t.

The importance of functions of finite variation lies in the fact that the map $f \mapsto \mu_f$ given by

$$f_t = \mu_f([0,t]) \quad \text{and} \quad \mu_f(\{0\}) = 0.$$

defines a one-to-one correspondence between right-continuous functions f having finite variation and Radon measures μ on $[0,\infty)$ (recall that a Radom measure on $[0,\infty)$ is a measure μ on the Borel σ-field $\mathcal{B}([0,\infty))$ such that $\mu(K) < \infty$ for every compact set $K \subset [0,\infty)$). The measure $\mu_{V(f)}$ induced by the total variation function $V(f)$ corresponds to the total variation $|\mu_f|$ of the measure μ and the measures induced by f^+ and f^- correspond to the minimal decomposition of μ_f into its positive and negative parts respectively [7]. Define

$$\Delta f_t = f_t - f_{t-}$$

and note that f can be decomposed into the sum of its continuous and atomic parts as follows:

$$f = f_0 + f^c + f^d$$

where

$$f_t^c = f_{t-} \quad \text{and} \quad f^d = \sum_{0<s\leq t} \Delta f_s.$$

Since $f_{t-} = \mu_f([0,t))$ and $\Delta f_t = \mu_f(\{t\})$, this corresponds to the decomposition of the measure μ_f into its absolutely continuous and singular parts (with respect to Lebesgue measure).

Given a function f of finite variation and a locally bounded Borel-measurable function g on $[0,\infty)$ (or, more generally, a function that lies in $L^1([0,\infty),\mathcal{B}([0,\infty)),|\mu_f|)$ the *Lebesgue-Stieltjes integral* of g with respect to f is defined to be the Lebesgue integral of g with respect to μ_f:

$$(g \cdot f)_t \doteq \int_0^t g_s df_s \doteq \int_{(0,t]} g_s d\mu_f(s), \tag{A.4}$$

Note that the jump at zero does not come into play, in the sense that

$$\int_0^t df_s = f_t - f_0.$$

This representation, along with the fact that $\mu_f = \mu_{f^+} - \mu_{f^-}$, implies that the map $t \to (g \cdot f)_t$ is also right-continuous and of finite variation. Moreover, as a consequence of the Radon-Nikodym theorem, it follows that if f is an absolutely continuous function then μ_f is absolutely continuous with respect to Lebesgue measure and the integral $g \cdot f$ is a continuous function of t. If g is a continuous function on $[0,t]$, then the Riemann-Stieltjes integral of f with respect to g on $[0,t]$ is well-defined and equals the Lebesgue-Stieltjes integral, that is

$$\int_{[0,t]} g(s)df(s) = \lim_{n\to\infty} \sum_{k=1}^{N_n} g(s_k^n)(f(t_k^n) - f(t_{k-1}^n)), \tag{A.5}$$

for any sequence of partitions $\pi = \{t_k, k = 1, \ldots, N_n\}$ where $s_k^n \in [t_{k-1}^n, t_k^n]$ and $\max_{k=1}^{N_n} |t_k - t_{k-1}| \to 0$ as $n \to \infty$.

We now recall the integration by parts and change-of-variable formulas for Lebesgue-Stieltjes integrals.

Lemma A.4.1. *(Integration by Parts for FV functions) If f and g are two right-continuous functions with finite variation, then for any t*

$$f_t g_t = f_0 g_0 + \int_0^t f_s dg_s + \int_0^t g_{s-} df_s = \int_0^t f_{s-} dg_s + \int_0^t g_{s-} df_s + [f,g]_t, \quad (A.6)$$

where

$$[f,g]_t \doteq \sum_{0<s\leq t} \Delta f_s \Delta g_s.$$

Proof. Consider the product signed measure $\mu_f \times \mu_g$ on $(0,\infty) \times (0,\infty)$ and notice that both sides of the equality (A.6) are equal to $(\mu_f \otimes \mu_g)([0,t]^2)$. It is not hard to see that $\int_0^t f_s dg_s$ is the measure of the upper triangle including the diagonal, $\int_0^{t-} g_{s-} df_s$ is the measure of the lower triangle excluding the diagonal and $f_0 g_0 = (\mu_f \otimes \mu_g)(\{0,0\})$. Indeed, evaluating the signed measures using the representation

$$\begin{aligned} (0,t] \times (0,t] \quad &= \quad \{(u,v) : 0 < v \leq t; 0 < u < v\} \cup \{(u,v) : 0 < u = v \leq t\} \\ &\quad\;\; \cup \{(u,v) : 0 < u \leq t; 0 < v < u\}, \end{aligned}$$

one obtains

$$(f_t - f_0)(g_t - g_0) = \int_{(0,t]} (f_{v-} - f_0) dg_u + \int_{(0,t]} (g_{v-} - g_0) df_u + [f,g]_t.$$

The above can be rearranged to give the lemma. □

Lemma A.4.2. *Let F be a C^1 function on $[0,\infty)$ and let g be a finite variation function on $[0,\infty)$. Then*

$$F(g_t) - F(g_0) = \int_{(0,t]} F'(g_{s-}) dg_s + \sum_{0<s\leq t} \left[F(g_s) - F(g_{s-}) - F'(g_{s-})\Delta g_s\right]. \quad (A.7)$$

Proof. Fix the finite variation function g. Let $\mathcal{C}$ be the class of functions F such that (A.7) holds. Then it is clear that $\mathcal{C}$ is a vector space. Moreover, as shown below, $\mathcal{C}$ is also an algebra. Indeed, let $F, G \in \mathcal{C}$ and set $H = FG$. Then by assumption F and G satisfy (A.7) and therefore in differential notation we can write

$$d(F(g_t)) = F'(g_{t-}) dg_t + [F(g_t) - F(g_{t-}) - F'(g_{t-})\Delta g_t]$$

and likewise

$$d(G(g))_t = G'(g_{t-}) dg_t + [G(g_t) - G(g_{t-}) - G'(g_{t-})\Delta g_t].$$

On the other hand by the integration-by-parts formula

$$H(g_t) - H(g_0) = \int_{(0,t]} F(g_{s-})d(G(g))_s + \int_{(0,t]} G(g_{s-})d(F(g))_s + [F(g), G(g)]_t,$$

from which it follows that

$$\begin{aligned}\Delta(H(g))_t &= [H(g_t) - H(g_{t-})] \\ &= F(g_{t-})\,[G(g_t) - G(g_{t-})] + G(g_{t-})\,[F(g_t) - F(g_{t-})] \\ &\quad + [F(g_t) - F(g_{t-})]\,[G(g_t) - G(g_{t-})]\,.\end{aligned}$$

Thus combining the last four displays we have

$$\begin{aligned}d(H(g))_t &= [F(g_{t-})G'(g_{t-}) + G(g_{t-})F'(g_{t-})]\,dg_t + F(g_{t-})[G(g_t) - G(g_{t-})] \\ &\quad +G(g_{t-})[F(g_t) - F(g_{t-})] - [F(g_{t-})G'(g_{t-}) + G(g_{t-})G'(g_{t-})]\,\Delta g_t \\ &\quad +[F(g_t) - F(g_{t-})][G(g_t) - G(g_{t-})]. \\ &= H'(g_{t-})dg_t + \Delta(H(g))_t - H'(g_{t-})\Delta g_t.\end{aligned}$$

Thus

$$d(H(g))_t = H'(g_{t-})dg_t + [\Delta(H(g))_t - H'(g_{t-})\Delta g_t]\,,$$

which shows that $h \in \mathcal{C}$ and thus C is closed under multiplication. Since $\mathcal{C}$ trivially contains the function $F(x) = x$, it follows that $\mathcal{C}$ contains all polynomials. Now let F be any C^1 function. It is enough to prove (A.7) for all $t \in [0, t_0]$ for any $t_0 < \infty$. Since g is right-continuous there exists M such that $g(t) \in [-M, M]$ for all $t \in [0, t_0]$. Moreover, there exist polynomials p_n such that $p_n \to F$ and $p'_n \to F$ uniformly on $[-M, M]$. Since (A.7) is true for each p_n, it follows from standard convergence theorems that (A.7) is satisfied by all C^1 functions F. □

One can similarly establish the N-dimensional generalization. Given an $\mathbb{R}^N$-valued function $g = (g^1, g^2, \ldots, g^N)$ of finite variation and function F on $\mathbb{R}^n$ with continuous first-order partial derivatives $D_i f$, it follows that

$$F(g_t)-F(g_0) = \sum_{i=1}^{N} \int_{(0,t]} D_i F(g_{s-}) dg^i_s + \sum_{0<s\leq t} \left[F(g_s) - F(g_{s-}) - \sum_{i=1}^{N} D_i f(g_{s-}) \Delta g^i_s \right].$$

A.5 Other Useful Results

1. Borel-Cantelli Lemmas

The Borel Cantelli Lemma, is an extremely powerful result that is particularly useful for establishing a.s. limit theorems. Given a sequence of events E_n, $n \in \mathbb{N}$, recall that the set of points that lie in infinitely number n (referred to as the event that E_n occurs infinitely often) is given by

$$\limsup E_n \doteq \cap_m \cup_{n\geq m} E_n.$$

Lemma A.5.1. *(Borel-Cantelli) Let $\{E_n\}$ be a sequence of events such that $\sum_n P(E_n) < \infty$. Then*

$$P(\limsup E_n) = P(E_n \textit{ infinitely often }) = 0.$$

The proof is easy and left to the reader.
A partial converse of this also holds.

Lemma A.5.2. *(Borel-Cantelli 2) Let E_n be a sequence of independent events such that $\sum_n P(E_n) = \infty$. Then*

$$P(E_n \textit{ infinitely often }) = 1.$$

2. Monotone Class Theorems

Dynkin system. A collection $\mathcal{D}$ of subsets of Ω is called a *Dynkin system* or a *d*-system if

1. $\Omega \in \mathcal{D}$;
2. $A, B \in \mathcal{D}$ and $B \subseteq A$ imply $A \setminus B \in \mathcal{D}$;
3. $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{D}$ and $A_1 \subseteq A_2 \subseteq ...$ imply $\cup_{n=1}^{\infty} A_n \in \mathcal{D}$.

The Dynkin system generated by a collection of subsets $\mathcal{C}$ (which is defined to be the intersection of all Dynkin systems containing $\mathcal{C}$ – check for yourself that this intersection is indeed still a Dynkin system) is denoted by $d(\mathcal{C})$.

π-system. A collection $\mathcal{I}$ of subsets of Ω is called a π-system if $\mathcal{I}$ is closed under finite intersections.

Theorem A.5.3. *(**Dynkin**) If $\mathcal{I}$ is a π-system then*

$$d(\mathcal{I}) = \sigma(\mathcal{I}).$$

This theorem is extremely powerful and is often used in the following context. Often one wants to show that a certain property A holds for all sets in a σ-algebra $\mathcal{F}$. In that case, one first shows that the property holds for a π-system, and for Ω and then shows that the property is "closed" under the operations (ii) and (iii) described in the Dynkin system theorem.

For example, showing that a function is $(\mathcal{F},\mathcal{S})$ measurable involves showing that for all sets in $\mathcal{S}$, the inverse image of the set under the function lies in $\mathcal{F}$ – the fact that this collection of sets is a Dynkin system is a simple consequence of the fact that $\mathcal{F}$ is a σ-algebra. Thus it is sufficient to show that the inverse image of any π-system that generates $\mathcal{S}$ lies in $\mathcal{F}$. Likewise, showing that the conditional expectation of a random variable with respect to a σ-algebra $\mathcal{F}$ is a.s. equal to some other random variable can also be phrased in this manner. Here one wants to show that for all sets in $\mathcal{F}$ the integral of the two random variables over the sets is equal. The fact that this collection of sets is a d-system follows easily from elementary properties of the integral. Thus once again, the Dynkin system theorem shows us that it is sufficient to check equality only for some simple π-system that generates $\mathcal{F}$.

Remark. Note that if you first show that the property is true for a collection of sets in an algebra (and not just a π-system) then properties (1) and (2) of a Dynkin system are automatically satisfied. And hence one only has to check that if the property holds for a countable sequence of increasing sets, then the property also holds for the union.

A.6 Some Functional Analytic Results

A.6.1 Some Useful Inequalities

Lemma A.6.1. (Cauchy-Schwarz and Hölder) *If $f, g \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, then $fg \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and*

$$\|fg\|_1 \leq \|f\|_2 \|g\|_2 .$$

More generally, if $f \in L^p(\Omega, \mathcal{F}, \mathbb{P})$, $g \in L^q(\Omega, \mathcal{F}, \mathbb{P})$ where $1/p + 1/q = 1$, then

$$\|fg\|_1 \leq \|f\|_p \|g\|_p .$$

A.6.2 The Banach-Steinhauss Theorem

The Banach-Steinhauss theorem or the principle of "uniform boundedness" is one of the main results in functional analysis. It states that for a family of continuous linear operators on a Banach space, pointwise boundedness is equivalent to boundedness in operator norm.

Definition A.6.2. *A Banach space is a complete normed vector space.*

Definition A.6.3. *An operator $T : X \to Y$ from one real vector space to another is said to be linear if*

$$T(ax + by) = aT(x) + bT(y)$$

for every $a, b \in \mathbb{R}$ and $x, y \in X$. A linear operator is continuous if it maps bounded sets into bounded sets.

Definition A.6.4. *The norm $\|T\|$ of an operator $T : X \to Y$ from one normed (real) vector space to another is defined as follows*

$$\|T\| \doteq \min\{c : \|Tx\| \leq c \|x\| \text{ for all } x \in X\}.$$

Theorem A.6.5. (Banach-Steinhauss Theorem) *Suppose X is a Banach space and Y is a normed vector space. Suppose F is a collection of continuous, linear operators from X to Y. If*

$$\sup_{T \in F} \|T(x)\| < \infty \quad \text{for every } x \in X$$

then

$$\sup_{T \in F} \|T\| < \infty.$$

Bibliography

[1] Z. Ciesielski. Hölder continuity for realizations of Gaussian processes. *Trans. Amer. Math. Soc.*, **103**, 434–450, 1962.

[2] S.N. Ethier and T.G. Kurtz, *Markov processes: Characterization and Convergence*, Wiley, 1986.

[3] K. Ito and H. Mckean. *Diffusion Processes and their Sample Paths.* Springer-Verlag, 1974.

[4] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus.* Springer-Verlag, New York, 1991.

[5] P. Lévy. Sur certains processus stochastiques homegenes. *Compositio Math.* **7**, 283–339, 1939.

[6] D. Revuz and M. Yor *Continuous Martingales and Brownian Motion* Springer, 1999

[7] H.L.Royden *Real Analysis* Prentice-Hall, 1988.

[8] D. Williams. Probability with Martingales, Cambridge University Press, 1991.

[9] W. Whitt. *Stochastic-Process Limits*, Springer-Verlag, 2001.