36-752: Lecture 1

How is this course different from your earlier probability courses? There are some problems that simply can't be handled with finite-dimensional sample spaces and random variables that are either discrete or have densities.

Example 1. Try to express the strong law of large numbers without using an infinite-dimensional space. Oddly enough, the weak law of large numbers requires only a sequence of finite-dimensional spaces, but the strong law concerns entire infinite sequences.

Example 2. Consider a distribution whose cumulative distribution function (cdf) increases continuously part of the time but has some jumps. Such a distribution is neither discrete nor continuous. How do you define the mean of such a random variable? Is there a way to treat such distributions together with discrete and continuous ones in a unified manner?

General Measures. Both of the above examples are accommodated by a generalization of the theories of summation and integration. Indeed, summation becomes a special case of the more general theory of integration. It all begins with a generalization of the concept of "size" of a set.

Example 3. One way to measure the size of a set is to count its elements. All infinite sets would have the same size (unless you distinguish different infinite cardinals).

Example 4. Special subsets of Euclidean spaces can be measured by length, area, volume, etc. But what about sets with lots of holes in them? For example, how large is the set of irrational numbers between 0 and 1?

We will use measures to say how large sets are. First, we have to decide which sets we will measure.

Definition 5. Let Ω be a set. A collection $\mathcal{F}$ of subsets of Ω is called a *field* if it satisfies

- $\Omega \in \mathcal{F}$,
- for each $A \in \mathcal{F}$, $A^C \in \mathcal{F}$,
- for all $A_1, A_2 \in \mathcal{F}$, $A_1 \cup A_2 \in \mathcal{F}$.

A field $\mathcal{F}$ is a *σ-field* if, in addition, it satisfies

- for every sequence $\{A_k\}_{k=1}^{\infty}$ in $\mathcal{F}$, $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$.

We will define measures on fields and σ-field's. A set Ω together with a σ-field $\mathcal{F}$ is called a *measurable space* $(\Omega, \mathcal{F})$, and the elements of $\mathcal{F}$ are called *measurable sets* .

EXAMPLE 6. Let $\Omega = \mathbb{R}$ and define $\mathcal{U}$ to be the collection of all unions of finitely many disjoint intervals of the form $(a, b]$ or $(-\infty, b]$ or (a, ∞) or $(-\infty, \infty)$, together with $\varnothing$. Then $\mathcal{U}$ is a field.

EXAMPLE 7. (POWER SET) Let Ω be an arbitrary set. The collection of all subsets of Ω is a σ-field. It is denoted 2^Ω and is called the *power set of* Ω.

EXAMPLE 8. (TRIVIAL σ-FIELD) Let Ω be an arbitrary set. Let $\mathcal{F} = \{\Omega, \varnothing\}$. This is the *trivial σ-field.*

DEFINITION 9. The *extended reals* is the set of all real numbers together with ∞ and $-\infty$. We shall denote this set $\overline{\mathbb{R}}$. The *positive extended reals*, denoted $\overline{\mathbb{R}}^+$ is $(0, \infty]$, and the *nonnegative extended reals*, denoted $\overline{\mathbb{R}}^{+0}$ is $[0, \infty]$.

DEFINITION 10. Let $(\Omega, \mathcal{F})$ be a measurable space. Let $\mu : \mathcal{F} \to \overline{\mathbb{R}}^{+0}$ satisfy

- $\mu(\varnothing) = 0$,
- for every sequence $\{A_k\}_{k=1}^\infty$ of mutually disjoint elements of $\mathcal{F}$, $\mu(\bigcup_{k=1}^\infty A_k) = \sum_{k=1}^\infty \mu(A_k)$.

Then μ is called a *measure* on $(\Omega, \mathcal{F})$ and $(\Omega, \mathcal{F}, \mu)$ is a *measure space.* If $\mathcal{F}$ is merely a field, then a μ that satisfies the above two conditions whenever $\bigcup_{k=1}^\infty A_k \in \mathcal{F}$ is called a *measure on the field* $\mathcal{F}$.

EXAMPLE 11. Let Ω be arbitrary with $\mathcal{F}$ the trivial σ-field. Define $\mu(\varnothing) = 0$ and $\mu(\Omega) = c$ for arbitrary $c > 0$ (with $c = \infty$ possible).

EXAMPLE 12. (COUNTING MEASURE) Let Ω be arbitrary and $\mathcal{F} = 2^\Omega$. For each finite subset A of Ω, define $\mu(A)$ to be the number of elements of A. Let $\mu(A) = \infty$ for all infinite subsets. This is called *counting measure* on Ω.

36-752: Lecture 2

For every collection $\mathcal{C}$ of subsets of Ω, there is a smallest field containing $\mathcal{C}$ and a smallest σ-field containing $\mathcal{C}$. These are called the *field generated by* $\mathcal{C}$ and the *σ-fieldgenerated by* $\mathcal{C}$. Just check that the intersection of an arbitrary collection of fields is a field and the intersection of an arbitrary collection of σ-field's is a σ-field. These collections are nonempty because 2^Ω is always a σ-field that contains every collection of subsets of Ω. The σ-field generated by $\mathcal{C}$ is sometimes denoted $\sigma(\mathcal{C})$.

EXERCISE 13. Let $\mathcal{F}_1, \mathcal{F}_2, \ldots$ be classes of sets in a common space Ω such that $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ for each n. Show that if each $\mathcal{F}_n$ is a field, then $\cup_{n=1}^\infty \mathcal{F}_n$ is also a field.

If each $\mathcal{F}_n$ is a σ-field, then is $\cup_{n=1}^\infty \mathcal{F}_n$ also necessarily a σ-field? Think about the following case: Ω is the set of nonnegative integers and $\mathcal{F}_n \equiv \sigma(\{\{0\}, \{1\}, \ldots, \{n\}\})$.

EXAMPLE 14. Let $\mathcal{C} = \{A\}$ for some nonempty A that is not itself Ω. Then $\sigma(\mathcal{C}) = \{\varnothing, A, A^C, \Omega\}$.

EXAMPLE 15. Let $\Omega = \mathbb{R}$ and let $\mathcal{C}$ be the collection of all intervals of the form $(a, b]$. Then the field generated by $\mathcal{C}$ is $\mathcal{U}$ from Example 6 while $\sigma(\mathcal{C})$ is larger.

EXAMPLE 16. (BOREL σ-FIELD) Let Ω be a topological space and let $\mathcal{C}$ be the collection of open sets. Then $\sigma(\mathcal{C})$ is called the *Borel σ-field*. If $\Omega = \mathbb{R}$, the Borel σ-field is the same as $\sigma(\mathcal{C})$ in Example 15. The Borel σ-field of subsets of $\mathbb{R}^k$ is denoted $\mathcal{B}^k$.

EXERCISE 17. Give some examples of classes of sets $\mathcal{C}$ such that $\sigma(\mathcal{C}) = \mathcal{B}^1$.

EXERCISE 18. Are there subsets of $\mathbb{R}$ which are not in $\mathcal{B}^1$?

DEFINITION 19. Let $(\Omega, \mathcal{F}, P)$ be a measure space. If $P(\Omega) = 1$, then P is called a *probability*, $(\Omega, \mathcal{F}, P)$ is a *probability space*, and elements of $\mathcal{F}$ are called *events*.

Sometimes, if the name of the probability P is understood or is not even mentioned, we will denote $P(E)$ by $\Pr(E)$ for events E.

Infinite measures pose a few unique problems. Some infinite measures are just like finite ones.

DEFINITION 20. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $\mathcal{C} \subseteq \mathcal{F}$. Suppose that there exists a sequence $\{A_n\}_{n=1}^\infty$ of elements of $\mathcal{C}$ such that $\mu(A_n) < \infty$ for all n and $\Omega = \bigcup_{n=1}^\infty A_n$. Then we say that μ is *σ-finite on $\mathcal{C}$*. If μ is σ-finite on $\mathcal{F}$, we merely say that μ is *σ-finite*.

EXAMPLE 21. Let $\Omega = \mathbb{Z}$ with $\mathcal{F} = 2^\Omega$ and μ being counting measure. This measure is σ-finite. Counting measure on an uncountable space is not σ-finite.

EXERCISE 22. Prove the claims in Example 21.

Properties of Measures. There are several useful properties of measures that are worth knowing.

First, measures are countably subadditive in the sense that

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n), \tag{23}$$

for arbitrary sequences $\{A_n\}_{n=1}^{\infty}$. The proof of this uses a standard trick for dealing with countable sequences of sets. Let $B_1 = A_1$ and let $B_n = A_n \setminus \bigcup_{i=1}^{n-1} B_i$ for $n > 1$. The B_n's are disjoint and have the same finite and countable unions as the A_n's. The proof of (23) relies on the additional fact that $\mu(B_n) \leq \mu(A_n)$ for all n.

Next, if $\mu(A_n) = 0$ for all n, it follows that $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = 0$. This gets used a lot in proofs. Similarly, if μ is a probability and $\mu(A_n) = 1$ for all n, then $\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = 1$.

Definition 24. Suppose that some statement about elements of Ω holds for all $\omega \in A^C$ where $\mu(A) = 0$. Then we say that the statement holds *almost everywhere*, denoted a.e. $[\mu]$. If P is a probability, then almost everywhere is often replaced by *almost surely*, denoted a.s. $[P]$.

Example 25. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of functions from Ω to $\mathbb{R}$. To say that X_n converges to X a.s. $[P]$ (denoted $X_n \overset{\text{a.s.}}{\to} X$) means that there is a set A with $P(A) = 0$ and $P(\{\omega \in A^C : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$.

Proposition 26. *If $\mu_1, \mu_2, \ldots$ are all measures on $(\Omega, \mathcal{F})$ and if $\{a_n\}_{n=1}^{\infty}$ is a sequence of positive numbers, then $\sum_{n=1}^{\infty} a_n \mu_n$ is a measure on $(\Omega, \mathcal{F})$.*

Exercise 27. Prove Proposition 26.

Definition 28. Define the *indicator function* $I_A : \Omega \to \{0, 1\}$ for the set $A \subseteq \Omega$ as $I_A(\omega) = 1$ if $\omega \in A$ and $I_A(\omega) = 0$ if $\omega \in A^C$.

Definition 29. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. A sequence $\{A_n\}_{n=1}^{\infty}$ of elements of $\mathcal{F}$ is called *monotone increasing* if $A_n \subseteq A_{n+1}$ for each n. It is *monotone decreasing* if $A_n \supseteq A_{n+1}$ for each n. For a general sequence, we define

$$\begin{aligned} \limsup_{n\to\infty} A_n &= \bigcap_{i=1}^{\infty} \bigcup_{n=i}^{\infty} A_n, \\ \liminf_{n\to\infty} A_n &= \bigcup_{i=1}^{\infty} \bigcap_{n=i}^{\infty} A_n. \end{aligned}$$

If $\limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n$, the common set is called $\lim_{n\to\infty} A_n$. The set $\limsup_{n\to\infty} A_n$ is often called A_n *infinitely often* or A_n i.o. because a point ω is in that set if and only if ω is in infinitely many of the A_n sets. The set $\liminf_{n\to\infty} A_n$ is often called A_n *all but finitely often* or A_n *eventually* (A_n ev.). This set has all those ω such that ω is in all of the A_n except possibly finitely many of the A_n, i.e., eventually.

EXERCISE 30. What is the relationship between the definition of the $\limsup$ and $\liminf$ of a sequence of reals $\{x_n\}_{n=1}^{\infty}$ and this definition of the $\limsup$ and $\liminf$ of a sequence of sets?

EXERCISE 31. Define A_n to be the set $(-1/n, 1]$ if n is odd, and to be $(-1, 1/n]$ if n is even. What are $\limsup_{n\to\infty} A_n$ and $\liminf_{n\to\infty} A_n$?

It is easy to establish some simple facts about these limiting sets.

PROPOSITION 32. *Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of sets.*

- $\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n$, *if and only if, for each ω, $\lim_{n\to\infty} I_{A_n}(\omega)$ exists.*
- *If the sequence is monotone increasing, then $\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$.*
- *If the sequence is monotone decreasing, then $\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n$.*

EXERCISE 33. Prove Proposition 32.

36-752: Lecture 3

LEMMA 34. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let $\{A_n\}_{n=1}^{\infty}$ be a monotone sequence of elements of $\mathcal{F}$. Then $\lim_{n\to\infty} \mu(A_n) = \mu\left(\lim_{n\to\infty} A_n\right)$ if either of the following hold:*

- *the sequence is increasing,*
- *the sequence is decreasing and $\mu(A_k) < \infty$ for some k.*

PROOF. Define $A_\infty = \lim_{n\to\infty} A_n$. In the first case, write $B_1 = A_1$ and $B_n = A_n \setminus A_{n-1}$ for $n > 1$. Then $A_n = \bigcup_{k=1}^n B_k$ for all n (including $n = \infty$). Then $\mu(A_n) = \sum_{k=1}^n \mu(B_k)$, and

$$\mu\left(\lim_{n\to\infty} A_n\right) = \mu(A_\infty) = \sum_{k=1}^{\infty} \mu(B_k) = \lim_{n\to\infty} \sum_{k=1}^{n} \mu(B_k) = \lim_{n\to\infty} \mu(A_n).$$

In the second case, write $B_n = A_n \setminus A_{n+1}$ for all $n \geq k$. Then, for all $n > k$,

$$\begin{aligned} A_k \setminus A_n &= \bigcup_{i=k}^{n-1} B_i, \\ A_k \setminus A_\infty &= \bigcup_{i=k}^{\infty} B_i. \end{aligned}$$

By the first case,

$$\lim_{n\to\infty} \mu(A_k \setminus A_n) = \mu\left(\bigcup_{i=k}^{\infty} B_i\right) = \mu(A_k \setminus A_\infty).$$

Because $A_n \subseteq A_k$ for all $n > k$ and $A_\infty \subseteq A_k$, it follows that

$$\begin{aligned} \mu(A_k \setminus A_n) &= \mu(A_k) - \mu(A_n), \\ \mu(A_k \setminus A_\infty) &= \mu(A_k) - \mu(A_\infty). \end{aligned}$$

It now follows that $\lim_{n\to\infty} \mu(A_n) = \mu(A_\infty)$. $\Box$

EXERCISE 35. Construct a simple counterexample to show that the condition $\mu(A_k) < \infty$ is required in the second claim of Lemma 34.

Uniqueness of Measures. There is a popular method for proving uniqueness theorems about measures. The idea is to define a function μ on a convenient class $\mathcal{C}$ of sets and then prove that there can be at most one extension of μ to $\sigma(\mathcal{C})$.

EXAMPLE 36. Suppose it is given that for any $a \in \mathbb{R}$,

$$P((-\infty, a]) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} \exp\left(-u^2/2\right) du.$$

Does that uniquely define a probability measure on the class of Borel subsets of the line, $\mathcal{B}^1$?

Definition 37. A collection $\mathcal{A}$ of subsets of Ω is a *π-system* if, for all $A_1, A_2 \in \mathcal{A}$, $A_1 \cap A_2 \in \mathcal{A}$. A class $\mathcal{C}$ is a *λ-system* if

- $\Omega \in \mathcal{C}$,
- for each $A \in \mathcal{C}$, $A^C \in \mathcal{C}$,
- for each sequence $\{A_n\}_{n=1}^{\infty}$ of disjoint elements of $\mathcal{C}$, $\bigcup_{n=1}^{\infty} A_n \in \mathcal{C}$.

Example 38. The collection of all intervals of the form $(-\infty, a]$ is a π-system of subsets of $\mathbb{R}$. So too is the collection of all intervals of the form $(a, b]$ (together with $\varnothing$). The collection of all sets of the form $\{(x, y) : x \leq a, y \leq b\}$ is a π-system of subsets of $\mathbb{R}^2$. So too is the collection of all rectangles with sides parallel to the coordinate axes.

Some simple results about π-systems and λ-systems are the following.

Proposition 39. *If Ω is a set and $\mathcal{C}$ is both a π-system and a λ-system, then $\mathcal{C}$ is a σ-field.*

Proposition 40. *Let Ω be a set and let Λ be a λ-system of subsets. If $A \in \Lambda$ and $A \cap B \in \Lambda$ then $A \cap B^C \in \Lambda$.*

Exercise 41. Prove Propositions 39 and 40.

Lemma 42. ($\pi - \lambda$ theorem) *Let Ω be a set and let Π be a π-system and let Λ be a λ-system that contains Π. Then $\sigma(\Pi) \subseteq \Lambda$.*

Proof. Define $\lambda(\Pi)$ to be the smallest λ-system containing Π. For each $A \subseteq \Omega$, define $\mathcal{G}_A$ to be the collection of all sets $B \subseteq \Omega$ such that $A \cap B \in \lambda(\Pi)$.

First, we show that $\mathcal{G}_A$ is a λ-system for each $A \in \lambda(\Pi)$. To see this, note that $A \cap \Omega \in \lambda(\Pi)$, so $\Omega \in \mathcal{G}_A$. If $B \in \mathcal{G}_A$, then $A \cap B \in \lambda(\Pi)$, and Proposition 40 says that $A \cap B^C \in \lambda(\Pi)$, so $B^C \in \mathcal{G}_A$. Finally, $\{B_n\}_{n=1}^{\infty} \in \mathcal{G}_A$ with the B_n disjoint implies that $A \cap B_n \in \lambda(\Pi)$ with $A \cap B_n$ disjoint, so their union is in $\lambda(\Pi)$. But their union is $A \cap (\bigcup_{n=1}^{\infty} B_n)$. So $\bigcup_{n=1}^{\infty} B_n \in \mathcal{G}_A$.

Next, we show that $\lambda(\Pi) \subseteq \mathcal{G}_C$ for every $C \in \lambda(\Pi)$. Let $A, B \in \Pi$, and notice that $A \cap B \in \Pi$, so $B \in \mathcal{G}_A$. Since $\mathcal{G}_A$ is a λ-system containing Π, it must contain $\lambda(\Pi)$. It follows that $A \cap C \in \lambda(\Pi)$ for all $C \in \lambda(\Pi)$. If $C \in \lambda(\Pi)$, it then follows that $A \in \mathcal{G}_C$. So, $\Pi \subseteq \mathcal{G}_C$ for all $C \in \lambda(\Pi)$. Since $\mathcal{G}_C$ is a λ-system containing Π, it must contain $\lambda(\Pi)$.

Finally, if $A, B \in \lambda(\Pi)$, we just proved that $B \in \mathcal{G}_A$, so $A \cap B \in \lambda(\Pi)$ and hence $\lambda(\Pi)$ is also a π-system. By Proposition 39, $\lambda(\Pi)$ is a σ-field containing Π and hence must contain $\sigma(\Pi)$. Since $\lambda(\Pi) \subseteq \Lambda$, the proof is complete. $\Box$

The uniqueness theorem is the following.

Theorem 43. *Suppose that μ_1 and μ_2 are measures on $(\Omega, \mathcal{F})$ and $\mathcal{F}$ is the smallest σ-field containing the π-system Π. If μ_1 and μ_2 are both σ-finite on Π and they agree on Π, then they agree on $\mathcal{F}$.*

Proof. First, let $C \in \Pi$ be such that $\mu_1(C) = \mu_2(C) < \infty$, and define $\mathcal{G}_C$ to be the collection of all $B \in \mathcal{F}$ such that $\mu_1(B \cap C) = \mu_2(B \cap C)$. It is easy to see that $\mathcal{G}_C$ is a λ-system that contains Π, hence it equals $\mathcal{F}$ by Lemma 42. (For example, if $B \in \mathcal{G}_C$,

$$\mu_1(B^C \cap C) = \mu_1(C) - \mu_1(B \cap C) = \mu_2(C) - \mu_2(B \cap C) = \mu_2(B^C \cap C),$$

so $B^C \in \mathcal{G}_C$.)

Since μ_1 and μ_2 are σ-finite, there exists a sequence $\{C_n\}_{n=1}^\infty \in \Pi$ such that $\mu_1(C_n) = \mu_2(C_n) < \infty$, and $\Omega = \bigcup_{n=1}^\infty C_n$. (Since Π is only a π-system, we cannot assume that the C_n are disjoint.) For each $A \in \mathcal{F}$,

$$\mu_j(A) = \lim_{n\to\infty} \mu_j\left(\bigcup_{i=1}^n [C_i \cap A]\right) \text{ for } j = 1, 2.$$

Since $\mu_j\left(\bigcup_{i=1}^n [C_i \cap A]\right)$ can be written as a linear combination of values of μ_j at sets of the form $A \cap C$, where $C \in \Pi$ is the intersection of finitely many of $C_1, \dots, C_n$, it follows from $A \in \mathcal{G}_C$ that $\mu_1\left(\bigcup_{i=1}^n [C_i \cap A]\right) = \mu_2\left(\bigcup_{i=1}^n [C_i \cap A]\right)$ for all n, hence $\mu_1(A) = \mu_2(A)$. $\square$

Exercise 44. Return to Example 36. You should now be able to answer the question posed there.

Exercise 45. Suppose that $\Omega = \{a, b, c, d, e\}$ and I tell you the value of $P(\{a, b\})$ and $P(\{b, c\})$. For which subset of Ω do I need to define $P(\cdot)$ in order to have a unique extension of P to a σ-field of subsets of Ω?

36-752: Lecture 4

Measures Based on Increasing Functions. Let F be a cdf (nondecreasing, right-continuous, limits equal 0 and 1 at $-\infty$ and ∞ respectively). Let $\mathcal{U}$ be the field in Example 6. Define $\mu : \mathcal{U} \to [0,1]$ by $\mu(A) = \sum_{k=1}^{n} F(b_k) - F(a_k)$ when $A = \bigcup_{k=1}^{n}(a_k, b_k]$ and $\{(a_k, b_k]\}$ are disjoint. This set-function is well-defined and finitely additive. To see that it is well-defined, look at an alternative representation as $\mu(A) = \sum_{j=1}^{m} F(d_j) - F(c_j)$. Consider the partition of A into the refinement of the two partitions given. The sum over the refinement is the same as both of the two sums we started with. Is μ countably additive as probabilities are supposed to be? That is, if $A = \bigcup_{i=1}^{\infty} A_i$ where the A_i's are disjoint, each A_i is a union of finitely many disjoint intervals, and A itself is the union of finitely many disjoint intervals $(a_k, b_k]$ for $k = 1, \ldots, n$, does $\mu(A) = \sum_{i=1}^{\infty} \mu(A_i)$? First, take the collection of intervals that go into all of the A_i's and split them, if necessary, so that each is a subset of at most one of the $(a_k, b_k]$ intervals. Then apply the following result to each $(a_k, b_k]$.

LEMMA 46. *Let $(a,b] = \bigcup_{k=1}^{\infty}(c_k, d_k]$ with the $(c_k, d_k]$'s disjoint. Then $F(b) - F(a) = \sum_{k=1}^{\infty} F(d_k) - F(c_k)$.*

PROOF. Since $(a,b] \supseteq \bigcup_{k=1}^{n}(c_k, d_k]$ for all n, it follows that $F(b) - F(a) \geq \sum_{k=1}^{n} F(d_k) - F(c_k)$, hence $F(b) - F(a) \geq \sum_{k=1}^{\infty} F(d_k) - F(c_k)$. We need to prove the opposite inequality.

Suppose first that both a and b are finite. Let $\epsilon > 0$. For each k, there is $e_k > d_k$ such that

$$F(d_k) \leq F(e_k) \leq F(d_k) + \frac{\epsilon}{2^k}.$$

Also, there is $f > a$ such that $F(a) \geq F(f) - \epsilon$. Now, the interval $[f, b]$ is compact and $[f,b] \subseteq \bigcup_{k=1}^{\infty}(c_k, e_k)$. So there are finitely many (c_k, e_k)'s (suppose they are the first n) such that $[f,b] \subseteq \bigcup_{k=1}^{n}(c_k, e_k)$. Now,

$$F(b) - F(a) \leq F(b) - F(f) + \epsilon \leq \epsilon + \sum_{k=1}^{n} F(e_k) - F(c_k) \leq 2\epsilon + \sum_{k=1}^{n} F(d_k) - F(c_k).$$

It follows that $F(b) - F(a) \leq 2\epsilon + \sum_{k=1}^{\infty} F(d_k) - F(c_k)$. Since this is true for all $\epsilon > 0$, it is true for $\epsilon = 0$.

If $-\infty = a < b < \infty$, let $g > -\infty$ be such that $F(g) < \epsilon$. The above argument shows that

$$F(b) - F(g) \leq \sum_{k=1}^{\infty} F(d_k \vee g) - F(c_k \vee g) \leq \sum_{k=1}^{\infty} F(d_k) - F(c_k).$$

Since $\lim_{g \to -\infty} F(g) = 0$, it follows that $F(b) \leq \sum_{k=1}^{\infty} F(d_k) - F(c_k)$. Similar arguments work when $a < b = \infty$ and $-\infty = a < b = \infty$. $\square$

In Lemma 46 you can replace F by an arbitrary nondecreasing right-continuous function with only a bit more effort. (See the supplement following at the end of this lecture.)

The function μ defined in terms of a nondecreasing right-continuous function is a measure on the field $\mathcal{U}$. There is an extension theorem that gives conditions under which a measure on a field can be extended to a measure on the generated σ-field. Furthermore, the extension is unique.

Example 47. (Lebesgue measure) Start with the function $F(x) = x$, form the measure μ on the field $\mathcal{U}$ and extend it to the Borel σ-field. The result is called *Lebesgue measure*, and it extends the concept of "length" from intervals to more general sets.

Example 48. Every distribution function for a random variable has a corresponding probability measure on the real line.

Another concept that is occasionally useful is that of a complete measure space.

Definition 49. A measure space $(\Omega, \mathcal{F}, \mu)$ is *complete* if, for every $A \in \mathcal{F}$ such that $\mu(A) = 0$ and every $B \subseteq A$, $B \in \mathcal{F}$.

Theorem 50. (Caratheodory extension) *Let μ be a σ-finite measure on the field $\mathcal{C}$ of subsets of Ω. There exists a σ-field $\mathcal{A}$ that contains $\mathcal{C}$ and a unique extension μ^* of μ to a measure on $(\Omega, \mathcal{A})$. Furthermore $(\Omega, \mathcal{A}, \mu^*)$ is a complete measure space.*

Exercise 51. In this exercise, we prove Theorem 50.

First, for each $B \in 2^\Omega$, define

$$\mu^*(B) = \inf \sum_{i=1}^{\infty} \mu(A_i), \tag{52}$$

where the inf is taken over all $\{A_i\}_{i=1}^{\infty}$ such that $B \subseteq \bigcup_{i=1}^{\infty} A_i$ and $A_i \in \mathcal{C}$ for all i. Since $\mathcal{C}$ is a field, we can assume that the A_i's are mutually disjoint without changing the value of $\mu^*(B)$. Let

$$\mathcal{A} = \{B \in 2^\Omega : \mu^*(C) = \mu^*(C \cap B) + \mu^*(C \cap B^C), \text{ for all } C \in 2^\Omega\}.$$

Now take the following steps:

1. Show that μ^* extends μ, i.e. that $\mu^*(A) = \mu(A)$ for each $A \in \mathcal{C}$.
2. Show that μ^* is monotone and subadditive.
3. Show that $\mathcal{C} \subseteq \mathcal{A}$.
4. Show that $\mathcal{A}$ is a field.
5. Show that μ^* is finitely additive on $\mathcal{A}$.
6. Show that $\mathcal{A}$ is a σ-field.
7. Show that μ^* is countably additive on $\mathcal{A}$.
8. Show that μ^* is the unique extension of μ to a measure of $(\Omega, \mathcal{A})$.
9. Show that $(\Omega, \mathcal{A}, \mu^*)$ is a complete measure space.

Supplement: Measures from Increasing Functions

Lemma 46 deals only with functions F that are cdf's. Suppose that F is an unbounded nondecreasing function that is continuous from the right. If $-\infty < a < b < \infty$, then the proof of Lemma 46 still applies. Suppose that $(-\infty, b] = \bigcup_{k=1}^{\infty}(c_k, d_k]$ with $b < \infty$ nd all $(c_k, d_k]$ disjoint. Suppose that $\lim_{x\to-\infty} F(x) = -\infty$. We want to show that $\sum_{k=1}^{\infty} F(d_k) - F(c_k) = \infty$. If one $c_k = -\infty$, the proof is immediate, so assume that all $c_k > -\infty$. Then there must be a subsequence $\{k_j\}_{j=1}^{\infty}$ such that $\lim_{j\to\infty} c_{k_j} = -\infty$. For each j, let $\{(c'_{j,n}, d'_{j,n}]\}_{n=1}^{\infty}$ be the subsequence of intervals that cover $(c_{k_j}, b]$. For each j, the proof of Lemma 46 applies to show that

$$F(b) - F(c_{k_j}) = \sum_{n=1}^{\infty} F(d'_{j,n}) - F(c'_{j,n}). \tag{53}$$

As $j \to \infty$, the left side of (53) goes to ∞ while the right side eventually includes every interval in the original collection.

A similar proof works for an interval of the form (a, ∞) when $\lim_{x\to\infty} F(x) = \infty$. A combination of the two works for $(-\infty, \infty)$.

36-752: Lecture 5

Measurable Functions. Measurable functions are the types of functions that we can integrate with respect to measures in much the same way that continuous functions can be integrated "dx". Recall that the Riemann integral of a continuous function f over a bounded interval is defined as a limit of sums of lengths of subintervals times values of f on the subintervals. The measure of a set generalizes the length while elements of the σ-field generalize the intervals. Recall that a real-valued function is continuous if and only if the inverse image of every open set is open. This generalizes to the inverse image of every measurable set being measurable.

DEFINITION 54. Let $(\Omega, \mathcal{F})$ and $(S, \mathcal{A})$ be measurable spaces. Let $f : \Omega \to S$ be a function that satisfies $f^{-1}(A) \in \mathcal{F}$ for each $A \in \mathcal{A}$. Then we say that f is $\mathcal{F}/\mathcal{A}$-*measurable.* If the σ-field's are to be understood from context, we simply say that f is measurable.

EXAMPLE 55. Let $\mathcal{F} = 2^\Omega$. Then every function from Ω to a set S is measurable no matter what $\mathcal{A}$ is.

EXAMPLE 56. Let $\mathcal{A} = \{\varnothing, S\}$. Then every function from a set Ω to S is measurable, no matter what $\mathcal{F}$ is.

Proving that a function is measurable is facilitated by noticing that inverse image commutes with union, complement, and intersection. That is, $f^{-1}(A^C) = [f^{-1}(A)]^C$ for all A, and for arbitrary collections of sets $\{A_\alpha\}_{\alpha \in \aleph}$,

$$f^{-1}\left(\bigcup_{\alpha\in\aleph} A_\alpha\right) = \bigcup_{\alpha\in\aleph} f^{-1}(A_\alpha),$$

$$f^{-1}\left(\bigcap_{\alpha\in\aleph} A_\alpha\right) = \bigcap_{\alpha\in\aleph} f^{-1}(A_\alpha).$$

EXERCISE 57. Is the inverse image of a σ-field is a σ-field? That is, if $f : \Omega \to S$ and if $\mathcal{A}$ is a σ-field of subsets of S, then $f^{-1}(\mathcal{A})$ is a σ-field of subsets of Ω.

PROPOSITION 58. *If f is a continuous function from one topological space to another (each with Borel σ-field's) then f is measurable.*

The proof of this makes use of Lemma 60.

DEFINITION 59. Let $f : \Omega \to S$, where $(S, \mathcal{A})$ is a measurable space. The σ-field $f^{-1}(\mathcal{A})$ is called the *σ-field generated by f*. The σ-field $f^{-1}(\mathcal{A})$ is sometimes denoted $\sigma(f)$.

It is easy to see that $f^{-1}(\mathcal{A})$ is the smallest σ-field $\mathcal{C}$ such that f is $\mathcal{C}/\mathcal{A}$-measurable. We can now prove the following helpful result.

Lemma 60. *Let $(\Omega, \mathcal{F})$ and $(S, \mathcal{A})$ be measurable spaces and let $f : \Omega \to S$. Suppose that $\mathcal{C}$ is a collection of sets that generates $\mathcal{A}$. Then f is measurable if $f^{-1}(\mathcal{C}) \subseteq \mathcal{F}$.*

Exercise 61. Prove Proposition 60.

Definition 62. If $(\Omega, \mathcal{F}, P)$ is a probability space and $X : \Omega \to \overline{\mathbb{R}}$ is measurable, then X is called a *random variable*. In general, if $X : \Omega \to S$, where $(S, \mathcal{A})$ is a measurable space, we call X a *random quantity*.

Exercise 63. Prove the following. Let $S = \mathbb{R}$ in Lemma 60. Let D be a dense subset of $\mathbb{R}$, and let $\mathcal{C}$ be the collection of all intervals of the form $(-\infty, a)$, for $a \in D$. To prove that a real-valued function is measurable, one need only show that $\{\omega : f(\omega) < a\} \in \mathcal{F}$ for all $a \in D$. Similarly, we can replace $< a$ by $> a$ or $\leq a$ or $\geq a$.

Exercise 64. Show that a monotone increasing function is measurable.

Example 65. Suppose that $f : \Omega \to \overline{\mathbb{R}}$ takes values in the extended reals. Then $f^{-1}(\{-\infty, \infty\}) = [f^{-1}((-\infty, \infty))]^C$. Also

$$f^{-1}(\{\infty\}) = \bigcap_{n=1}^{\infty} \{\omega : f(\omega) > n\},$$

and similarly for $-\infty$. In order to check whether f is measurable, we need to see that the inverse images of all semi-infinite intervals are measurable sets. If we include the infinite endpoint in these intervals, then we don't need to check anything else. If we don't include the infinite endpoint, and if both infinite values are possible, then we need to check that at least one of $\{\infty\}$ or $\{-\infty\}$ has measurable inverse image.

Definition 66. A measurable function that takes at most finitely many values is called a *simple function*.

Example 67. Let $(\Omega, \mathcal{F})$ be a measurable space and let $A_1, \ldots, A_n$ be disjoint elements of $\mathcal{F}$, and let $a_1, \ldots, a_n$ be real numbers. Then $f = \sum_{i=1}^{n} a_i I_{A_i}$ defines a simple function since $f^{-1}((-\infty, a))$ is a union of at most finitely many measurable sets.

Definition 68. Let f be a simple function whose distinct values are $a_1, \ldots, a_n$, and let $A_i = \{\omega : f(\omega) = a_i\}$. Then $f = \sum_{i=1}^{n} a_i I_{A_i}$ is called the *canonical representation of f*.

Lemma 69. *Let f be a nonnegative measurable extended real-valued function from Ω. Then there exists a sequence $\{f_n\}_{n=1}^{\infty}$ of nonnegative (finite) simple functions such that $f_n \leq f$ for all n and $\lim_{n\to\infty} f_n(\omega) = f(\omega)$ for all ω.*

Proof. For each n, define $A_{n,k} = f^{-1}((k/n, (k+1)/n])$ for $k = 1, \ldots, n^2 - 1$ and $A_{n,0} = f^{-1}([0, 1/n] \cup (n, \infty))$. Let $A_\infty = f^{-1}(\{\infty\})$. Define $f_n(\omega) = \frac{1}{n} \sum_{k=0}^{n^2-1} k I_{A_{n,k}}(\omega) + n A_\infty$. The proof is easy to complete now. $\square$

Lemma 69 says that each nonnegative measurable function f can be approximated arbitrarily closely from below by simple functions. It is easy to see that if f is bounded the approximation is uniform once n is greater than the bound.

Many theorems about real-valued functions are easier to prove for nonnegative measurable functions. This leads to the common device of splitting a measurable function f as follows.

DEFINITION 70. Let f be a real-valued function. The *positive part* f^+ of f is defined as $f^+(\omega) = \max\{f(\omega), 0\}$. The *negative part* f^- of f is $f^-(\omega) = -\min\{f(\omega), 0\}$.

Notice that both the positive and negative parts of a function are nonnegative. It follows easily that $f = f^+ - f^-$. It is easy to prove that the positive and negative parts of a measurable function are measurable.

Here are some simple properties of measurable functions.

THEOREM 71. *Let $(\Omega, \mathcal{F})$, $(S, \mathcal{A})$, and $(T, \mathcal{B})$ be measurable spaces.*

1. *If f is an extended real-valued measurable function and a is a constant, then af is measurable.*
2. *If $f : \Omega \to S$ and $g : S \to T$ are measurable, then $g(f) : \Omega \to T$ is measurable.*
3. *If f and g are measurable real-valued functions, then $f + g$ and fg are measurable.*

PROOF. For $a = 0$, part 1 is trivial. Assume $a \neq 0$. Because

$$\{\omega : af(\omega) < c\} = \begin{cases} \{\omega : f(\omega) < c/a\} & \text{if } a > 0, \\ \{\omega : f(\omega) > c/a\} & \text{if } a < 0, \end{cases}$$

we see that af is measurable.

For part 2, just notice that $[g(f)]^{-1}(B) = f^{-1}(g^{-1}(B))$.

For part 3, let $h : \mathbb{R}^2 \to \mathbb{R}$ be defined by $h(x, y) = x + y$. This function is continuous, hence measurable. Then $f + g = h(f, g)$. We now show that $(f, g) : \Omega \to \mathbb{R}$ is measurable, where $(f, g)(\omega) = (f(\omega), g(\omega))$. To see that (f, g) is measurable, look at inverse images of sets that generate $\mathcal{B}^2$, namely sets of the form $(-\infty, a] \times (-\infty, b]$, and apply Lemma 60. We see that

$$(f, g)^{-1}((-\infty, a] \times (-\infty, b]) = f^{-1}((-\infty, a]) \cap g^{-1}((-\infty, b]),$$

which is measurable. Hence, (f, g) is measurable and $h(f, g)$ is measurable by part 2. Similarly fg is measurable. $\square$

You can also prove that f/g is measurable when the ratio is defined to be an arbitrary constant when $g = 0$. Similarly, part 3 can be extended to extended real-valued functions so long as care is taken to handle cases of $\infty - \infty$ and $\infty \times 0$.

36-752: Lecture 6

THEOREM 72. *Let $f_n : \Omega \to I\!R$ be measurable for all n. Then the following are measurable:*

1. $\limsup_{n\to\infty} f_n$,
2. $\liminf_{n\to\infty} f_n$,
3. $\{\omega : \lim_{n\to\infty} f_n \text{ exists}\}$.
4. $f = \begin{cases} \lim_{n\to\infty} f_n & \text{where the limit exists,} \\ 0 & \text{elsewhere.} \end{cases}$

EXERCISE 73. Prove Lemma 72.

Random Variables and Induced Measures.

EXAMPLE 74. Let $\Omega = (0, 1)$ with the Borel σ-field, and let μ be Lebesgue measure, a probability. Let $Z_0(\omega) = \omega$. For $n \geq 1$, define $X_n(\omega) = \lfloor 2Z_{n-1}(\omega) \rfloor$ and $Z_n(\omega) = 2Z_{n-1}(\omega) - X_n(\omega)$. All X_n's and Z_n's are random variables. Each X_n takes only two values, 0 and 1. It is easy to see that $\mu(\{\omega : X_n(\omega) = 1\}) = 1/2$. It is also easy to see that $\mu(\{\omega : Z_n(\omega) \leq c\}) = c$ for $0 \leq c \leq 1$.

Each measurable function from a measure space to another measurable space induces a measure on its range space.

LEMMA 75. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $(S, \mathcal{A})$ be a measurable space. Let $f : \Omega \to S$ be a measurable function. Then f induces a measure on $(S, \mathcal{A})$ defined by $\nu(A) = \mu(f^{-1}(A))$ for each $A \in \mathcal{A}$.*

PROOF. Clearly, $\nu \geq 0$ and $\nu(\varnothing) = 0$. Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of disjoint elements of $\mathcal{A}$. Then

$$\begin{aligned} \nu\left(\bigcup_{n=1}^{\infty} A_n\right) &= \mu\left(f^{-1}\left[\bigcup_{n=1}^{\infty} A_n\right]\right) \\ &= \mu\left(\bigcup_{n=1}^{\infty} f^{-1}[A_n]\right) \\ &= \sum_{n=1}^{\infty} \mu(f^{-1}[A_n]) \\ &= \sum_{n=1}^{\infty} \nu(A_n). \quad \square \end{aligned}$$

The measure ν in Lemma 75 is called the *measure induced on $(S, \mathcal{A})$ from μ by f*. This measure is only interesting in special cases. First, if μ is a probability then so is ν.

Definition 76. Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $(S, \mathcal{A})$ be a measurable space. Let $X : \Omega \to S$ be a random quantity. Then the measure induced on $(S, \mathcal{A})$ from P by X is called the *distribution of X*.

We typically denote the distribution of X by μ_X. In this case, μ_X is a measure on the space $(S, \mathcal{A})$.

Example 77. Consider the random variables in Example 74. The distribution of each X_n is the Bernoulli distribution with parameter $1/2$. The distribution of each Z_n is the uniform distribution on the interval $(0, 1)$. These were each computed in Example 74.

If μ is infinite and f is not one-to-one, then the induced measure may be of no interest at all.

Exercise 78. Either prove or create a counterexample to the following conjecture: If μ is a σ-finite on some measurable space $(\Omega, \mathcal{F})$, then for any measurable function f from Ω to S, the induced measure is also σ-finite.

Example 79. (Jacobians) If $\Omega = S = \mathbb{R}^k$ and f is one-to-one with a differentiable inverse, then ν is the measure you get from the usual change-of-variables formula using Jacobians.

We have just seen how to construct the distribution from a random variable. Oddly enough, the opposite construction is also available. First notice that every probability ν on $(\mathbb{R}, \mathcal{B}^1)$ has a distribution function F defined by $F(x) = \nu((-\infty, x])$. Now, we can construct a probability space $(\Omega, \mathcal{F}, P)$ and a random variable $X : \Omega \to \mathbb{R}$ such that $\nu = P(X^{-1})$.[1] Indeed, just let $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}^1$, $P = \nu$, and $X(\omega) = \omega$.

Integration. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. The definition of integral is done in three stages. We start with simple functions.

Definition 80. Let $f : \Omega \to \overline{\mathbb{R}}^{+0}$ be a simple function with canonical representation $f(\omega) = \sum_{i=1}^n a_i I_{A_i}(\omega)$ The *integral of f with respect to μ* is defined to be $\sum_{i=1}^n a_i \mu(A_i)$. The integral is denoted variously as $\int f d\mu$, $\int f(\omega)\mu(d\omega)$, or $\int f(\omega) d\mu(\omega)$.

The values $\pm\infty$ are allowed for an integral.

We use the following convention whenever necessary in defining an integral: $\pm\infty \times 0 = 0$. This applies to both the case when the function is 0 on a set of infinite measure and when the function is infinite on a set of 0 measure.

Proposition 81. *If $f \leq g$ and both are nonegative and simple, then $\int f d\mu \leq \int g d\mu$.*

Definition 82. We say that f is *integrable with respect to μ* if $\int f d\mu$ is finite.

Example 83. A real-valued simple function is always integrable with respect to a finite measure.

[1] **Notation:** When X is a random quantity and B is a set in the space where X takes its values, we use the following two symbols interchangeably: $X^{-1}(B)$ and $X \in B$. Both of these stand for $\{\omega : X(\omega) \in B\}$. Finally, for all B,

$$\mu_X(B) = \Pr(X \in B) = P(X^{-1}(B)).$$

36-752: Lecture 7

The second step in the definition of integral is to consider nonnegative measurable functions.

DEFINITION 84. For nonnegative measurable f, define the *integral of f with respect to μ* by

$$\int f d\mu = \sup_{\text{nonnegative finite simple } g \leq f} \int g d\mu.$$

That is, if f is nonnegative and measurable, $\int f d\mu$ is the least upper bound (possibly infinite) of the integrals of nonnegative finite simple functions $g \leq f$. Proposition 81 helps to show that Definition 80 is a special case of Definition 84, so the two definitions do not conflict when they both apply.

Finally, for arbitrary measurable f, we first split f into its positive and negative parts, $f = f^+ - f^-$.

DEFINITION 85. Let f be measurable. If either f^+ or f^- is integrable with respect to μ, we define the *integral of f with respect to μ* to be $\int f^+ d\mu - \int f^- d\mu$, otherwise the integral *does not exist*.

It is easy to see that Definition 84 is a special case of Definition 85, so the two definitions do not conflict when they both apply. The reason for splitting things up this way is to avoid ever having to deal with $\infty - \infty$.

One unfortunate consequence of this three-part definition is that many theorems about integrals must be proven in three steps. One fortunate consequence is that, for most of these theorems, at least some of the three steps are relatively straightforward.

DEFINITION 86. If $A \in \mathcal{F}$, we define $\int_A f d\mu$ by $\int I_A f d\mu$.

PROPOSITION 87. *If $f \leq g$ and both integrals are defined, then $\int f d\mu \leq \int g d\mu$.*

EXAMPLE 88. Let μ be counting measure on a set Ω. (This measure is not σ-finite unless Ω is countable.) If $A \subseteq \Omega$, then $\mu(A) = \#(A)$, the number of elements in A. If f is a nonnegative simple function, $f = \sum_{i=1}^n a_i I_{A_i}$, then

$$\int f d\mu = \sum_{i=1}^n a_i \#(A_i) = \sum_{\text{All } \omega} f(\omega).$$

It is not difficult to see that the equality of the first and last terms above continues to hold for all nonnegative functions, and hence for all integrable functions.

Before we study integration in detail, we should note that integration with respect to Lebesgue measure is the same as the Riemann integral in many cases.

THEOREM 89. *Let f be a continuous function on a closed bounded interval $[a, b]$. Let μ be Lebesgue measure. Then the Riemann integral $\int_a^b f(x)dx$ equals $\int_{[a,b]} f d\mu$.*

Exercise 90. Prove Theorem 89.

Example 91. A case in which the Riemann integral differs from the Lebesgue integral is that of "improper" Riemann integrals. These are defined as limits of Riemann integrals that are each defined in the usual way. For example, integrals of unbounded functions and integrals over unbounded regions cannot be defined in the usual way because the Riemann sums would always be ∞ or undefined. Consider the function $f(x) = \sin(x)/x$ over the interval $[1, \infty)$. It is not difficult to see that neither f^+ nor f^- is integrable with respect to Lebesgue measure. Hence, the integral that we have defined here does not exist. However, the improper Riemann integral is defined as $\lim_{T\to\infty} \int_1^T f(x)dx$, if the limit exists. In this case, the limit exists.

Some simple properties of integrals include the following:

- For c a constant, $\int cf d\mu = c \int f d\mu$ if the latter exists.
- If $f \geq 0$, then $\int f d\mu \geq 0$.
- If f is extended real-valued, then $\left|\int f d\mu\right| < \infty$ only if $\mu(f^{-1}(\{\pm\infty\})) = 0$.
- if $f = g$ a.e. $[\mu]$ and if either $\int f d\mu$ or $\int g d\mu$ exists, then so does the other, and they are equal. Similarly, if one of the integrals doesn't exist, then neither does the other.

Definition 92. If P is a probability and X is a random variable, then $\int X dP$ is called the *mean* of X, *expected value* of X, or *expectation* of X and denoted $\mathrm{E}(X)$. If $\mathrm{E}(X) = \mu$ is finite, then the *variance* of X is $\mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2]$.

The mean and variance of a random variable have an interesting relation to the tail of the distribution.

Proposition 93. (Markov inequality) *Let X be a nonnegative random variable. Then* $\Pr(X \geq c) \leq \mathrm{E}(X)/c$.

There is also a famous corollary.

Corollary 94. (Tchebychev inequality) *Let X have finite mean μ. Then* $\Pr(|X - \mu| \geq c) \leq \mathrm{Var}(X)/c^2$.

Exercise 95. Show that there is some random variable X for which $\Pr(|X - \mu| \geq c) = \mathrm{Var}(X)/c^2$. Thus, without additional assumptions, Tchebychev's inequality cannot be improved.

We would like to be able to prove that $\int (f+g)d\mu = \int f d\mu + \int g d\mu$ whenever at least two of them are finite. We could prove this for nonnegative simple functions now, but not in general.

Proposition 96. *Let f and g be nonnegative simple functions defined on a measure space $(\Omega, \mathcal{F}, \mu)$. Then $\int (f+g)d\mu = \int f d\mu + \int g d\mu$.*

The proof for general functions requires some limit theorems first.

36-752: Lecture 8

One of the famous limit theorems is the following.

THEOREM 97. (FATOU'S LEMMA) *Let $\{f_n\}_{n=1}^\infty$ be a sequence of nonnegative measurable functions. Then*

$$\int \liminf_n f_n d\mu \le \liminf_n \int f_n d\mu.$$

The proof of Theorem 97 is given in a separate document. Here is an outline of the proof. Let $f = \liminf_n f_n$, and let ϕ be an arbitrary nonnegative simple function such that $\phi \le f$. We need to show that $\int \phi d\mu \le \liminf_n \int f_n d\mu$. The set $\{\omega : \phi(\omega) > 0\}$ can be written as the union of the sets

$$A_n = \{\omega : f_k(\omega) > (1-\epsilon)\phi(\omega), \text{ for all } k \ge n\}.$$

For each n, $\int f_n d\mu \ge (1-\epsilon)\int_{A_n} \phi d\mu$. The liminf of the right sides can be shown to equal $(1-\epsilon)\int \phi d\mu$. Since $\liminf_n \int f_n d\mu \ge (1-\epsilon)\int \phi d\mu$ for all $\epsilon > 0$, we have what we need.

The first of the two most useful limit theorems is the following.

THEOREM 98. (MONOTONE CONVERGENCE THEOREM) *Let $\{f_n\}_{n=1}^\infty$ be a sequence of measurable nonnegative functions, and let f be a measurable function such that $f_n \le f$ a.e. $[\mu]$ and $\lim_{n\to\infty} f_n = f(x)$ a.e. $[\mu]$. Then,*

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

PROOF. Since $f_n \le f$ for all n, $\int f_n d\mu \le \int f d\mu$ for all n. Hence

$$\liminf_{n\to\infty} \int f_n d\mu \le \limsup_{n\to\infty} \int f_n d\mu \le \int f d\mu.$$

By Fatou's lemma, $\int f d\mu \le \liminf_{n\to\infty} \int f_n d\mu$. $\Box$

EXERCISE 99. Why is it called the "monotone" convergence theorem?

We are now in a position to prove that the integral of the sum is the sum of the integrals.

THEOREM 100. *If $\int f d\mu$ and $\int g d\mu$ are defined and they are not both infinite and of opposite signs, then $\int [f+g] d\mu = \int f d\mu + \int g d\mu$.*

PROOF. If $f, g \ge 0$, then by Lemma 69, there exist sequences of nonnegative simple functions $\{f_n\}_{n=1}^\infty$ and $\{g_n\}_{n=1}^\infty$ such that $f_n \uparrow f$ and $g_n \uparrow g$. Then $(f_n + g_n) \uparrow (f+g)$ and $\int [f_n + g_n] d\mu = \int f_n d\mu + \int g_n d\mu$ by Proposition 96. The result now follows from the monotone convergence theorem. For integrable f and g, note that $(f+g)^+ + f^- + g^- = (f+g)^- + f^+ + g^+$. What we just proved for nonnegative functions implies that

$$\begin{aligned} &\int (f+g)^+ d\mu + \int f^- d\mu + \int g^- d\mu \\ &= \int [(f+g)^+ + f^- + g^-] d\mu \\ &= \int [(f+g)^- + f^+ + g^+] d\mu \\ &= \int (f+g)^- d\mu + \int f^+ d\mu + \int g^+ d\mu. \end{aligned}$$

Rearranging the terms in the first and last expressions gives the desired result. If both f and g have infinite integral of the same sign, then it follows easily that $f+g$ has infinite integral of the same sign. Finally, if only one of f and g has infinite integral, it also follows easily that $f+g$ has infinite integral of the same sign. □

For proving theorems about integrals, there is a common sequence of steps that is often called the *standard machinery* or *standard machine*. It is illustrated in the next result, the measure-theoretic version of the change-of-variables formula.

LEMMA 101. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $(S, \mathcal{A})$ be a measurable space. Let $f : \Omega \to S$ be a measurable function. Let ν be the measure induced on $(S, \mathcal{A})$ by f from μ. (See Definition 76.) Let $g : S \to I\!R$ be $\mathcal{A}/\mathcal{B}^1$ measurable. Then*

$$\int g d\nu = \int g(f) d\mu, \tag{102}$$

if either integral exists.

PROOF. First, assume that $g = I_A$ for some $A \in \mathcal{A}$. Then (102) becomes $\nu(A) = \mu(f^{-1}(A))$, which is the definition of ν. Next, if g is a nonnegative simple function, then (102) holds by linearity of integrals. If g is a nonnegative function, then use the monotone convergence theorem and a sequence of nonnegative simple functions converging to g from below to see that (102) holds. Finally, for general g, (102) holds if either g^+ or g^- is integrable. □

EXERCISE 103. Suppose that f_n is integrable for each n and $\sup_n \int f_n d\mu < \infty$. Show that, if $f_n \uparrow f$, then f is integrable and $\int f_n d\mu \to \int f d\mu$.

EXERCISE 104. Show that if f and g are integrable, then

$$\left| \int f d\mu - \int g d\mu \right| \leq \int |f - g| \, d\mu.$$

EXERCISE 105. Assume the sequence of functions f_n is defined on a measure space $(\Omega, \mathcal{F}, \mu)$ such that $\mu(\Omega) < \infty$. Further, suppose that the f_n are uniformly bounded and that $f_n \to f$ uniformly. Show that $\int f_n d\mu \to \int f d\mu$.

Fatou's Lemma

Theorem 97. (Fatou's lemma) *Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of nonnegative measurable functions. Then*

$$\int \liminf_n f_n d\mu \le \liminf_n \int f_n d\mu.$$

Proof. Let $f(\omega) = \liminf_{n\to\infty} f_n(\omega)$. Because

$$\int f d\mu = \sup_{\text{finite simple } \phi \le f} \int \phi d\mu,$$

we need only prove that, for every finite simple $\phi \le f$,

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

Let $\phi \le f$ be finite and simple, and let $\epsilon > 0$. For each n, define

$$A_n = \{\omega \in A : f_k(\omega) \ge (1-\epsilon)\phi(\omega), \text{ for all } k \ge n\}.$$

Since $(1-\epsilon)\phi(\omega) \le f(\omega)$ for all ω with strict inequality wherever either side is positive, $\bigcup_{n=1}^{\infty} A_n = \Omega$ and $A_n \subseteq A_{n+1}$ for all n. Let $B_n = A \cap A_n^C$.

$$\int f_n d\mu \ge \int_{A_n} f_n d\mu \ge (1-\epsilon) \int_{A_n} \phi d\mu. \tag{106}$$

Let the canonical representation of ϕ be $\sum_{i=1}^{m} c_i I_{C_i}$. Then, for all n.

$$\int_{A_n} \phi d\mu = \sum_{i=1}^{m} c_i \mu(C_i \cap A_n).$$

Because the A_n's form an increasing sequence whose union is Ω, $\lim_{n\to\infty} \mu(C_i \cap A_n) = \mu(C_i)$ for all i. Taking the $\liminf_n$ of both sides of (106) yields

$$\liminf_n \int f_n d\mu \ge (1-\epsilon) \sum_{i=1}^{m} c_i \mu(C_i) = (1-\epsilon) \int \phi d\mu.$$

Since this is true for every $\epsilon > 0$,

$$\liminf_{n\to\infty} \int f_n d\mu \ge \int \phi d\mu. \quad \square$$

36-752: Lecture 9

Lemma 101 has a widely-used corollary.

COROLLARY 107. (LAW OF THE UNCONSCIOUS STATISTICIAN) *If $X : \Omega \to S$ is a random quantity with distribution μ_X and if $f : S \to I\!R$ is measurable, then* $\mathrm{E}[f(X)] = \int f d\mu_X$.

Another useful application of monotone convergence is the following.

THEOREM 108. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $f : \Omega \to \overline{I\!R}^{+0}$ be measurable. Then $\nu(A) = \int_A f d\mu$ is a measure on $(\Omega, \mathcal{F})$.*

EXERCISE 109. Prove Theorem 108.

If μ is σ-finite and if f is finite a.e. $[\mu]$, then ν in Theorem 108 is σ-finite.

What goes wrong with the conclusion to Theorem 108 if f is integrable but not necessarily nonnegative? If f can take negative values then $\nu(A) = \int_A f d\mu$ might be negative. Let $A = \{\omega : f(\omega) < 0\}$. Suppose that $\mu(A) > 0$. Write $A = \bigcup_{n=1}^{\infty} A_n$, where $A_n = \{\omega : f(\omega) < -1/n\}$. If $\mu(A) > 0$, then there exists n such that $\mu(A_n) > 0$. (This argument is used often in proving probability results.) Then

$$-\nu(A) = \int I_A(-f) d\mu \geq \int I_{A_n}(-f) d\mu \geq \frac{1}{n}\mu(A_n) > 0.$$

Here is another application of the standard machinery.

THEOREM 110. *Assume the same conditions as Theorem 108. Integrals with respect to ν can be computed as $\int g d\nu = \int g f d\mu$, if either exists.*

PROOF. We prove the result in four stages. First, assume that g is a indicator I_A of some set $A \in \mathcal{F}$. Then the definition of ν says that $\int g d\nu = \nu(A) = \int I_A f d\mu$. Second, assume that g is a nonnegative simple function. The result holds for g by linearity of integrals. Third, assume that g is nonnegative. Approximate g from below by nonnegative simple functions $\{g_n\}_{n=1}^{\infty}$. Then $\int g_n d\nu = \int g_n f d\mu$ for each n and the monotone convergence theorem says that the left side converges to $\int g d\nu$ and the right side converges to $\int g f d\mu$. Finally, if g is measurable, write $g = g^+ - g^-$ (the positive and negative parts). Then $\int g^+ d\nu = \int g^+ f d\mu$ and $\int g^- d\nu = \int g^- f d\mu$. We see that $\int g d\nu$ exists if and only if $\int g f d\mu$ exists, and if either exists they are equal. $\square$

The standard machinery corresponds to the three stages in defining integrals. The first stage is split into indicators and nonnegative simple functions to make four steps in the standard machinery.

DEFINITION 111. The function f in Theorem 108 is called the *density of ν with respect to μ.*

EXAMPLE 112. (PROBABILITY DENSITY FUNCTIONS) Consider a continuous random variable X having a density f. That is,

$$\Pr(X \leq a) = \int_{-\infty}^{a} f(x)dx.$$

Then the distribution of X, defined by $\mu_X(B) = \Pr(X \in B)$ for $B \in \mathcal{B}^1$, satisfies

$$\mu_X(B) = \int_B f d\lambda,$$

where λ is Lebesgue measure. That is, the probability density functions of the usual continuous distributions that you learned about in earlier courses are also densities with respect to Lebesgue measure in the sense defined above.

EXAMPLE 113. (PROBABILITY MASS FUNCTIONS) Consider a typical discrete random variable X with mass function f, i.e., $f(x) = \Pr(X = x)$ for all x. There are at most countably many x such that $f(x) > 0$. Let μ_X be the distribution of X. For each set B, we know that

$$\mu_X(B) = \Pr(X \in B) = \sum_{x \in B} f(x).$$

The rightmost term in this equation is $\int f d\mu$, where μ is counting measure on the range space of X. So, f is the density of μ_X with respect to μ.

The other major limit theorem is the following.

THEOREM 114. (DOMINATED CONVERGENCE THEOREM) *Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of measurable functions, and let f and g be measurable functions such that $f_n \to f$ a.e. $[\mu]$, $|f_n| \leq g$ a.e. $[\mu]$, and $\int g d\mu < \infty$. Then,*

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

PROOF. We have $-g \leq f_n \leq g$ a.e. $[\mu]$, hence

$$\begin{aligned}
g + f_n &\geq 0, \quad \text{a.e. } [\mu], \\
g - f_n &\geq 0, \quad \text{a.e. } [\mu], \\
\lim_{n\to\infty} [g + f_n] &= g + f \quad \text{a.e. } [\mu], \\
\lim_{n\to\infty} [g - f_n] &= g - f \quad \text{a.e. } [\mu].
\end{aligned}$$

It follows from Fatou's lemma and Theorem 100 that

$$\begin{aligned}
\int [g + f] d\mu &\leq \liminf_{n\to\infty} \int [g + f_n] d\mu \\
&= \int g d\mu + \liminf_{n\to\infty} \int f_n d\mu, \\
\int f d\mu &\leq \liminf_{n\to\infty} \int f_n d\mu.
\end{aligned}$$

Similarly, it follows that

$$
\begin{aligned}
\int [g-f]d\mu &\leq \liminf_{n\to\infty} \int [g-f_n]d\mu \\
&= \int g d\mu - \limsup_{n\to\infty} \int f_n d\mu, \\
\int f d\mu &\geq \limsup_{n\to\infty} \int f_n d\mu.
\end{aligned}
$$

Together, these imply the conclusion of the theorem. □

Example 115. Let μ be a finite measure. Then limits and integrals can be interchanged whenever the functions in the sequence are uniformly bounded.

An alternate version of the dominated convergence theorem is the following.

Proposition 116. *Let* $\{f_n\}_{n=1}^{\infty}$, $\{g_n\}_{n=1}^{\infty}$ *be sequences of measurable functions such that* $|f_n| \leq g_n$, *a.e.* $[\mu]$. *Let* f *and* g *be measurable functions such that* $\lim_{n\to\infty} f_n = f$ *and* $\lim_{n\to\infty} g_n = g$, *a.e.* $[\mu]$. *Suppose that* $\lim_{n\to\infty} \int g_n d\mu = \int g d\mu < \infty$. *Then,* $\lim_{n\to\infty} \int f_n d\mu = \int f d\mu$.

The proof is the same as the proof of Theorem 114, except that g_n replaces g in the first three lines and wherever g appears with f_n and a limit is being taken.

For finite measure spaces (i.e. $(\Omega, \mathcal{F}, \mu)$ with $\mu(\Omega) < \infty$), the minimal condition that guarantees convergence of integrals is *uniform integrability* .

Definition 117. A sequence of integrable functions $\{f_n\}_{n=1}^{\infty}$ is *uniformly integrable (with respect to* μ*)* if $\lim_{c\to\infty} \sup_n \int_{\{\omega:|f_n(\omega)|>c\}} |f_n| d\mu = 0$.

Theorem 118. *Let* μ *be a finite measure. Let* $\{f_n\}_{n=1}^{\infty}$ *be a sequence of integrable functions such that* $\lim_{n\to\infty} f_n = f$ *a.e.* $[\mu]$. *Suppose that* $\{f_n\}_{n=1}^{\infty}$ *is uniformly integrable. Then* $\lim_{n\to\infty} \int f_n d\mu = \int f d\mu$.

If the f_n's in Theorem 118 are nonnegative and integrable and $f_n \to f$, then $\lim_{n\to\infty} \int f_n d\mu = \int f d\mu$ implies that $\{f_n\}_{n=1}^{\infty}$ are uniformly integrable. We will not use this result, however.

36-752: Lecture 10

Here are some more useful properties of integrals.

THEOREM 119. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let f and g be measurable extended real-valued functions.*

1. *If f is nonnegative and $\mu(\{\omega : f(\omega) > 0\}) > 0$, then $\int f d\mu > 0$.*
2. *If f and g are integrable and if $\int_A f d\mu = \int_A g d\mu$ for all $A \in \mathcal{F}$, then $f = g$ a.e. $[\mu]$.*
3. *If μ is σ-finite and if $\int_A f d\mu = \int_A g d\mu$ for all $A \in \mathcal{F}$, then $f = g$ a.e. $[\mu]$.*
4. *Let Π be a π-system that generates $\mathcal{F}$. Suppose that Ω is a finite or countable union of elements of Π. If f and g are integrable and if $\int_A f d\mu = \int_A g d\mu$ for all $A \in \Pi$, then $f = g$ a.e. $[\mu]$.*

PROOF.

1. Let $A_c = \{\omega : f(\omega) > c\}$ for each $c \geq 0$. Because $\mu(A_0) > 0$ and $A_0 = \bigcup_{n=1}^{\infty} A_{1/n}$, it follows from Lemma 34 that there exists n such that $\mu(A_{1/n}) > 0$. Since $f \geq f I_{A_{1/n}}$, we have $\int f d\mu \geq \int_{A_{1/n}} f d\mu$. But $(1/n)I_{A_{1/n}}$ is a simple function that is $\leq f I_{A_{1/n}}$ and $\int (1/n) I_{A_{1/n}} d\mu = \mu(A_{1/n}) > 0$. It follows that $\int f d\mu > 0$.
2. This will appear on a homework assignment.
3. First, assume that f and g are real-valued. Let $\{A_n\}_{n=1}^{\infty}$ be disjoint elements of $\mathcal{F}$ such that $\mu(A_n) < \infty$ and $\bigcup_{n=1}^{\infty} A_n = \Omega$. Let $B_m = \{\omega : |f(\omega)| < m, |g(\omega)| < m\}$ for each integer m. For each pair (n, m), $f I_{A_n \cap B_m}$ and $g I_{A_n \cap B_m}$ satisfy the conditions of the previous part, so $f I_{A_n \cap B_m} = g I_{A_n \cap B_m}$ a.e. $[\mu]$. Let $C = \{\omega : f(\omega) \neq g(\omega)\}$. Since
$$C = \bigcup_{n=1}^{\infty} \bigcup_{m=1}^{\infty} [C \cap B_m \cap A_n],$$
and each $\mu(C \cap B_m \cap A_n) = 0$, it follows that $\mu(C) = 0$.

   Next, suppose that f and/or g is extended real-valued. Let $E = \{f = \infty\}\Delta\{g = \infty\}$, the set where one function is ∞ but the other is not. If $\mu(E) > 0$, then there is a subset A of E such that $0 < \mu(A) < \infty$ and one of the functions is bounded above on A while the other is infinite. This contradicts $\int_A f d\mu = \int_A g d\mu$. A similar result holds for $-\infty$.
4. Define $\nu_1^+(A) = \int_A f^+ d\mu$, $\nu_2^+(A) = \int_A g^+ d\mu$, $\nu_1^-(A) = \int_A f^- d\mu$, and $\nu_2^-(A) = \int_A g^- d\mu$. These are all finite measures according to Theorem 108. The additional condition implies that they are all σ-finite on Π. The equality of the integrals implies that $\nu_1^+ + \nu_2^- = \nu_1^- + \nu_2^+$ for all sets in Π. Theorem 43 implies that $\nu_1^+ + \nu_2^- = \nu_1^- + \nu_2^+$ for all sets in $\mathcal{F}$. Hence, the condition of part 2 hold and the result is proven. □

The condition about unions in part 4 of the above theorem holds for the π-systems in Example 38.

Corollary 120. *If μ is σ-finite and ν is related to μ as in Theorem 108, then the density of ν with respect to μ is unique, a.e. $[\mu]$.*

There is an interesting characterization of σ-finite measures in terms of integrals.

Theorem 121. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Then μ is σ-finite if and only if there exists a strictly positive integrable function.*

Exercise 122. Prove Theorem 121.

Absolute Continuity. There is a special relationship between measures on the same space that is very useful in Probability theory.

Definition 123. Let ν and μ be measures on the space $(\Omega, \mathcal{F})$. We say that $\nu \ll \mu$ (read ν *is absolutely continuous with respect to* μ) if for every $A \in \mathcal{F}$, $\mu(A) = 0$ implies $\nu(A) = 0$.

That is, $\nu \ll \mu$ if and only if every measure 0 set under μ is also a measure 0 set under ν.

Example 124. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let f be a nonnegative function, and define $\nu(A) = \int_A f d\mu$. Then ν is a measure and $\nu \ll \mu$. If $f < \infty$ a.e. $[\mu]$ and if μ is σ-finite, then ν is σ-finite as well.

Example 125. Let μ_1 and μ_2 be measures on the same space. Let $\mu = \mu_1 + \mu_2$. Then $\mu_i \ll \mu$ for $i = 1, 2$.

Absolute continuity has a connection with continuity of functions.

Proposition 126. *Let ν and μ be measures on the space $(\Omega, \mathcal{F})$. Suppose that, for every $\epsilon > 0$, there exists δ such that for every $A \in \mathcal{F}$, $\mu(A) < \delta$ implies $\nu(A) < \epsilon$. Then $\nu \ll \mu$.*

A concept related to absolute continuity is singularity.

Definition 127. Two measures μ and ν on the same space $(\Omega, \mathcal{F})$ are *(mutually) singular* (denoted $\mu \perp \nu$) if there exist disjoint sets S_μ and S_ν such that $\mu(S_\mu^C) = \nu(S_\nu^C) = 0$.

Example 128. Let f and g be nonnegative functions such that $fg = 0$ a.e. $[\mu]$. Define $\nu_1(A) = \int_A f d\mu$ and $\nu_2(A) = \int_A g d\mu$. Then $\nu_1 \perp \nu_2$.

The main theoretical result on absolute continuity is the Radon-Nikodym theorem which says that, in the σ-finite case, all absolute continuity is of the type in Example 124.

Theorem 129. (Radon-Nikodym) *Let μ and ν be σ-finite measures on the space $(\Omega, \mathcal{F})$. Then $\nu \ll \mu$ if and only if there exists a nonnegative measurable f such that $\nu(A) = \int_A f d\mu$ for all $A \in \mathcal{F}$. The function f is unique a.e. $[\mu]$.*

One proof of this result is given in a separate course document. Another proof is given later after we introduce conditional expectation.

Definition 130. The function f in Theorem 129 is called a *Radon-Nikodym derivative of ν with respect to μ*. It is denoted $d\nu/d\mu$. Each such function is called a *version* of $d\nu/d\mu$.

36-752: Lecture 11

The uniqueness of Radon-Nikodym derivatives is only a.e. $[\mu]$. If $f = d\nu/d\mu$, then every measurable function that equals f a.e. $[\mu]$ could also be called $d\nu/d\mu$. All of these functions are called *versions* of the Radon-Nikodym derivative.

Definition 131. If $\mu \ll \nu$ and $\nu \ll \mu$, we say that μ and ν are *equivalent.*

If μ and ν are equivalent, then

$$\frac{d\mu}{d\nu} = \frac{1}{\frac{d\nu}{d\mu}}.$$

If $\nu \ll \mu \ll \eta$, then the chain rule for R-N derivatives says

$$\frac{d\nu}{d\eta} = \frac{d\nu}{d\mu}\frac{d\mu}{d\eta}.$$

Absolute continuity plays an important role in statistical inference. Parametric families are collections of probability measures that are all absolutely continuous with respect to a single measure.

Theorem 132. *Let $(\Omega, \mathcal{F}, \mu)$ be a σ-finite measure space. Let $\{\mu_\theta : \theta \in \Theta\}$ be a collection of measures on $(\Omega, \mathcal{F})$ such that $\mu_\theta \ll \mu$ for all $\theta \in \Theta$. Then there exists a sequence of nonnegative numbers $\{c_n\}_{n=1}^\infty$ and a sequence of elements $\{\theta_n\}_{n=1}^\infty$ of Θ such that $\sum_{n=1}^\infty c_n$ and $\mu_\theta \ll \sum_{n=1}^\infty c_n \mu_{\theta_n}$ for all $\theta \in \Theta$.*

We will not prove this theorem here. (See Theorem A.78 in Schervish 1995.)

Random Vectors. In Definition 62 we defined random variables and random quantities. A special case of the latter and generalization of the former is a random vector.

Definition 133. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X : \Omega \to \mathbb{R}^k$ be a measurable function. Then X is called a *random vector* .

There arises, in this definition, the question of what σ-field of subsets of $\mathbb{R}^k$ should be used. When left unstated, we always assume that the σ-field of subsets of a multidimensional real space is the Borel σ-field, namely the smallest σ-field containing the open sets. However, because $\mathbb{R}^k$ is also a product set of k sets, each of which already has a natural σ-field associated with it, we might try to use a σ-field that corresponds to that product in some way.

Product Spaces. The set $\mathbb{R}^k$ has a topology in its own right, but it also happens to be a product set. Each of the factors in the product comes with its own σ-field. There is a way of constructing σ-field's of subsets of product sets directly without appealing to any additional structure that they might have.

DEFINITION 134. Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be measurable spaces. Let $\mathcal{F}_1 \otimes \mathcal{F}_2$ be the smallest σ-field of subsets of $\Omega_1 \times \Omega_2$ containing all sets of the form $A_1 \times A_2$ where $A_i \in \mathcal{F}_i$ for $i = 1, 2$. Then $\mathcal{F}_1 \otimes \mathcal{F}_2$ is the *product σ-field.*

LEMMA 135. *Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be measurable spaces. Suppose that $\mathcal{C}_i$ is a π-system that generates $\mathcal{F}_i$ for $i = 1, 2$. Let $\mathcal{C} = \{C_1 \times C_2 : C_1 \in \mathcal{C}_1, C_2 \in \mathcal{C}_2\}$. Then $\sigma(\mathcal{C}) = \mathcal{F}_1 \otimes \mathcal{F}_2$, and $\mathcal{C}$ is a π-system.*

PROOF. Because $\sigma(\mathcal{C})$ is a σ-field, it contains all sets of the form $C_1 \times A_2$ where $A_2 \in \mathcal{F}_2$. For the same reason, it must contain all sets of the form $A_1 \times A_2$ for $A_i \in \mathcal{F}_i$ $(i = 1, 2)$. Because

$$(C_1 \times C_2) \cap (D_1 \times D_2) = (C_1 \cap D_1) \times (C_2 \times D_2),$$

we see that $\mathcal{C}$ is a π-system. $\Box$

EXAMPLE 136. Let $\Omega_i = \mathrm{I\!R}$ for $i = 1, 2$, and let $\mathcal{F}_1$ and $\mathcal{F}_2$ both be $\mathcal{B}^1$. Let $\mathcal{C}_i$ be the collection of all intervals centered at rational numbers with rational lengths. Then $\mathcal{C}_i$ generates $\mathcal{F}_i$ for $i = 1, 2$ and the product topology is the smallest topology containing $\mathcal{C}$ as defined in Lemma 135. It follows that $\mathcal{F}_1 \otimes \mathcal{F}_2$ is the smallest σ-field containing the product topology. We call this σ-field $\mathcal{B}^2$.

EXAMPLE 137. This time, let $\Omega_1 = \mathrm{I\!R}^2$ and $\Omega_2 = \mathrm{I\!R}$. The product set is $\mathrm{I\!R}^3$ and the product σ-field is called $\mathcal{B}^3$. It is also the smallest σ-field containing all open sets in $\mathrm{I\!R}^3$. The same idea extends to each finite-dimensional Euclidean space, with Borel σ-field's $\mathcal{B}^k$, for $k = 1, 2, \ldots$.

LEMMA 138. *Let $(\Omega_i, \mathcal{F}_i)$ and $(S_i, \mathcal{A}_i)$ be measurable spaces for $i = 1, 2$. Let $f_i : \Omega_i \to S_i$ be a function for $i = 1, 2$. Define $g(\omega_1, \omega_2) = (f_1(\omega_1), f_2(\omega_2))$, which is a function from $\Omega_1 \times \Omega_2$ to $S_1 \times S_2$. Then f_i is $\mathcal{F}_i/\mathcal{A}_i$-measurable for $i = 1, 2$ if and only if g is $\mathcal{F}_1 \otimes \mathcal{F}_2/\mathcal{A}_1 \otimes \mathcal{A}_2$-measurable.*

PROOF. For the "only if" direction, assume that each f_i is measurable. It suffices to show that for each product set $A_1 \times A_2$ (with $A_i \in \mathcal{A}_i$ for $i = 1, 2$) $g^{-1}(A_1 \times A_2) \in \mathcal{F}_1 \otimes \mathcal{F}_2$. But, it is easy to see that $g^{-1}(A_1 \times A_2) = f_1^{-1}(A_1) \times f_2^{-1}(A_2) \in \mathcal{F}_1 \otimes \mathcal{F}_2$.

For the "if" direction, suppose that g is measurable. Then for every $A_1 \in \mathcal{A}_1$, $g^{-1}(A_1 \times S_2) \in \mathcal{F}_1 \otimes \mathcal{F}_2$. But $g^{-1}(A_1 \times S_2) = f_1^{-1}(A_1) \times \Omega_2$. The fact that $f_1^{-1}(A_1) \in \mathcal{F}_1$ will now follow from the first claim in Proposition 140. (Sorry for the forward reference.) So f_1 is measurable. Similarly, f_2 is measurable. $\Box$

PROPOSITION 139. *Let $(\Omega, \mathcal{F})$, $(S_1, \mathcal{A}_1)$, and $(S_2, \mathcal{A}_2)$ be measurable spaces. Let $X_i : \Omega \to S_i$ for $i = 1, 2$. Define $X = (X_1, X_2)$ a function from Ω to $S_1 \times S_2$. Then X_i is $\mathcal{F}/\mathcal{A}_i$ measurable for $i = 1, 2$ if and only if X is $\mathcal{F}/\mathcal{A}_1 \otimes \mathcal{A}_2$ measurable.*

Lemma 138 and Proposition 139 extend to higher-dimensional products as well.

The product σ-field is also the smallest σ-field such that the coordinate projection functions are measurable. The coordinate projection functions for a product set $S_1 \times S_2$ are the functions $f_i : S_1 \times S_2 \to S_i$ (for $i = 1, 2$) defined by $f_i(s_1, s_2) = s_i$ (for $i = 1, 2$).

Infinite-dimensional product spaces pose added complications that we will not consider until later in the course.

There are a number of facts about product spaces that we might take for granted.

PROPOSITION 140. *Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be measurable spaces.*

- *For each $B \in \mathcal{F}_1 \otimes \mathcal{F}_2$ and each $\omega_1 \in \Omega_1$, the ω_1-section of B, $B_{\omega_1} = \{\omega_2 : (\omega_1, \omega_2) \in B\}$ is in $\mathcal{F}_2$.*

- *If μ_2 is a σ-finite measure on $(\Omega_2, \mathcal{F}_2)$, then $\mu_2(B_{\omega_1})$ is a measurable function from Ω_1 to $\mathbb{R}$.*

- *If $f : \Omega_1 \times \Omega_2 \to S$ is measurable, then for every $\omega_1 \in \Omega_1$, the function $f_{\omega_1} : \Omega_2 \to S$ defined by $f_{\omega_1}(\omega_2) = f(\omega_1, \omega_2)$ is measurable.*

- *If μ_2 is a σ-finite measure on $(\Omega_2, \mathcal{F}_2)$ and if $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ is nonnegative, then $\int f(\omega_1, \omega_2)\mu_2(d\omega_2)$ defines a measurable (possibly infinite valued) function of ω_1.*

To prove results like these, start with product sets or indicators of product sets and then show that the collection of sets that satisfy the results is a σ-field. Then, if necessary, proceed with the standard machinery. For example, consider the second claim. For the case of finite μ_2, the claim is true if B is a product set. It is easy to show that the collection $\mathcal{C}$ of all sets B for which $\mu_2(B_{\omega_1})$ is measurable is a λ-system. Then use Lemma 42. Here is the proof that the second claim holds for σ-finite measures once it is proven that it holds for finite measures. Let $\{A_n\}_{n=1}^{\infty}$ be elements of $\mathcal{F}_2$ that cover Ω_2 and have finite μ_2 measure. Define $\mathcal{F}_{2,n} = \{C \cap A_n : C \in \mathcal{F}_2\}$ and $\mu_{2,n}(C) = \mu_2(A_n \cap C)$ for all $C \in \mathcal{F}_2$. Then $(A_n, \mathcal{F}_{2,n}, \mu_{2,n})$ is a finite measure space for each n and $\mu_{2,n}(B_{\omega_1})$ is measurable for all n and all B in the product σ-field. Finally, notice that

$$\mu_2(B_{\omega_1}) = \sum_{n=1}^{\infty} \mu_2(B_{\omega_1} \cap A_n) = \sum_{n=1}^{\infty} \mu_{2,n}(B_{\omega_1}),$$

a sum of nonnegative measurable functions, hence measurable. The standard machinery can be used to prove the third and fourth claims. (Even though the third claim does not involve integrals, the steps in the proof are similar to those of the standard machinery.)

Radon-Nikodym Theorem

Theorem 129. (Radon-Nikodym) *Let μ and ν be σ-finite measures on the space $(\Omega, \mathcal{F})$. Then $\nu \ll \mu$ if and only if there exists a nonnegative measurable f such that $\nu(A) = \int_A f d\mu$ for all $A \in \mathcal{F}$. The function f is unique a.e. $[\mu]$.*

The proof of this result relies upon the theory of signed measures.

Definition 141. Let $(\Omega, \mathcal{F})$ be a measurable space. Let $\eta : \mathcal{F} \to \overline{\mathbb{R}}$. We call η a *signed measure* if

- $\eta(\varnothing) = 0$,
- for every sequence $\{A_k\}_{k=1}^{\infty}$ of mutually disjoint elements of $\mathcal{F}$, $\eta(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \eta(A_k)$.
- η takes at most one of the two values $\pm\infty$.

Example 142. Let μ_1 and μ_2 be measures on the same space such that at most one of them is infinite. Then $\mu_1 - \mu_2$ is a signed measure.

Example 143. Let f be integrable with respect to μ, and define $\eta(A) = \int_A f d\mu$. Then f is a finite signed measure. If the integral of f is merely defined, but not finite, then $\int_A f d\mu$ is a signed measure.

The nice thing about σ-finite signed measures is that they divide up nicely into positive and negative parts just like measurable functions.

Theorem 144. (Hahn and Jordan decompositions) *Let η be a finite signed measure on $(\Omega, \mathcal{F})$. Then there exists a set A^+ such that every subset A of A^+ has $\eta(A) \geq 0$ and every subset B of A^{+C} has $\eta(B) \leq 0$. Also, there exist finite mutually singular measures η_+ and η_- such that $\eta = \eta_+ - \eta_-$.*

Proof. Let $\alpha = \sup_{A \in \mathcal{F}} \eta(A)$. Let $\lim_{n\to\infty} \eta(A_n) = \alpha$. Although the sequence $\{\bigcup_{i=1}^{n} A_i\}_{n=1}^{\infty}$ is monotone increasing and signed measures do satisfy Lemma 1 of the course notes, $\eta\left(\bigcup_{i=1}^{n} A_i\right)$ is not necessarily as large as $\eta(A_n)$. However, the following trick replaces $\bigcup_{i=1}^{n} A_i$ by a sequence of sets whose signed measures do increase. For each n, partition Ω using the sets $A_1, \ldots, A_n$ and their complements. Let C_n be the union of all of the component sets that have positive signed measure. Since the $n+1$st partition is a refinement of the nth partition, we see that $C_{n+1} \cap C_n^C$ is a union of sets with positive signed measure and

$$\eta(A_n) \leq \eta(C_n) \leq \eta\left(C_n \bigcup C_{n+1}\right).$$

By induction, we then show that $A^+ = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} C_n$ has $\eta(A^+) = \alpha$. The conclusions now follow easily. □

Theorem 144 has an interesting consequence.

Lemma 145. *Suppose that μ and ν are finite and not mutually singular. Then there exists $\epsilon > 0$ and a set A with $\mu(A) > 0$ and $\epsilon\mu(E) \leq \nu(E)$ for every $E \subseteq A$.*

Proof. For each n, let $\eta_n = \nu - (1/n)\mu$. Let $\beta = \nu(\Omega)$. Let A_n^+ and be the set called A^+ in Theorem 144 when η is η_n. Let $M = \cap_{n=1}^{\infty} A_n^{+C}$. Since $\eta_n(E) \le 0$ for every subset of A_n^{+C}, we have $\eta_n(M) \le 0$ for all n and $\nu(M) \le (1/n)\mu(M)$. It follows that $\nu(M) = 0$ and $\nu(M^C) = \beta$. Since μ and ν are not mutually singular, $\mu(M^C) > 0$ and at least one $\mu(A_n^+) > 0$. Let $A = A_n^+$ and $\epsilon = 1/n$. □

Proof.Theorem 129 The σ-finite case follows easily from the finite case, so assume that μ and ν are finite with $\nu \ll \mu$. Let $\mathcal{G}$ be the set of all nonnegative measurable functions g such that $\int_E g d\mu \le \nu(E)$ for all $E \in \mathcal{F}$. Because $0 \in \mathcal{G}$, we know that $\mathcal{G}$ is nonempty. If g_1 and g_2 are in $\mathcal{G}$, we know that $\{g_1 \le g_2\}$ is measurable, hence it is easy to see that $\max\{g_1, g_2\} \in \mathcal{G}$. Also, if $g_n \in \mathcal{G}$ for all n and $g_n \uparrow g$, then the monotone convergence theorem implies that $g \in \mathcal{G}$. So, let $\alpha = \sup_{g \in \mathcal{G}} \int g d\mu$ and let $\lim_{n\to\infty} \int g_n d\mu = \alpha$. Let $f_n = \max\{g_1, \ldots, g_n\}$ so that there is f such that $f_n \uparrow f$, $f_n \in \mathcal{G}$ for all n, and $\lim_{n\to\infty} \int f_n d\mu = \alpha$. It follows that $\int f d\mu = \alpha$ and $f \in \mathcal{G}$. Define $\nu_1(E) = \int_E f d\mu$ and $\nu_2 = \nu - \nu_1$, which is a measure since $\nu_1 \le \nu$. If ν_2 and μ were not mutually singular, there would exist $\epsilon > 0$ and a set A with $\mu(A) > 0$ and $\epsilon\mu(E) \le \nu_2(E)$ for all $E \subseteq A$. For each $E \in \mathcal{F}$,

$$
\begin{aligned}
\int_E (f + \epsilon I_A) d\mu &= \int_E f d\mu + \epsilon\mu(E \cap A) \\
&\le \nu_1(E) + \nu_2(E \cap A) \le \nu_1(E) + \nu_2(E) = \nu(E).
\end{aligned}
$$

Hence $h = f + \epsilon I_A \in \mathcal{G}$, but $\int h d\mu = \alpha + \epsilon\mu(A) > \alpha$, a contradiction. It follows that ν_2 and μ are mutually singular. Hence, there exists S such that $\nu_2(S) = \mu(S^C) = 0$. Since $\nu \ll \mu$, we have $\nu(S^C) = 0$. Because $\nu_2 \le \nu$, we have $\nu_2(S^C) = 0$ and $\nu_2(\Omega) = 0$. It follows that $\nu = \nu_1$ and the proof of existence is complete. Uniqueness follows from Theorem 119 in the class notes. □

Notice that absolute continuity was not used in the proof until the final steps.

Definition 146. The decomposition of ν into $\nu_1 \ll \mu$ and $\nu_2 \perp \mu$ in the proof of Theorem 129 is called the *Lebesgue decomposition* of ν into an absolutely continuous part and a singular part relative to μ.

36-752: Lecture 12

THEOREM 147. *Let* $(\Omega_i, \mathcal{F}_i)$ *for* $i = 1, 2, 3$ *be measurable spaces. Let* $f : \Omega_1 \to \Omega_2$ *be a measurable onto function. Suppose that* $\mathcal{F}_3$ *contains all singletons. Let* $\mathcal{A}_1 = \sigma(f)$. *Let* $g : \Omega_1 \to \Omega_3$ *be* $\mathcal{F}_1/\mathcal{F}_3$*-measurable. Then* g *is* $\mathcal{A}_1/\mathcal{F}_3$*-measurable if and only if there exists a* $\mathcal{F}_2/\mathcal{F}_3$*-measurable* $h : \Omega_2 \to \Omega_3$ *such that* $g = h(f)$.

PROOF. For the "if" part, assume that there is a measurable $h : \Omega_2 \to \Omega_3$ such that $g(\omega) = h(f(\omega))$ for all $\omega \in \Omega_1$. Let $B \in \mathcal{F}_3$. We need to show that $g^{-1}(B) \in \mathcal{A}_1$. Since h is measurable, $h^{-1}(B) \in \mathcal{F}_2$, so $h^{-1}(B) = A$ for some $A \in \mathcal{F}_2$. Since $g^{-1}(B) = f^{-1}(h^{-1}(B))$, it follows that $g^{-1}(B) = f^{-1}(A) \in \mathcal{A}_1$.

For the "only if" part, assume that g is $\mathcal{A}_1$ measurable. For each $t \in \Omega_3$, let $C_t = g^{-1}(\{t\})$. Since g is measurable with respect to $\mathcal{A}_1 = f^{-1}(\mathcal{F}_2)$, every element of $g^{-1}(\mathcal{F}_3)$ is in $f^{-1}(\mathcal{F}_2)$. So let $A_t \in \mathcal{F}_2$ be such that $C_t = f^{-1}(A_t)$. Define $h(\omega) = t$ for all $\omega \in A_t$. (Note that if $t_1 \neq t_2$, then $A_{t_1} \cap A_{t_2} = \varnothing$, so h is well defined.) To see that $g(\omega) = h(f(\omega))$, let $g(\omega) = t$, so that $\omega \in C_t = f^{-1}(A_t)$. This means that $f(\omega) \in A_t$, which in turn implies $h(f(\omega)) = t = g(\omega)$.

To see that h is measurable, let $A \in \mathcal{F}_3$. We must show that $h^{-1}(A) \in \mathcal{F}_2$. Since g is $\mathcal{A}_1$ measurable, $g^{-1}(A) \in \mathcal{A}_1$, so there is some $B \in \mathcal{F}_2$ such that $g^{-1}(A) = f^{-1}(B)$. We will show that $h^{-1}(A) = B \in \mathcal{F}_2$ to complete the proof. If $\omega \in h^{-1}(A)$, let $t = h(\omega) \in A$ and $\omega = f(x)$ (because f is onto). Hence, $x \in C_t \subseteq g^{-1}(A) = f^{-1}(B)$, so $f(x) \in B$. Hence, $\omega \in B$. This implies that $h^{-1}(A) \subseteq B$. Lastly, if $\omega \in B$, $\omega = f(x)$ for some $x \in f^{-1}(B) = g^{-1}(A)$ and $h(\omega) = h(f(x)) = g(x) \in A$. So, $h(\omega) \in A$ and $\omega \in h^{-1}(A)$. This implies $B \subseteq h^{-1}(A)$. □
The condition that f be onto can be relaxed at the expense of changing the domain of h to be the image of f, i.e. $h : f(\Omega_1) \to \Omega_3$, with a different σ-field. The proof is slightly more complicated due to having to keep track of the image of f, which might not be a measureable set in $\mathcal{F}_2$.

The following is an example to show why the condition that $\mathcal{F}_3$ contains all singletons is included in Theorem 147.

EXAMPLE 148. Let $\Omega_i = \mathbb{R}$ for all i and let $\mathcal{F}_1 = \mathcal{F}_2 = \mathcal{B}^1$, while $\mathcal{F}_3 = \{\mathbb{R}, \varnothing\}$. Then every function $g : \Omega_1 \to \Omega_3$ is $\sigma(f)/\mathcal{F}_3$-measurable, no matter what $f : \Omega_1 \to \Omega_2$ is. For example, let $f(x) = x^2$ and $g(x) = x$ for all x. Then $g^{-1}(\mathcal{F}_3) \subseteq \sigma(f)$ but g is not a function of f.

The reason that we need the condition about singletons is the following. Suppose that there are two points $t_1, t_2 \in \Omega_3$ such that $t_1 \in A$ implies $t_2 \in A$ and vice versa for every $A \in \mathcal{F}_3$. Then there can be a set $A \in \mathcal{F}_3$ that contains both t_1 and t_2, and g can take both of the values t_1 and t_2, but f is constant on $g^{-1}(A)$ and all the measurability conditions still hold. In this case, g is not a function of f.

Product Measures. Product measures are measures on product spaces that arise from individual measures on the component spaces. Product measures are just like joint distributions of independent random variables, as we shall see after we define both concepts.

THEOREM 149. *Let* $(\Omega_i, \mathcal{F}_i, \mu_i)$ *for* $i = 1, 2$ *be* σ*-finite measure spaces. There exists a unique measure* μ *defined on* $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ *that satisfies* $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ *for all* $A_1 \in \mathcal{F}_1$ *and* $A_2 \in \mathcal{F}_2$*.*

PROOF. The uniqueness will follow from Theorem 43 since any two such measures will agree on the π-system of product sets. For the existence, consider the measurable function $\mu_2(B_{\omega_1})$ defined in Proposition 140. For $B \in \mathcal{F}_1 \otimes \mathcal{F}_2$, define

$$\mu(B) = \int \mu_2(B_{\omega_1})\mu_1(d\omega_1).$$

Because $\mu_2(B_{\omega_1}) \geq 0$, μ is a σ-finite measure. (See Example 124.) If B is a product set $A_1 \times A_2$, then $B_{\omega_1} = A_2$ for all ω_1, and

$$\mu(B) = \int \mu_2(A_2) I_{A_1}(\omega_1)\, \mu_1(d\omega_1) = \mu_1(A_1)\mu_2(A_2).$$

It follows that μ is the desired measure. □

DEFINITION 150. The measure μ in Theorem 149 is called the *product measure of* μ_1 *and* μ_2 and is sometimes denoted $\mu_1 \times \mu_2$.

How to integrate with respect to a product measure is an interesting question. For nonnegative functions, there is a simple answer.

THEOREM 151. (FUBINI/TONELLI THEOREM) *Let* $(\Omega_1, \mathcal{F}_1, \mu_1)$ *and* $(\Omega_2, \mathcal{F}_2, \mu_2)$ *be* σ*-finite measure spaces. Let* $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ *be a nonnegative* $\mathcal{F}_1 \otimes \mathcal{F}_2/\mathcal{B}^1$*-measurable function. Then*

$$\int f d\mu_1 \times \mu_2 = \int \left[\int f(\omega_1, \omega_2)\mu_1(d\omega_1)\right] \mu_2(d\omega_2) = \int \left[\int f(\omega_1, \omega_2)\mu_2(d\omega_2)\right] \mu_1(d\omega_1). \tag{152}$$

PROOF. We will use the standard machinery. If f is the indicator of a set B, then all three integrals in (152) equal $\mu_1 \times \mu_2(B)$, as in the poof of Theorem 149. By linearity of integrals, the three integrals are the same for all nonnegative simple functions. Next, let $\{f_n\}_{n=1}^{\infty}$ be a sequence of nonnegative simple functions all $\leq f$ such that $\lim_{n\to\infty} f_n = f$. We have just shown that, for each n,

$$\int f_n d\mu_1 \times \mu_2 = \int \left[\int f_n(\omega_1, \omega_2)\mu_1(d\omega_1)\right] \mu_2(d\omega_2).$$

For each ω_2, the monotone convergence theorem says

$$\lim_{n\to\infty} \int f_n(\omega_1, \omega_2)\mu_1(d\omega_1) = \int f(\omega_1, \omega_2)\mu_1(d\omega_1).$$

Again, the monotone convergence theorem says that

$$\lim_{n\to\infty} \int \left[\int f_n(\omega_1, \omega_2)\mu_1(d\omega_1)\right] \mu_2(d\omega_2) = \int \left[\lim_{n\to\infty} \int f_n(\omega_1, \omega_2)\mu_1(d\omega_1)\right] \mu_2(d\omega_2).$$

Combining these last three equations proves that the first two integrals in (152) are equal. A similar argument shows that the first and third are equal. □

36-752: Lecture 13

Theorem 151 says that nonnegative product-measurable functions can be integrated in either order to get the integral with respect to product measure. A similar result holds for integrable product-measurable functions.

COROLLARY 153. *Let $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ be σ-finite measure spaces. Let $f : \Omega_1 \times \Omega_2 \to I\!R$ be a function that is integrable with respect to $\mu_1 \times \mu_2$. Then (152) holds.*

The only sticky point in the proof of Corollary 153 is making sure that $\infty - \infty$ occurs with measure zero in the iterated integrals. But if $\infty(-\infty)$ occurs with positive measure for $f^+(f^-)$ in either of the iterated integrals, that iterated integral would be infinite and $f^+(f^-)$ would not be integrable.

EXERCISE 154. Let X be a nonnegative random variable defined on a probability space $(\Omega, \mathcal{F}, P)$ having distribution function F. Show that $\mathrm{E}(X) = \int_0^\infty [1 - F(x)]dx$.

EXAMPLE 155. This example satisfies neither the conditions of Theorem 151 nor those of Corollary 153. Let

$$f(x,y) = \begin{cases} x\exp(-[1+x^2]y/2) & \text{if } y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned} \int f(x,y)dx &= \exp(-y/2)\int x\exp(-x^2y/2) \\ &= 0, \\ \int f(x,y)dy &= x\int_0^\infty \exp(-[1+x^2]y/2)dy \\ &= \frac{2x}{1+x^2}. \end{aligned}$$

The iterated integral in one direction is 0 and is undefined in the other direction.

These results extend to arbitrary finite products.

EXAMPLE 156. The product of k copies of Lebesgue measure on $\mathbb{R}^1$ is Lebesgue measure on $\mathbb{R}^k$. Theorem 151 and Corollary 153 give conditions under which integrals can be performed in any desired order.

Independence. We shall define what it means for collections of events and random quantities to be independent.

DEFINITION 157. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be subsets of $\mathcal{F}$. We say that $\mathcal{C}_1$ and $\mathcal{C}_2$ are *independent* if $P(A_1 \cap A_2) = P(A_1)P(A_2)$ for all $A_1 \in \mathcal{C}_1$ and $A_2 \in \mathcal{C}_2$.

EXAMPLE 158. If each of $\mathcal{C}_1$ and $\mathcal{C}_2$ contains only one event, then $\mathcal{C}_1$ being independent of $\mathcal{C}_2$ is the same as those events being independent.

DEFINITION 159. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $(S_i, \mathcal{A}_i)$ for $i = 1, 2$ be measurable spaces. Let $X_i : \Omega \to S_i$ be $\mathcal{F}/\mathcal{A}_i$ measurable for $i = 1, 2$. We say that X_1 and X_2 are *independent* if the σ-field's $\sigma(X_1)$ and $\sigma(X_2)$ (see Definition 59) are independent.

PROPOSITION 160. *If $\mathcal{C}_1$ and $\mathcal{C}_2$ are independent π-systems then $\sigma(\mathcal{C}_1)$ and $\sigma(\mathcal{C}_2)$ are independent.*

EXAMPLE 161. Let f_1 and f_2 be densities with respect to Lebesgue measure. Let P be defined on $(\mathbb{R}^2, \mathcal{B}^2)$ by $P(C) = \int_C \int f_1(x) f_2(y) dx dy$. Then the following two σ-field's are independent :

$$\begin{aligned} \mathcal{C}_1 &= \{A \times \mathbb{R} : A \in \mathcal{B}^1\}, \\ \mathcal{C}_2 &= \{\mathbb{R} \times A : A \in \mathcal{B}^1\}. \end{aligned}$$

Also, the following two random variables are independent: $X_1(x, y) = x$ and $X_2(x, y) = y$, the coordinate projection functions. Indeed, $\mathcal{C}_i = \sigma(X_i)$ for $i = 1, 2$.

EXAMPLE 162. Let X_1 and X_2 be two random variables defined on the same probability space $(\Omega, \mathcal{F}, P)$. Suppose that the joint distribution of (X_1, X_2) has a density $f(x, y)$ that factors into $f(x, y) = f_1(x) f_2(y)$, the two marginal densities. Then, for each product set $A \times B$ with $A, B \in \mathcal{B}^1$,

$$\begin{aligned} \Pr(X_1 \in A, X_2 \in B) &= \Pr((X_1, X_2) \in A \times B) \\ &= \int_A \int_B f_1(x) f_2(y) dy dx \\ &= \int_A f_1(x) dx \int_B f_2(y) dy \\ &= \Pr(X_1 \in A) \Pr(X_2 \in B). \end{aligned}$$

So, X_1 and X_2 are independent. The same reasoning would apply if the two random variables were discrete. It would also apply if one were discrete and the other continuous.

These definitions extend to more than two collections of events and more than two random variables.

DEFINITION 163. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $\{\mathcal{C}_\alpha : \alpha \in \aleph\}$ be a collection of subsets of $\mathcal{F}$. We say that the $\mathcal{C}_\alpha$'s are *(mutually) independent* if, for every finite integer $n \geq 2$ and no more than the cardinality of $\aleph$, and for all distinct $\alpha_1, \ldots, \alpha_n \in \aleph$, and $A_{\alpha_i} \in \mathcal{C}_{\alpha_i}$ for $i = 1, \ldots, n$,

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

Definition 164. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $\{(S_\alpha, \mathcal{A}_\alpha) : \alpha \in \aleph\}$ be measurable spaces. Let $X_\alpha : \Omega \to S_\alpha$ be $\mathcal{F}/\mathcal{A}_\alpha$ measurable for each $\alpha \in \aleph$. We say that $\{X_\alpha : \alpha \in \aleph\}$ are *(mutually) independent* if the σ-field's $\{\sigma(X_\alpha) : \alpha \in \aleph\}$ are mutually independent.

Theorem 165. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $(S_i, \mathcal{A}_i)$ for $i = 1, 2$ be measurable spaces. Let $X_1 : \Omega \to S_1$ and $X_2 : \Omega \to S_2$ be random quantities. Define $X = (X_1, X_2)$. The distribution of $X : \Omega \to S_1 \times S_2$, μ_X, is the product measure $\mu_{X_1} \times \mu_{X_2}$ if and only if X_1 and X_2 are independent.*

Proof. For the "if" direction, suppose that X_1 and X_2 are independent. Then for every product set $A_1 \times A_2$,

$$\begin{aligned} \mu_X(A_1 \times A_2) &= \Pr(X_1 \in A_1, X_2 \in A_2) = \Pr(X_1 \in A_1)\Pr(X_2 \in A_2) \\ &= \mu_{X_1}(A_1)\mu_{X_2}(A_2). \end{aligned}$$

It follows from the uniqueness of product measure that μ_X is the product measure.

For the "only if" direction, suppose that $\mu_X = \mu_{X_1} \times \mu_{X_2}$. Then, for every $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$,

$$\begin{aligned} \Pr(X_1 \in A_1, X_2 \in A_2) &= \mu_X(A_1 \times A_2) = \mu_{X_1}(A_1)\mu_{X_2}(A_2) \\ &= \Pr(X_1 \in A_1)\Pr(X_2 \in A_2). \quad \square \end{aligned}$$

Theorem 166. (First Borel-Cantelli lemma) *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. If $\sum_{n=1}^\infty \mu(A_n) < \infty$ then $\mu(\limsup_{n\to\infty} A_n) = 0$.*

Proof. Let $B_i = \bigcup_{n=i}^\infty A_n$. Then $\{B_i\}_{i=1}^\infty$ is a decreasing sequence of sets, each of which has finite measure, so the second part of Lemma 34 says that

$$\lim_{i\to\infty} \mu(B_i) = \mu\left(\lim_{i\to\infty} B_i\right) = \mu\left(\bigcap_{i=1}^\infty B_i\right) = \mu\left(\limsup_{n\to\infty} A_n\right).$$

Since $\sum_{n=1}^\infty \mu(A_n) < \infty$, it follows that $\lim_{i\to\infty} \sum_{n=i}^\infty \mu(A_n) = 0$. Since $\mu(B_i) \le \sum_{n=i}^\infty \mu(A_n)$, $\lim_{i\to\infty} \mu(B_i) = 0$, and the result follows. $\square$

Theorem 167. (Second Borel-Cantelli lemma) *Let $(\Omega, \mathcal{F}, P)$ be a probability space. If $\sum_{n=1}^\infty P(A_n) = \infty$ and if $\{A_n\}_{n=1}^\infty$ are mutually independent, then $P(\limsup_{n\to\infty} A_n) = 1$.*

Proof. Let $B = \limsup_{n\to\infty} A_n$. We shall prove that $P(B^C) = 0$. Let $C_i = \bigcap_{n=i}^\infty A_n^C$. Then $B^C = \bigcup_{i=1}^\infty C_i$. So, we shall prove that $P(C_i) = 0$ for all i. Now, for each i and $k > i$,

$$P(C_i) = P\left(\bigcap_{n=i}^\infty A_n^C\right) \le P\left(\bigcap_{n=i}^k A_n^C\right) = \prod_{n=i}^k [1 - P(A_n)].$$

Use the fact that $\log(1-x) \le -x$ for all $0 \le x \le 1$ to see that, for every $k > i$,

$$\log[P(C_i)] \le \sum_{n=i}^{k} \log[1 - P(A_n)] \le -\sum_{n=i}^{k} P(A_n).$$

Since this is true for all $k > i$, it follows that $\log[P(C_i)] \le -\sum_{n=i}^{\infty} P(A_n) = -\infty$. Hence, $P(C_i) = 0$ for all i. $\square$

THEOREM 168. (KOLMOGOROV 0-1 LAW) *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of independent random quantities. Define $\mathcal{T}_n = \sigma(\{X_i : i \ge n\})$ and $\mathcal{T} = \bigcap_{n=1}^{\infty} \mathcal{T}_n$. Then every event in $\mathcal{T}$ has probability either 0 or 1.*

PROOF. Let $\mathcal{U}_n = \sigma(\{X_i : i \le n\})$, and let $\mathcal{U} = \bigcup_{n=1}^{\infty} \mathcal{U}_n$. Let $A \in \mathcal{U}$ and $B \in \mathcal{T}$. There exists n such that $A \in \mathcal{U}_n$. Because $B \in \mathcal{T}_{n+1}$, it follows that A and B are independent. So $\mathcal{U}$ and $\mathcal{T}$ are independent. It follows from Proposition 160 that $\sigma(\mathcal{U}) = \sigma(\{X_n\}_{n=1}^{\infty})$ and $\mathcal{T}$ are independent. Since $\mathcal{T} \subseteq \sigma(\mathcal{U})$, it follows that $\mathcal{T}$ is independent of itself, hence for all $B \in \mathcal{T}$, $\Pr(B) \in \{0,1\}$ by a homework problem. $\square$

DEFINITION 169. The σ-field $\mathcal{T}$ in Theorem 168 is called the *tail σ-field* of the sequence $\{X_n\}_{n=1}^{\infty}$.

EXERCISE 170. Let $X_1, X_2, \ldots$ be independent, real-valued random variables defined on a probability space. Let $S_n = X_1 + X_2 + \cdots + X_n$. Which of the following is in $\mathcal{T}$?

1. $\{\lim_{n\to\infty} S_n \ \text{ exists}\}$
2. $\{\limsup_{n\to\infty} S_n > 0\}$
3. $\{\limsup_{n\to\infty} S_n/c_n > x\}$ where $c_n \to \infty$

36-752: Lecture 13b

Stochastic Processes. A stochastic process is an indexed collection of random quantities.

DEFINITION 171. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $\aleph$ be a set. Suppose that, for each $\alpha \in \aleph$, there is a measurable space $(\mathcal{X}_\alpha, \mathcal{A}_\alpha)$ and a random quantity $X_\alpha : \Omega \to \mathcal{X}_\alpha$. The collection $\{X_\alpha : \alpha \in \aleph\}$ is called a *stochastic process*, and $\aleph$ is called the *index set*.

The most popular stochastic processes are those for which $\mathcal{X}_\alpha = \mathbb{R}$ for all α. Among those, there are two very commonly used index sets, namely $\aleph = \mathbb{Z}^+$ (sequences of random variables) and $\aleph = \mathbb{R}^{+0}$ (continuous-time stochastic processes). There are, however, many more general index sets than these, and they are all handled in the same general fashion.

EXAMPLE 172. (RANDOM VECTOR) Let $\aleph = \{1, \ldots, k\}$ and for each $i \in \aleph$, let X_i be a random variable (all defined on the same probability space). Then $(X_1, \ldots, X_k)$ is one way to represent $\{X_i : i \in \{1, \ldots, k\}\}$.

EXAMPLE 173. (RANDOM PROBABILITY MEASURE) Let $\Theta : \Omega \to \mathbb{R}^k$ be a random vector with distribution μ_Θ. Let $f : \mathbb{R} \times \mathbb{R}^k \to \mathbb{R}^{+0}$ be a measurable function such that $\int f(x, \theta) dx = 1$ for all $\theta \in \mathbb{R}^k$. Let $\aleph = \mathcal{B}^1$, the Borel σ-field of subsets of $\mathbb{R}$. For each $B \in \aleph$, define

$$X_B(\omega) = \int_B f(x, \Theta(\omega)) dx.$$

The stochastic process $\{X_B : B \in \mathcal{B}^1\}$ is a random probability measure.

The distribution of a stochastic process is the probability measure induced on its range space. Unfortunately, if $\aleph$ is an infinite set, the range space of a stochastic process is an infinite-dimensional product set. We need to be able to construct a σ-field of subsets of such a set.

An infinite product of sets is usually defined as a set of functions.

DEFINITION 174. Let $\aleph$ be a set. Suppose that, for each $\alpha \in \aleph$, there is a set $\mathcal{X}_\alpha$. The *product set* $\mathcal{X} = \prod_{\alpha \in \aleph} \mathcal{X}_\alpha$ is defined to be the set of all functions $f : \aleph \to \bigcup_{\alpha \in \aleph} \mathcal{X}_\alpha$ such that, for every α, $f(\alpha) \in \mathcal{X}_\alpha$. When each $\mathcal{X}_\alpha$ is the same set $\mathcal{Y}$, then the product set is denoted $\mathcal{Y}^\aleph$.

The above definition applies to all product sets, not just infinite ones.

EXAMPLE 175. It is easy to see that finite product sets can be considered sets of functions also. Each k-tuple is a function f from $\{1, \ldots, k\}$ to some space, where the ith coordinate is $f(i)$. For example, the notation $\mathbb{R}^k$ can be thought of as a shorthand for $\mathbb{R}^{\{1,\ldots,k\}}$. A vector $(x_1, \ldots, x_k)$ is the function f such that $f(i) = x_i$ for $i = 1, \ldots, k$.

EXAMPLE 176. (RANDOM PROBABILITY MEASURE) In Example 173, let $\mathcal{X}_B = [0, 1]$ for all $B \in \aleph$. Then each random variable X_B takes values in $\mathcal{X}_B$. The infinite product set is $[0, 1]^{\mathcal{B}^1}$. Each probability measure on $(\mathbb{R}, \mathcal{B}^1)$ is a function from $\mathcal{B}^1$ into $[0, 1]$. The product set contains other functions that are not probabilities. For example, the function $f(B) = 1$ for all $B \in \mathcal{B}^1$ is in the product set, but is not a probability.

We want the σ-field of subsets of a product space to be large enough so that all of the coordinate projection functions are measurable.

Definition 177. Let $\aleph$ be a set. For each $\alpha \in \aleph$, let $(\mathcal{X}_\alpha, \mathcal{A}_\alpha)$ be a measurable space. Let $\mathcal{X} = \prod_{\alpha \in \aleph} \mathcal{X}_\alpha$ be the product set. For each $\alpha \in \aleph$, the *α-coordinate projection function* $p_\alpha : \mathcal{X} \to \mathcal{X}_\alpha$ is defined as $p_\alpha(f) = f(\alpha)$. A *one-dimensional cylinder set is a set of the form* $\prod_{\alpha \in \aleph} B_\alpha$ *where there exists one* $\alpha_0 \in \aleph$ *and* $B \in \mathcal{A}_{\alpha_0}$ *such that* $B_{\alpha_0} = B$ *and* $B_\alpha = \mathcal{X}_\alpha$ *for all* $\alpha \neq \alpha_0$. *Define* $\otimes_{\alpha \in \aleph} \mathcal{A}_\alpha$ *to be the σ-field generated by the one-dimensional cylinder sets, and call this the* product σ-field.

Example 178. Let $\mathcal{X} = \mathbb{R}^k$ for finite k. For $1 \leq i \leq k$, the i-coordinate projection function is $p_i(x_1, \ldots, x_k) = x_i$. An example of a one-dimensional cylinder set (in the case $k = 3$) is $\mathbb{R} \times [-3.7, 4.2) \times \mathbb{R}$.

Example 179. (Random probability measure) In Example 176, let Q be a probability on $\mathcal{B}^1$. Then Q is an element of the infinite product set $[0,1]^{\mathcal{B}^1}$. For each $B \in \aleph$, the B-coordinate projection function evaluated at Q is $p_B(Q) = Q(B)$.

Lemma 180. *The product σ-field is the smallest σ-field such that all p_α are measurable.*

Proof. Notice that, for each $\alpha_0 \in \aleph$ and each $B_{\alpha_0} \in \mathcal{A}_{\alpha_0}$, $p_{\alpha_0}^{-1}(B_{\alpha_0})$ is the one-dimensional cylinder set $\prod_{\alpha \in \aleph} B_\alpha$ where $B_\alpha = \mathcal{X}_\alpha$ for all $\alpha \neq \alpha_0$. This makes every p_α measurable. Notice also that the sets required to make all the p_α measurable generate the product σ-field, hence the product σ-field is the smallest σ-field such that the p_α are all measurable. $\Box$

A stochastic process can be thought of as a random function. When a product space is explicitly considered a function space, the coordinate projection functions are sometimes called *evaluation functionals*.

Theorem 180. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $\aleph$ be a set. For each $\alpha \in \aleph$, let $(\mathcal{X}_\alpha, \mathcal{A}_\alpha)$ be a measurable space and let $X_\alpha : \Omega \to \mathcal{X}_\alpha$ be a function. Let $\mathcal{X} = \prod_{\alpha \in \aleph} \mathcal{X}_\alpha$. Define $\boldsymbol{X} : \Omega \to \mathcal{X}$ by setting $\boldsymbol{X}(\omega)$ to be the function f defined by $f(\alpha) = X_\alpha(\omega)$ for all α. Then $\boldsymbol{X}$ is $\mathcal{F}/\otimes_{\alpha \in \aleph} \mathcal{A}_\alpha$-measurable if and only if each $X_\alpha : \Omega \to \mathcal{X}_\alpha$ is $\mathcal{F}/\mathcal{A}_\alpha$-measurable.*

Proof. For the "if" direction, assume that each X_α is measurable. Let $\mathcal{C}$ be the collection of one-dimensional cylinder sets, which generates the product σ-field. Let $C \in \mathcal{C}$. Then there exists α_0 and $B \in \mathcal{A}_{\alpha_0}$ such that $C = \prod_{\alpha \in \aleph} B_\alpha$ where $B_{\alpha_0} = B$ and $B_\alpha = \mathcal{X}_\alpha$ for all $\alpha \neq \alpha_0$. It follows that $\boldsymbol{X}^{-1}(C) = X_{\alpha_0}^{-1}(B) \in \mathcal{F}$. So, $\boldsymbol{X}$ is measurable by Lemma 60.

For the "only if" direction, assume that $\boldsymbol{X}$ is measurable. Let p_α be the α coordinate projection function for each $\alpha \in \aleph$. It is trivial to see that $X_\alpha = p_\alpha(\boldsymbol{X})$. Since each p_α is measurable, it follows that each X_α is measurable. $\Box$

The function $\boldsymbol{X}$ defined in Theorem 180 is an alternative way to represent the stochastic process $\{X_\alpha : \alpha \in \aleph\}$. That is, instead of thinking of a stochastic process as an indexed set of random quantities, think of it as just another random quantity, but one whose range space is itself a function space. In this way, stochastic processes can be thought of as random functions. The idea is that, instead of thinking of X_α as a function of ω for each α, think of $\boldsymbol{X}(\omega)$ as a function of α for each ω.

Here are some examples of how to think of stochastic processes as random functions and vice-versa.

EXAMPLE 181. Let β_0 and β_1 be random variables. Let $\aleph = \mathbb{R}$. For each $x \in \mathbb{R}$, define $X_x(\omega) = \beta_0(\omega) + \beta_1(\omega)x$. Define $\boldsymbol{X}$ as in Theorem 180. Then $\boldsymbol{X}$ is a random linear function. This means that, for every ω, $\boldsymbol{X}(\omega)$ is a linear function from $\mathbb{R}$ to $\mathbb{R}$. Indeed, it is the function that maps the number x to the number $\beta_0(\omega) + \beta_1(\omega)x$.

EXAMPLE 182. (RANDOM PROBABILITY MEASURE) In Example 173, define $\boldsymbol{X}(\omega)$ to be the function (element of the product set) that maps each set B to $\int_B f(x, \Theta(\omega))dx$. To see that $\boldsymbol{X} : \Omega \to [0,1]^{\mathcal{B}^1}$ is measurable, let C be the one-dimensional cylinder set $\prod_{B\in\aleph} C_B$ where each $C_B = [0,1]$ except $C_{B_0} = D$. Define $g(\theta) = \int_{B_0} f(x,\theta)dx$. We know that $g : \mathbb{R}^k \to [0,1]$ is measurable. Hence $g(\Theta) : \Omega \to [0,1]$ is measurable. It follows that $\boldsymbol{X}^{-1}(C) = g^{-1}(D)$, a measurable set.

Clearly, there must exist probability measures on product spaces such as $\left(\prod_{\alpha\in\aleph} \mathcal{X}_\alpha, \otimes_{\alpha\in\aleph}\mathcal{A}_\alpha\right)$. If we start with a stochastic process $\{X_\alpha : \alpha \in \aleph\}$ and represent it as a random function $\boldsymbol{X}$, then the distribution of $\boldsymbol{X}$ is a probability measure on the product space. This distribution has the obvious marginal distributions for the individual X_α's. But, in general, nothing much can be said about other aspects of the joint distribution.

When a stochastic process is a sequence of independent random quantities, then we can say more.

THEOREM 183. (KOLMOGOROV 0-1 LAW) *Let $\{X_n\}_{n=1}^\infty$ be a sequence of independent random quantities. Define $\mathcal{T}_n = \sigma(\{X_i : i \geq n\})$ and $\mathcal{T} = \bigcap_{n=1}^\infty \mathcal{T}_n$. Then every event in $\mathcal{T}$ has probability either 0 or 1.*

PROOF. Let $\mathcal{U}_n = \sigma(\{X_i : i \leq n\})$, and let $\mathcal{U} = \bigcup_{n=1}^\infty \mathcal{U}_n$. Let $A \in \mathcal{U}$ and $B \in \mathcal{T}$. There exists n such that $A \in \mathcal{U}_n$. Because $B \in \mathcal{T}_{n+1}$, it follows that A and B are independent. So $\mathcal{U}$ and $\mathcal{T}$ are independent. It follows from Proposition 160 that $\sigma(\mathcal{U}) = \sigma(\{X_n\}_{n=1}^\infty)$ and $\mathcal{T}$ are independent. Since $\mathcal{T} \subseteq \sigma(\mathcal{U})$, it follows that $\mathcal{T}$ is independent of itself, hence for all $B \in \mathcal{T}$, $\Pr(B) \in \{0,1\}$ by a homework problem. $\square$

DEFINITION 184. The σ-field $\mathcal{T}$ in Theorem 168 is called the *tail σ-field* of the sequence $\{X_n\}_{n=1}^\infty$.

There is such a thing as product measure on an infinite product space, but to prove it, we need a little more machinery. There is a theorem that says that finite-dimensional distributions that satisfy a certain intuitive condition will determine a unique joint distribution on the product space. This theorem is stated and proven in another course document.

36-752: Lecture 14

DEFINITION 185. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and suppose f is a Borel measurable function defined on this space. Define, for $1 \le p < \infty$,

$$\|f\|_p = \left[\int |f|^p \, d\mu\right]^{1/p}$$

and

$$\|f\|_\infty = \inf\left[\alpha\colon \mu[\omega\colon |f(\omega)| > \alpha] = 0\right].$$

Let $L^p(\Omega, \mathcal{F}, \mu)$ denote the class of all Borel measurable functions f such that $\|f\|_p < \infty$. When the measure space is clear, we usually write only L^p to denote this space of functions.

Convergence of Random Variables. Let $(\Omega, \mathcal{F}, P)$ be a probability space. We have already discussed convergence a.s., in the context of what a.s. means.

Each L^p space has a sense of convergence.

DEFINITION 186. Suppose $f_1, f_2, \ldots$ is a sequence of Borel measurable functions defined on $(\Omega, \mathcal{F}, \mu)$ and each $f_n \in L^p$. Let f be another Borel measurable function on $(\Omega, \mathcal{F}, \mu)$. Then we say that f_n **converges in** L^p to f if $\|f_n - f\|_p \to 0$. Write this as $f_n \xrightarrow{L^p} f$.

EXERCISE 187. Assume $(\Omega, \mathcal{F}, \mu)$ is a measure space with $\mu(\Omega) < \infty$. Show that, for $f, f_1, f_2, \ldots$ real-valued functions defined on this space, $f_n \xrightarrow{L^p} f$ implies $f_n \xrightarrow{L^r} f$ for $r < p$.

Convergence in L^p is different from convergence a.s.

EXAMPLE 188. Let $\Omega = (0, 1)$ with P being Lebesgue measure. Consider the sequence of functions $1, I_{(0,1/2]}, I_{(1/2,1)}, I_{(0,1/3]}, I_{(1/3,2/3]}, \ldots$. These functions converge to 0 in L^p for all finite p since the integrals of their absolute values go to 0. But they clearly don't converge to 0 a.s. since every ω has $f_n(\omega) = 1$ infinitely often. These functions are in L^∞, but they don't converge to 0 in L^∞. because their L^∞ norms are all 1.

EXAMPLE 189. Let $\Omega = (0, 1)$ with P being Lebesgue measure. Consider the sequence of functions

$$f_n(\omega) = \begin{cases} 0 & \text{if } 0 < \omega < 1/n, \\ 1/\omega & \text{if } 1/n \le \omega < 1. \end{cases}$$

Each f_n is in L^p for all p, and $\lim_{n\to\infty} f_n(\omega) = 1/\omega$ a.s. But the limit function is not in L^p for even a single p. Clearly, $\{f_n\}_{n=1}^\infty$ does not converge in L^p.

EXAMPLE 190. Let $\Omega = (0, 1)$ with P being Lebesgue measure. Consider the sequence of functions

$$f_n(\omega) = \begin{cases} n & \text{if } 0 < \omega < 1/n, \\ 0 & \text{otherwise.} \end{cases}$$

Then f_n converges to 0 a.s. but not in L^p since $\int |f_n|^p dP = n^{p-1}$ for all n and finite p. In this case, the a.e. limit is in L^p, but it is not an L^p limit.

Oddly enough convergence in L^∞ does imply convergence a.e., the reason being that L^∞ convergence is "almost" uniform convergence.

PROPOSITION 191. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. If f_n converges to f in L^∞, then $\lim_{n\to\infty} f_n = f$, a.e. $[\mu]$.*

There are other modes of convergence besides those mentioned above.

DEFINITION 192. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let f and $\{f_n\}_{n=1}^\infty$ be measurable functions that take values in a metric space with metric d. We say that f_n converges to f *in measure* if, for every $\epsilon > 0$,

$$\lim_{n\to\infty} \mu(\{\omega : d(f_n(\omega), f(\omega)) > \epsilon\}) = 0.$$

When μ is a probability, convergence in measure is called *convergence in probability*, denoted $f_n \xrightarrow{P} f$.

Convergence in measure is different from a.e. convergence. Example 188 is a classic example of a sequence that converges in measure (in probability in that example) but not a.e. Here is an example of a.e. convergence without convergence in measure (only possible in infinite measure spaces).

EXAMPLE 193. Let $\Omega = \mathbb{R}$ with μ being Lebesgue measure. Let $f_n(x) = I_{[n,\infty)}(x)$ for all n. Then f_n converges to 0 a.e. $[\mu]$. However, f_n does not converge in measure to 0, because $\mu(\{|f_n| > \epsilon\}) = \infty$ for every n.

Example 190 is an example of convergence in probability but not in L^p. Indeed convergence in probability is weaker than L^p convergence.

PROPOSITION 194. *If X_n converges to X in L^p for some $p \geq 1$, then $X_n \xrightarrow{P} X$.*

Convergence in probability is also weaker than converges a.s.

LEMMA 195. *If $X_n \to X$ a.s., then $X_n \xrightarrow{P} X$.*

PROOF. Let $\epsilon > 0$. Let $C = \{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$, and define $C_n = \{\omega : d(X_k(\omega), X(\omega)) < \epsilon$, for all $k \geq n\}$. Clearly, $C \subseteq \bigcup_{n=1}^\infty C_{n,\epsilon}$. Because $\Pr(C) = 1$ and $\{C_n\}_{n=1}^\infty$ is an increasing sequence of events, $\Pr(C_n) \to 1$. Because $\{\omega : d(X_n(\omega), X(\omega)) > \epsilon\} \subseteq C_n^C$,

$$\Pr(d(X_n, X) > \epsilon) \to 0. \quad \square$$

A partial converse of this lemma is true.

LEMMA 196. *If $X_n \xrightarrow{P} X$, then there is a subsequence $\{X_{n_k}\}_{k=1}^\infty$ such that $X_{n_k} \xrightarrow{\text{a.s.}} X$.*

PROOF. Let n_k be large enough so that $n_k > n_{k-1}$ and $\Pr(d(X_{n_k}, X) > 1/2^k) < 1/2^k$. Because $\sum_{k=1}^\infty \Pr(d(X_{n_k}, X) > 1/2^k) < \infty$, we know that $\Pr(d(X_{n_k}, X) > 1/2^k \text{ i.o.}) = 0$. Let $A = \{d(X_{n_k}, X) > 1/2^k \text{ i.o.}\}$. Then $\Pr(A^C) = 1$ and $\lim_{k\to\infty} X_{n_k}(\omega) = X(\omega)$ for every $\omega \in A^C$. $\square$

There is an even weaker form of convergence that we will discuss later in the course.

36-752: Lecture 15

Let $\{X_n\}_{n=1}^\infty$ be a sequence of random variables. As we pointed out earlier, the tail σ-field contains all events of the form $\{X_n \text{ converges}\}$ or $\{X_n \text{ converges to } c\}$. Because $\frac{1}{n}\sum_{i=1}^n X_i$ converges if and only if $\frac{1}{n}\sum_{i=\ell}^n X_i$ converges for all $\ell = 1, 2, \ldots$, we see that $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i$, if it exists, is measurable with respect to the tail σ-field. The Kolmogorov 0-1 law says that the tail σ-field of an independent sequence has all probabilities 0 and 1. So, the sample averages of an independent sequence must converge a.s. to constants if they converge at all. Also, $\sum_{i=1}^n X_i$ must converge a.s. or with probability 0, although it will not necessarily be measurable with respect to the tail σ-field. Next, we will begin study of sums of independent random variables, finding conditions under which sums and averages converge or don't.

Sums of Independent Random Variables. There are several useful theorems about sums of independent random variables. All of these make use of a common setup. Let $\{X_n\}_{n=1}^\infty$ be a sequence of random variables, and define, for each n, $S_n = \sum_{k=1}^n X_k$. First, there is the weak law of large numbers, this version of which does not assume that the X_n's are independent.

Theorem 197. (Weak law of large numbers) *Let $\{X_n\}_{n=1}^\infty$ be uncorrelated random variables with mean 0 and such that $\sum_{i=1}^n \mathrm{Var}(X_i) = o(n^2)$. Then $\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{P} 0$.*

Proof. Since the X_n's are uncorrelated,

$$\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i),$$

which we have assumed goes to 0 as $n \to \infty$. According to Tchebychev's inequality (Corollary 94)

$$\Pr\left(\left|\frac{1}{n}\sum_{i=1}^n X_i\right| > \epsilon\right) \le \frac{1}{\epsilon^2}\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right),$$

which we just showed goes to 0 as $n \to \infty$. $\square$

There are various strong laws of large numbers that conclude that the average converges almost surely. A proof of Theorem 198 is given in another course document.

Theorem 198. (Strong law of large numbers) *Assume that $\{X_k\}_{k=1}^\infty$ are independent and identically distributed random variables with finite mean μ. Then $\lim_{n\to\infty} S_n/n = \mu$, a.s.*

We will prove a stronger law than Theorem 198 later in the course. For now, we will concentrate on sums of independent random variables.

Theorem 199. (Kolmogorov's maximal inequality) *Let $\{X_k\}_{k=1}^n$ be a finite collection of independent random variables with finite variance and mean 0. Define $S_k = \sum_{i=1}^k X_i$ for all k. Then*

$$\Pr\left(\max_{1\le k\le n} |S_k| \ge \epsilon\right) \le \frac{\mathrm{Var}(S_n)}{\epsilon^2}.$$

PROOF. For $n = 1$, the result is just Chebyshev's inequality. So assume that $n > 1$ for the rest of the proof. Let A_k be the event that $|S_k| \geq \epsilon$ but $|S_j| < \epsilon$ for $j < k$. Then $\{A_k\}_{k=1}^n$ are disjoint and

$$\left\{ \max_{1 \leq k \leq n} |S_k| \geq \epsilon \right\} = \bigcup_{k=1}^n A_k. \tag{200}$$

It follows that

$$\begin{aligned}
\mathrm{E}(S_n^2) &\geq \sum_{k=1}^n \int_{A_k} S_n^2 dP \\
&= \sum_{k=1}^n \int_{A_k} \left[S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2\right] dP \\
&\geq \sum_{k=1}^n \int_{A_k} [S_k^2 + 2S_k(S_n - S_k)] dP \\
&= \sum_{k=1}^n \int_{A_k} S_k^2 dP \\
&\geq \epsilon^2 \sum_{k=1}^n \Pr(A_k) \\
&= \epsilon^2 \Pr\left(\max_{1 \leq k \leq n} |S_k| \geq \epsilon \right),
\end{aligned}$$

where the first two inequalities and the first equality are obvious. The second inequality follows from the fact that $I_{A_k} S_k$ is independent of $(S_n - S_k)$ which has mean 0. The third inequality follows since $S_k^2 \geq \epsilon^2$ on A_k, and the third equality follows from (200). □

The reason that this theorem works is that whenever the maximum $|S_k|$ is large, it most likely is $|S_n|$ that is large. There is another inequality like that of Kolmogorov that is often used in proofs, but we will not discuss it in this class:

PROPOSITION 201. (ETEMADI LEMMA) *Let $\{X_n\}_{n=1}^\infty$ be a sequence if independent random variables. Then, for each $\epsilon > 0$ and each finite or infinite m,*

$$\Pr\left(\max_{1 \leq n \leq m} |S_n| > 3\epsilon \right) \leq 3 \max_{1 \leq n \leq m} \Pr(|S_n| > \epsilon).$$

The first theorem on the convergence of sums has a simple condition.

THEOREM 202. *Let $\{X_n\}_{n=1}^\infty$ be independent with mean 0 and suppose that $\sum_{n=1}^\infty \mathrm{Var}(X_n) < \infty$. Then S_n converges a.s.*

PROOF. The proof is to show that S_n is a Cauchy sequence a.s. The sequence $\{S_n(\omega)\}_{n=1}^\infty$ is not Cauchy if and only if there exists a rational $\epsilon > 0$ such that for every n, $\sup_{j,k>n} |S_j(\omega) -$

$S_k(\omega)| \geq \epsilon$. For each n and $\epsilon > 0$, let

$$\begin{aligned} B_{n,\epsilon} &= \{\sup_{j,k>n} |S_j - S_k| \geq \epsilon\}, \\ C_{n,\epsilon} &= \left\{\sup_{k\geq 1} |S_{n+k} - S_n| \geq \frac{\epsilon}{2}\right\}, \end{aligned}$$

So that $B_{n,\epsilon} \subseteq C_{n,\epsilon}$ for all n and $\epsilon > 0$. Then $\{S_n(\omega)\}_{n=1}^{\infty}$ is not a Cauchy sequence if and only if

$$\omega \in \bigcup_{\text{rational } \epsilon > 0} \bigcap_{n=1}^{\infty} B_{n,\epsilon} \subseteq \bigcup_{\text{rational } \epsilon > 0} \bigcap_{n=1}^{\infty} C_{n,\epsilon}.$$

So, it suffices to show that

$$\lim_{n\to\infty} \Pr\left(\sup_{k\geq 1} |S_{n+k} - S_n| \geq \frac{\epsilon}{2}\right) = 0. \tag{203}$$

To show (203), use Theorem 199 to see that, for each $r \geq 1$,

$$\Pr\left(\max_{1\leq k\leq r} |S_{n+k} - S_n| \geq \frac{\epsilon}{2}\right) \leq \frac{4}{\epsilon^2}\sum_{k=1}^{r} \mathrm{Var}(X_{n+k}).$$

The sets whose probabilities are on the left side increase with r, so we can take a limit on both sides as $r \to \infty$:

$$\Pr\left(\sup_{k\geq 1} |S_{n+k} - S_n| \geq \frac{\epsilon}{2}\right) \leq \frac{4}{\epsilon^2}\sum_{k=1}^{\infty} \mathrm{Var}(X_{n+k}) = \frac{4}{\epsilon^2}\sum_{j=n+1}^{\infty} \mathrm{Var}(X_j).$$

Since $\sum_{j=1}^{\infty} \mathrm{Var}(X_j) < \infty$, the tail sums must go to 0, and this implies (203). $\square$

Corollary 204. *Let $\{X_n\}_{n=1}^{\infty}$ be independent. Suppose that $\sum_{n=1}^{\infty} \mathrm{Var}(X_n) < \infty$ and $\sum_{k=1}^{n} \mathrm{E}(X_k)$ converges. Then S_n converges a.s.*

Proof. Let $\mu_n = \sum_{k=1}^{n} \mathrm{E}(X_k)$. Write $S_n = (S_n - \mu_n) + \mu_n$. Theorem 202 says that $S_n - \mu_n$ converges a.s., and we are assuming that μ_n converges, hence the sum of the two converges a.s.$\square$

Example 205. Let the X_n's have normal distribution with mean $1/n^2$ and variance $1/n^2$. Then Theorem 202 says that the partial sums $\sum_{i=1}^{n}(X_i - 1/i^2)$ converge a.s. It follows easily that $\sum_{i=1}^{n} X_i$ converges a.s. as well. Later we will be able to prove that the distribution of the limit is normal with mean and variance equal to $\sum_{n=1}^{\infty} 1/n^2$.

Example 206. Let the X_n's have uniform distributions on the intervals $[-1/n, 1/n]$. Then $\sum_{i=1}^{n} X_i$ converges a.s.

Strong Law of Large Numbers

The preliminary results in this document are numbered locally because they do not figure in the course notes.

Lemma 207. (Kronecker's lemma) *Let $\{x_k\}_{k=1}^\infty$ and $\{b_k\}_{k=1}^\infty$ be sequences of real numbers such that $\sum_{k=1}^\infty x_k = s < \infty$ and $b_k \uparrow \infty$. Then*

$$\lim_{n\to\infty} \frac{1}{b_n} \sum_{k=1}^n b_k x_k = 0.$$

Proof. Define $r_n = \sum_{k=n+1}^\infty x_k$ so that $r_0 = s$. Then $x_k = r_{k-1} - r_k$ for all k. So

$$\begin{aligned}
\sum_{k=1}^n b_k x_k &= \sum_{k=1}^n b_k(r_{k-1} - r_k) \\
&= \sum_{k=0}^{n-1} b_{k+1} r_k - \sum_{k=1}^n b_k r_k \\
&= \sum_{k=1}^{n-1} (b_{k+1} - b_k) r_k + b_1 s - b_n r_n.
\end{aligned}$$

Take absolute values to conclude that

$$\left| \sum_{k=1}^n b_k x_k \right| \le \sum_{k=1}^{n-1} (b_{k+1} - b_k)|r_k| + b_1|s| + b_n|r_n|.$$

Let $\epsilon > 0$. Because $|r_n| \to 0$, there exists N such that for all $k \ge N$, $|r_k| < \epsilon$. It follows that

$$\begin{aligned}
\lim_{n\to\infty} \left| \frac{1}{b_n} \sum_{k=1}^n b_k x_k \right| &\le \lim_{n\to\infty} \frac{\epsilon}{b_n} \sum_{k=N}^{n-1} (b_{k+1} - b_k) \\
&= \epsilon \lim_{n\to\infty} \left(1 - \frac{b_N}{b_n}\right) = \epsilon.
\end{aligned}$$

Since this is true for all $\epsilon > 0$, the limit is 0. □

Theorem 198. (Strong law of large numbers) *Assume that $\{X_k\}_{k=1}^\infty$ are independent and identically distributed random variables with finite mean μ. Then $\lim_{n\to\infty} S_n/n = \mu$, a.s.*

Proof. Define $Y_k = X_k I_{[-k,k]}(X_k)$, $S_n^* = \sum_{k=1}^n Y_k$, and $\mu_k = \mathrm{E}(Y_k)$. Recall that $\mathrm{Var}(Y_k) \le \mathrm{E}(Y_k^2)$. Also,

$$\sum_{k=1}^\infty \frac{1}{k^2} \mathrm{E}(Y_k^2) = \sum_{k=1}^\infty \frac{1}{k^2} \int_{|x|<k} x^2 d\mu_X(x)$$

$$
\begin{aligned}
&= \sum_{k=1}^{\infty} \sum_{j=1}^{k} \frac{1}{k^2} \int_{j-1<|x|\le j} x^2 d\mu_X(x) \\
&= \sum_{j=1}^{\infty} \left(\int_{j-1<|x|\le j} x^2 d\mu_X(x) \sum_{k=j}^{\infty} \frac{1}{k^2} \right) \\
&< \sum_{j=1}^{\infty} \frac{2}{j} \int_{j-1<|x|\le j} x^2 d\mu_X(x) \\
&\le 2\mathrm{E}(|X_1|) < \infty,
\end{aligned}
$$

where the first inequality follows from the fact that $\sum_{k=j}^{\infty} 1/k^2 < 2/j$. So, $\sum_{k=1}^{n} \mathrm{Var}(Y_k)/k^2$ converges It follows from Theorem 202 in the class notes that $\sum_{k=1}^{n}(Y_k - \mu_k)/k$ converges a.s. Now, apply Lemma 207 to conclude that $\frac{1}{n}\sum_{k=1}^{n}(Y_k - \mu_k)$ converges a.s. Since $\mu_k \to \mu$, it follows that $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mu_k = \mu$. So $\frac{1}{n}\sum_{k=1}^{n} Y_k$ converges a.s. to μ.

Notice that $\Pr(Y_k \neq X_k) = \Pr(|X_k| > k)$. Let μ_X denote the distribution of each X_i. Recall that

$$
\begin{aligned}
\mathrm{E}(|X_1|) &= \int_0^{\infty} \Pr(|X_1| > t)dt \\
&\ge \sum_{k=1}^{\infty} \Pr(|X_k| > k).
\end{aligned}
$$

Because $\mathrm{E}(|X_1|) < \infty$, the first Borel-Cantelli lemma says that $\Pr(Y_k \neq X_k \text{ i.o.}) = 0$. Hence $\sum_{k=1}^{\infty}(Y_k - X_k)$ is finite a.s. and $\frac{1}{n}\sum_{k=1}^{n} X_k$ converges a.s. to μ. $\Box$

36-752: Lecture 16

Another interesting theorem about sums of independent random variables is the following. It gives necessary and sufficient conditions for convergence of S_n. For each $c > 0$ and each n, let $X_n^{(c)}(\omega) = X_n(\omega)I_{[0,c]}(|X_n(\omega)|)$. We will prove only the sufficiency part of the result. The necessity proof is not included here but can be found in another course document.

THEOREM 208. (THREE-SERIES THEOREM) *Suppose that $\{X_n\}_{n=1}^\infty$ are independent. For each $c > 0$, consider the following three series:*

$$\sum_{n=1}^\infty \Pr(|X_n| > c), \quad \sum_{n=1}^\infty \mathrm{E}(X_n^{(c)}), \quad \sum_{n=1}^\infty \mathrm{Var}(X_n^{(c)}). \tag{209}$$

A necessary condition for S_n to converge a.s. is that all three series are finite for all $c > 0$. A sufficient condition is that all three series converge for some $c > 0$.

EXAMPLE 210. Let X_n have a uniform distribution on the interval $[a_n, b_n]$. A necessary condition for convergence of S_n is that $\sum_{n=1}^\infty (b_n - a_n)^2 < \infty$ (the third series). Another necessary condition is that $\sum_{n=1}^\infty (a_n + b_n)$ converge (the second series). It follows that a_n and b_n must both converge to 0 so that the first series also converges for all $c > 0$. That the two conditions above are sufficient for the convergence of S_n follows from Corollary 204.

PROOF.Theorem 208 First, define some notation. For each $c > 0$ and each n, define

$$\begin{aligned} S_n^{(c)} &= \sum_{k=1}^n X_k^{(c)}, \\ M_n^{(c)} &= \sum_{k=1}^n \mathrm{E}(X_k^{(c)}), \\ s_n^{(c)} &= \sqrt{\sum_{k=1}^n \mathrm{Var}(X_k^{(c)})}. \end{aligned}$$

For sufficiency, assume that all three series converge for some $c > 0$. Because the second and third series in (209) converge, Corollary 204 says that $S_n^{(c)}$ converges a.s. We know that $\Pr(X_n \neq X_n^{(c)}) = \Pr(|X_n| > c)$. Since the first series in (209) converges, the first Borel-Cantelli lemma says that $\Pr(X_n \neq X_n^{(c)} \text{ i.o.}) = 0$. Hence, for almost all ω, there exists $N(\omega)$ such that $S_n(\omega) - S_n^{(c)}(\omega)$ is the same for all $n \geq N(\omega)$. Hence $S_n(\omega)$ converges for almost all ω. $\square$

EXAMPLE 211. Let

$$\Pr(X_n = x) = \begin{cases} \frac{1}{2n^2} & \text{if } x = n \text{ or } x = -n, \\ \frac{1}{2} - \frac{1}{2n^2} & \text{if } x = -1/n \text{ or } x = 1/n, \\ 0 & \text{otherwise.} \end{cases}$$

Then $\mathrm{E}(X_n) = 0$ and $\mathrm{Var}(X_n) = 1 + 1/n^2 - 1/n^4$. So Theorem 202 does not imply that S_n converges a.s. However, for $c > 0$, $\mathrm{E}(X_n^{(c)}) = 0$ and $\mathrm{Var}(X_n^{(c)})$ eventually equals $1/n^2 - 1/n^4$ while $\Pr(|X_n| > c)$ eventually equals $1/n^2$, so the three-series theorem does imply that S_n converges a.s.

Conditional Expectation. The measure-theoretic definition of conditional expectation is a bit unintuitive, but we will show how it matches what we already know from earlier study.

Definition 212. Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $\mathcal{C} \subseteq \mathcal{F}$ be a sub-σ-field. Let X be a random variable whose mean is defined. We use the symbol $\mathrm{E}(X|\mathcal{C})$ to stand for any function $h : \Omega \to \mathrm{I\!R}$ that is $\mathcal{C}/\mathcal{B}^1$ measurable and that satisfies

$$\int_C h dP = \int_C X dP, \text{ for all } C \in \mathcal{C}. \tag{213}$$

We call such a function h, a *version* of the *conditional expectation of X given $\mathcal{C}$.*

Equation (213) can also be written $\mathrm{E}(I_C h) = \mathrm{E}(I_C X)$ for all $C \in \mathcal{C}$. Any two versions of $\mathrm{E}(X|\mathcal{C})$ must be equal a.s. according to Theorem 119 (part 3). Also, any $\mathcal{C}/\mathcal{B}^1$-measurable function that equals a version of $\mathrm{E}(X|\mathcal{C})$ a.s. is another version.

Example 214. If X is itself $\mathcal{C}/\mathcal{B}^1$ measurable, then X is a version of $\mathrm{E}(X|\mathcal{C})$.

Example 215. If $X = a$ a.s., then $\mathrm{E}(X|\mathcal{C}) = a$ a.s.

Let Y be a random quantity and let $\mathcal{C} = \sigma(Y)$. We will use the notation $\mathrm{E}(X|Y)$ to stand for $\mathrm{E}(X|\mathcal{C})$. According to Theorem 147, $\mathrm{E}(X|Y)$ is some function $g(Y)$ because it is $\sigma(Y)/\mathcal{B}^1$-measurable. We will also use the notation $\mathrm{E}(X|Y = y)$ to stand for $g(y)$.

Example 216. (Joint Densities) Let (X, Y) be a pair of random variables with a joint density $f_{X,Y}$ with respect to Lebesgue measure. Let $\mathcal{C} = \sigma(Y)$. The usual marginal and conditional densities are

$$\begin{aligned} f_Y(y) &= \int f_{X,Y}(x, y) dx, \\ f_{X|Y}(x|y) &= \frac{f_{X,Y}(x, y)}{f_Y(y)}. \end{aligned}$$

The traditional calculation of the conditional mean of X given $Y = y$ is

$$g(y) = \int x f_{X|Y}(x|y) dx.$$

That is, $\mathrm{E}(X|Y) = g(Y)$ is the traditional definition of conditional mean of X given Y. We also use the symbol $\mathrm{E}(X|Y = y)$ to stand for $g(y)$. We can prove that $h = g(Y)$ is a version of the conditional mean according to Definition 212. Since $g(Y)$ is a function of Y, we know that it is $\mathcal{C}/\mathcal{B}^1$ measurable. We need to show that (213) holds. Let $C \in \mathcal{C}$ so that there exists $B \in \mathcal{B}^1$ so that $C = Y^{-1}(B)$. Then $I_C(\omega) = I_B(Y(\omega))$ for all ω. Then

$$\int_C h dP = \int I_C h dP$$

$$
\begin{aligned}
&= \int I_B(Y)g(Y)dP \\
&= \int I_B g d\mu_Y \\
&= \int I_B(y)g(y)f_Y(y)dy \\
&= \int I_B(y)\int x f_{X|Y}(x|y)dx f_Y(y)dy \\
&= \int\int I_B(y)x f_{X,Y}(x,y)dxdy \\
&= \mathrm{E}(I_B(Y)X) = \mathrm{E}(I_C X).
\end{aligned}
$$

Example 216 can be extended easily to handle two more general cases. First, we could find $\mathrm{E}(r(X)|Y)$ by virtually the same calculation. Second, the use of conditional densities extends to the case in which the joint distribution of (X,Y) has a density with respect to an arbitrary product measure.

Three-Series Theorem

THEOREM 208. (THREE-SERIES THEOREM) *Suppose that $\{X_n\}_{n=1}^\infty$ are independent. For each $c > 0$, consider the following three series:*

$$\sum_{n=1}^\infty \Pr(|X_n| > c), \quad \sum_{n=1}^\infty \mathrm{E}(X_n^{(c)}), \quad \sum_{n=1}^\infty \mathrm{Var}(X_n^{(c)}). \tag{209}$$

A necessary condition for S_n to converge a.s. is that all three series are finite all $c > 0$. A sufficient condition is that all three series converge for some $c > 0$.

PROOF. Recall some notation. For each $c > 0$ and each n, define

$$\begin{aligned} m_n^{(c)} &= \mathrm{E}(X_n^{(c)}), \\ S_n^{(c)} &= \sum_{k=1}^n X_k^{(c)}, \\ M_n^{(c)} &= \sum_{k=1}^n m_k^{(c)}, \\ s_n^{(c)} &= \sqrt{\sum_{k=1}^n \mathrm{Var}(X_k^{(c)})}. \end{aligned}$$

The sufficiency was proved in the course notes. For necessity, suppose that S_n converges a.s. Let $c > 0$. For each ω such that $S_n(\omega)$ converges, we must have $\lim_{n\to\infty} X_n(\omega) = 0$. It follows that $X_n(\omega) = X_n^{(c)}(\omega)$ for all but finitely many n and so $S_n^{(c)}(\omega)$ converges. Since $\Pr(X_n \neq X_n^{(c)}) = \Pr(|X_n| > c)$ the contrapositive of the second Borel-Cantelli lemma says that the first series in (209) converges. Suppose that the third series in (209) diverges. Since $X_n^{(c)} - m_n^{(c)}$ are uniformly bounded and $s_n^{(c)} \to \infty$, the central limit theorem says that, for all $y > x$,

$$\lim_{n\to\infty} \Pr\left(x < \frac{S_n^{(c)} - M_n^{(c)}}{s_n^{(c)}} \leq y \right) = \Phi(y) - \Phi(x), \tag{217}$$

where Φ is the standard normal df. Since $S_n^{(c)}$ converges a.s., we have $\lim_{n\to\infty} S_n^{(c)}/s_n^{(c)} = 0$ a.s. Hence, $S_n^{(c)}/s_n^{(c)} \xrightarrow{P} 0$. For each $1/2 > \epsilon > 0$,

$$\Pr\left(x < \frac{S_n^{(c)} - M_n^{(c)}}{s_n^{(c)}} \leq y, \left|\frac{S_n^{(c)}}{s_n^{(c)}}\right| < \epsilon \right) \tag{218}$$

$$\geq \Pr\left(x < \frac{S_n^{(c)} - M_n^{(c)}}{s_n^{(c)}} \leq y \right) - \Pr\left(\left|\frac{S_n^{(c)}}{s_n^{(c)}}\right| \geq \epsilon \right). \tag{219}$$

Notice that the event on the left side of (218) can occur only if $x - \epsilon < -M_n^{(c)}/s_n^{(c)} \leq y + \epsilon$. Hence, the event on the left side of (218) cannot occur for both of the pairs $(x, y) = (\epsilon - 1, -\epsilon)$ and $(x, y) = (\epsilon, 1 - \epsilon)$. Let $\delta > 0$ be smaller than both $\Phi(-\epsilon) - \Phi(\epsilon - 1)$ and $\Phi(1 - \epsilon) - \Phi(\epsilon)$.

Then (217) says that there exists $N_1(x,y)$ large enough so that $n \geq N_1(x,y)$ implies that the first probability on the right of (219) is within $\delta/2$ of $\Phi(y) - \Phi(x)$. Also, since $S_n^{(c)}/s_n^{(c)} \xrightarrow{P} 0$, there exists N_2 so that $n \geq N_2$ implies that the second probability on the right of (219) is at most $\delta/2$. So, if

$$n \geq \max\{N_1(\epsilon - 1, -\epsilon), N_1(\epsilon, 1-\epsilon), N_2\},$$

we have

$$\begin{aligned}
&\Pr\left(\epsilon - 1 < \frac{S_n^{(c)} - M_n^{(c)}}{s_n^{(c)}} \leq -\epsilon, \left|\frac{S_n^{(c)}}{s_n^{(c)}}\right| < \epsilon\right) \\
&\quad \geq \ \Phi(-\epsilon) - \Phi(\epsilon - 1) - \delta > 0, \\
&\Pr\left(\epsilon < \frac{S_n^{(c)} - M_n^{(c)}}{s_n^{(c)}} \leq 1 - \epsilon, \left|\frac{S_n^{(c)}}{s_n^{(c)}}\right| < \epsilon\right) \\
&\quad \geq \ \Phi(1-\epsilon) - \Phi(\epsilon) - \delta > 0.
\end{aligned}$$

This contradicts the fact that at least one of the two events on the far left sides of these inequalities is impossible. Hence, $s_n^{(c)}$ cannot diverge and the third series in (209) converges. Theorem 202 now says that $S_n^{(c)} - M_n^{(c)}$ converges a.s. Since we already showed that $S_n^{(c)}$ converges a.s., the second series in (209) must converge. □

36-752: Lecture 17

All of the familiar results about conditional expectation are special cases of the general definition. Here is an unfamiliar example.

EXAMPLE 220. Let X_1, X_2 be independent with $U(0,\theta)$ distribution for some known θ. Let $Y = \max\{X_1, X_2\}$ and $X = X_1$. Find the conditional mean of X given Y. In this case (X, Y) do not have a joint density with respect to any product measure. But we can argue what the conditional distribution, and hence conditional mean, of X given Y should be. With probability 1/2, $X = Y$. With probability 1/2, X is the min of X_1 and X_2 and ought to be uniformly distributed between 0 and Y. The mean of this hybrid distribution is $Y/2 + Y/4 = 3Y/4$. Let's verify this.

First, we see that $h = 3Y/4$ is measurable with respect to $\mathcal{C} = \sigma(Y)$. Next, let $C \in \mathcal{C}$. We need to show that $\mathrm{E}(XI_C) = \mathrm{E}([3Y/4]I_C)$. Theorem 119 (part 4) says that we only need to check this for sets of the form $C = Y^{-1}([0,d])$ with $0 < d < \theta$. Rewrite these expectations as integrals with respect to the joint distribution of (X_1, X_2). We need to show that

$$\int_0^d \int_0^d \frac{x_1}{\theta^2} dx_1 dx_2 = \int_0^d \frac{3y}{4} \frac{2y}{\theta^2} dy, \tag{221}$$

for all $0 < d < \theta$. It is easy to see that both sides of (221) equal $d^3/[2\theta^2]$.

A reminder about versions: If two functions h_1 and h_2 are both $\mathcal{C}/\mathcal{B}^1$-measurable and if they both satisfy $\mathrm{E}(h_i I_C) = \mathrm{E}(XI_C)$ for all $C \in \mathcal{C}$, then they are both versions of $\mathrm{E}(X|\mathcal{C})$. Similarly, any function h' that equals a version of $\mathrm{E}(X|\mathcal{C})$ a.s. and is $\mathcal{C}/\mathcal{B}^1$-measurable is another version.

EXAMPLE 222. In Example 220,

$$h' = \begin{cases} 3Y/4 & \text{if } Y \text{ is irrational,} \\ 0 & \text{otherwise.} \end{cases}$$

is another version of $\mathrm{E}(X|Y)$.

The following fact is immediate by letting $C = \Omega$.

PROPOSITION 223. $\mathrm{E}(\mathrm{E}(X|\mathcal{C})) = \mathrm{E}(X)$.

Here is a generalization of Proposition 223, which is sometimes called the *tower property* of conditional expectations.

PROPOSITION 224. (LAW OF TOTAL PROBABILITY) *If* $\mathcal{C}_1 \subseteq \mathcal{C}_2 \subseteq \mathcal{F}$ *are sub-σ-field's and* $\mathrm{E}(X)$ *exists, then* $\mathrm{E}(X|\mathcal{C}_1)$ *is a version of* $\mathrm{E}(\mathrm{E}(X|\mathcal{C}_2)|\mathcal{C}_1)$.

PROOF. By definition $\mathrm{E}(X|\mathcal{C}_1)$ is $\mathcal{C}_1/\mathcal{B}^1$-measurable. We need to show that, for every $C \in \mathcal{C}_1$,

$$\int_C \mathrm{E}(X|\mathcal{C}_1) dP = \int_C \mathrm{E}(X|\mathcal{C}_2) dP.$$

The left side is $\mathrm{E}(XI_C)$ by definition of conditional mean. Similarly, because $C \in \mathcal{C}_2$ also, the right side is $\mathrm{E}(XI_C)$ as well. □

Example 225. Let (X, Y, Z) be a triple of random variables. Then $\mathrm{E}(X|Y)$ is a version of $\mathrm{E}(\mathrm{E}(X|(Y, Z))|Y)$.

Here is a simple property that extends from expectations to conditional expectations.

Lemma 226. *If* $X_1 \leq X_2$ *a.s., then* $\mathrm{E}(X_1|\mathcal{C}) \leq \mathrm{E}(X_2|\mathcal{C})$ *a.s.*

Proof. Suppose that both $\mathrm{E}(X_1|\mathcal{C})$ and $\mathrm{E}(X_2|\mathcal{C})$ exist. Let

$$\begin{aligned} C_0 &= \{\infty > \mathrm{E}(X_1|\mathcal{C}) > \mathrm{E}(X_2|\mathcal{C})\}, \\ C_1 &= \{\infty = \mathrm{E}(X_1|\mathcal{C}) > \mathrm{E}(X_2|\mathcal{C})\}. \end{aligned}$$

Then, for $i = 0, 1$,

$$0 \leq \int_{C_i} [\mathrm{E}(X_1|\mathcal{C}) - \mathrm{E}(X_2|\mathcal{C})] dP = \int_{C_i} (X_1 - X_2) dP \leq 0.$$

It follows that all terms in this string are 0 and $P(C_i) = 0$ for $i = 0, 1$. Since $C_0 \cup C_1 = \{\mathrm{E}(X_1|\mathcal{C}) > \mathrm{E}(X_2|\mathcal{C})\}$, the result is proven. $\square$

We can prove that versions of conditional expectations exist by the Radon-Nikodym theorem. However, the "modern" way to prove the existence of conditional expectations is through the theory of Hilbert spaces.

36-752: Lecture 18

The following corollary to Proposition 224 is sometimes useful.

COROLLARY 227. *Assume that* $\mathcal{C}_1 \subseteq \mathcal{C}_2 \subseteq \mathcal{F}$ *are sub-*σ*-field's and* $\mathrm{E}(X)$ *exists. If a version of* $\mathrm{E}(X|\mathcal{C}_2)$ *is* $\mathcal{C}_1/\mathcal{B}^1$*-measurable, then* $\mathrm{E}(X|\mathcal{C}_1)$ *is a version of* $\mathrm{E}(X|\mathcal{C}_2)$ *and* $\mathrm{E}(X|\mathcal{C}_2)$ *is a version of* $\mathrm{E}(X|\mathcal{C}_1)$.

EXAMPLE 228. Suppose that X and Y have a joint conditional density given Θ that factors,

$$f_{X,Y|\Theta}(x,y|\theta) = f_{X|\Theta}(x|\theta)f_{Y|\Theta}(y|\theta).$$

Then, the conditional density of X given (Y,Θ) is

$$f_{X|Y,\Theta}(x|y,\theta) = \frac{f_{X,Y|\Theta}(x,y|\theta)}{f_{Y|\Theta}(y|\theta)} = f_{X|\Theta}(x|\theta).$$

With $\mathcal{C}_1 = \sigma(\Theta)$ and $\mathcal{C}_2 = \sigma(Y,\Theta)$, we see that $\mathrm{E}(r(X)|\mathcal{C}_1)$ will be a version of $\mathrm{E}(r(X)|\mathcal{C}_2)$ for every function $r(X)$ with defined mean.

Here is another example of a result that extends from expectations to conditional expectations.

LEMMA 229. *If* $\mathrm{E}(X)$, $\mathrm{E}(Y)$, *and* $\mathrm{E}(X+Y)$ *all exist, then* $\mathrm{E}(X|\mathcal{C})+\mathrm{E}(Y|\mathcal{C})$ *is a version of* $\mathrm{E}(X+Y|\mathcal{C})$.

PROOF. Clearly $\mathrm{E}(X|\mathcal{C}) + \mathrm{E}(Y|\mathcal{C})$ is $\mathcal{C}/\mathcal{B}^1$-measurable. We need to show that for all $C \in \mathcal{C}$,

$$\int_C \mathrm{E}(X|\mathcal{C}) + \mathrm{E}(Y|\mathcal{C})dP = \int_C (X+Y)dP. \tag{230}$$

The left side of (230) is $\int_C XdP + \int_C YdP = \int_C (X+Y)dP$ because $\mathrm{E}(I_C X)$, $\mathrm{E}(I_C Y)$ and $\mathrm{E}(I_C[X+Y])$ all exist. □

The following theorem is used extensively in later results.

THEOREM 231. *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and let* $\mathcal{C}$ *be a sub-*σ*-field of* $\mathcal{F}$. *Suppose that* $\mathrm{E}(Y)$ *and* $\mathrm{E}(XY)$ *exist and that* X *is* $\mathcal{C}/\mathcal{B}^1$*-measurable. Then* $\mathrm{E}(XY|\mathcal{C}) = X\mathrm{E}(Y|\mathcal{C})$.

PROOF. Clearly, $X\mathrm{E}(Y|\mathcal{C})$ is $\mathcal{C}/\mathcal{B}^1$-measurable. We will use the standard machinery on X. If $X = I_B$ for a set $B \in \mathcal{C}$, then

$$\mathrm{E}(I_C XY) = \mathrm{E}(I_{C\cap B}Y) = \mathrm{E}(I_{C\cap B}\mathrm{E}(Y|\mathcal{C})) = \mathrm{E}(I_C X\mathrm{E}(Y|\mathcal{C})), \tag{232}$$

for all $C \in \mathcal{C}$. Hence, $X\mathrm{E}(Y|\mathcal{C}) = \mathrm{E}(XY|\mathcal{C})$. By linearity of expectation, the extreme ends of (232) are equal for every nonnegative simple function, X. Next, suppose that X is nonnegative and let $\{X_n\}$ be a sequence of nonnegative simple functions converging to X from below. Then

$$\begin{aligned}
\mathrm{E}(I_C X_n Y^+) &= \mathrm{E}(I_C X_n \mathrm{E}(Y^+|\mathcal{C})), \\
\mathrm{E}(I_C X_n Y^-) &= \mathrm{E}(I_C X_n \mathrm{E}(Y^-|\mathcal{C})),
\end{aligned}$$

for each n and each $C \in \mathcal{C}$. Apply the monotone convergence theorem to all four sequences above to get

$$\begin{aligned} \mathrm{E}(I_C XY^+) &= \mathrm{E}(I_C X\mathrm{E}(Y^+|\mathcal{C})), \\ \mathrm{E}(I_C XY^-) &= \mathrm{E}(I_C X\mathrm{E}(Y^-|\mathcal{C})), \end{aligned}$$

for all $C \in \mathcal{C}$. It now follows easily from Lemma 229 that $X\mathrm{E}(Y|\mathcal{C}) = \mathrm{E}(XY|\mathcal{C})$. Finally, if X is general, use what we just proved to see that $X^+\mathrm{E}(Y|\mathcal{C}) = \mathrm{E}(X^+Y|\mathcal{C})$ and $X^-\mathrm{E}(Y|\mathcal{C}) = \mathrm{E}(X^-Y|\mathcal{C})$. Apply Lemma 229 one last time. □

In all of the proofs so far, we have proven that the defining equation for conditional expectation holds for all $C \in \mathcal{C}$. Sometimes, this is too difficult and the following result can simplify a proof.

PROPOSITION 233. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{C}$ be a sub-σ-field of $\mathcal{F}$. Let Π be a π-system that generates $\mathcal{C}$. Assume that Ω is the finite or countable union of sets in Π. Let Y be a random variable whose mean exists. Let Z be a $\mathcal{C}/\mathcal{B}^1$-measurable random variable such that $\mathrm{E}(I_C Z) = \mathrm{E}(I_C Y)$ for all $C \in \Pi$. Then Z is a version of $\mathrm{E}(Y|\mathcal{C})$.*

One proof of this result relies on signed measures, and is very similar to the proof of Theorem 43.

Conditional Probability. For $A \in \mathcal{F}$, define $\Pr(A|\mathcal{C}) = \mathrm{E}(I_A|\mathcal{C})$. That is, treat I_A as a random variable X and define the conditional probability of A to be the conditional mean of X. We would like to show that conditional probabilities behave like probabilities. The first thing we can show is that they are additive. That is a consequence of the following result.

It follows easily from Lemma 229 that $\Pr(A|\mathcal{C}) + \Pr(B|\mathcal{C}) = \Pr(A \cup B|\mathcal{C})$ a.s. if A and B are disjoint. The following additional properties are straightforward, and we will not do them all in class. They are similar to Lemma 229.

EXAMPLE 234. (PROBABILITY AT MOST 1) We shall show that $\Pr(A|\mathcal{C}) \le 1$ a.s. Let $B = \{\omega : \Pr(A|\mathcal{C}) > 1\}$. Then $B \in \mathcal{C}$, and

$$P(B) \le \int_B \Pr(A|\mathcal{C})dP = \int_B I_A dP = P(A \cap B) \le P(B),$$

where the first inequality is strict if $P(B) > 0$. Clearly, neither of the inequalities can be strict, hence $P(B) = 0$.

EXAMPLE 235. (COUNTABLE ADDITIVITY) Let $\{A_n\}_{n=1}^\infty$ be disjoint elements of $\mathcal{F}$. Let $W = \sum_{n=1}^\infty \Pr(A_n|\mathcal{C})$. We shall show that W is a version of $\Pr(\bigcup_{n=1}^\infty A_n|\mathcal{C})$. Let $C \in \mathcal{C}$.

$$\begin{aligned} \mathrm{E}\left[I_C I_{\cup_{n=1}^\infty A_n}\right] &= P\left(C \cap \left[\bigcup_{n=1}^\infty A_n\right]\right) \\ &= \sum_{n=1}^\infty P\left(C \cap A_n\right) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{n=1}^{\infty} \int_C \Pr(A_n|\mathcal{C}) dP \\
&= \int_C \sum_{n=1}^{\infty} \Pr(A_n|\mathcal{C}) dP \\
&= \int_C W dP,
\end{aligned}
$$

where the sum and integral are interchangeable by the monotone convergence theorem.

We could also prove that $\Pr(A|\mathcal{C}) \geq 0$ a.s. and $\Pr(\Omega|\mathcal{C}) = 1$, a.s. But there are generally uncountably many different $A \in \mathcal{F}$ and uncountably many different sequences of disjoint events. Although countable additivity holds a.s. separately for each sequence of disjoint events, how can we be sure that it holds simultaneously for all sequences a.s.?

Definition 236. Let $\mathcal{A} \subseteq \mathcal{F}$ be a sub-σ-field. We say that a collection of versions $\{\Pr(A|\mathcal{C}) : A \in \mathcal{A}\}$ are *regular conditional probabilities* if, for each ω, $\Pr(\cdot|\mathcal{C})(\omega)$ is a probability measure on $(\Omega, \mathcal{A})$.

Rarely do regular conditional probabilities exist on $(\Omega, \mathcal{F})$, but there are lots of common sub-σ-field's $\mathcal{A}$ such that regular conditional probabilities exist on $(\Omega, \mathcal{A})$. Oddly enough, the existence of regular conditional probabilities doesn't seem to depend on $\mathcal{C}$.

Example 237. (Joint Densities) Use the same setup as in Example 216. For each y such that $f_Y(y) = 0$, define $f_{X|Y}(x|y) = \phi(x)$, the standard normal density. For each y such that $f_Y(y) > 0$, define $f_{X|Y}$ as in Example 216. Next, for each $A \in \sigma(X)$, define

$$
h(y) = \int_B f_{X|Y}(x|y) dx,
$$

for all y, where $A = X^{-1}(B)$. Finally, define $\Pr(A|\mathcal{C})(\omega) = h(Y(\omega))$. The calculation done in Example 216 shows that this is a version of the conditional mean of I_A given $\mathcal{C}$. But it is easy to see that for each ω, $\Pr(\cdot|\mathcal{C})(\omega)$ is a probability measure on $(\Omega, \sigma(X))$.

36-752: Lecture 19

Convergence in Distribution. Let $\mathcal{X}$ be a topological space and let $\mathcal{B}$ be the Borel σ-field. Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X_n : \Omega \to \mathcal{X}$ be $\mathcal{F}/\mathcal{B}$-measurable. Also, let $X : \Omega \to \mathcal{X}$ be another random quantity. This will be the standard setup for all discussions of convergence in distribution.

DEFINITION 238. We say that X_n *converges in distribution to* X if

$$\lim_{n\to\infty} \mathrm{E}[f(X_n)] = \mathrm{E}[f(X)],$$

for all bounded continuous functions $f : \mathcal{X} \to \mathbb{R}$. We denote this property $X_n \xrightarrow{\mathcal{D}} X$.

EXAMPLE 239. Let $\Omega = \mathbb{R}^\infty$ with $\mathcal{F} = \mathcal{B}^\infty$ and P being the joint distribution of a sequence $\{X_n\}_{n=1}^\infty$ of iid standard normal random variables. Let $X_n(\omega) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \omega_j$. Let $X = X_1$. Then $X_n \xrightarrow{\mathcal{D}} X$ in a trivial way.

There are several conditions that are all equivalent to $X_n \xrightarrow{\mathcal{D}} X$.

THEOREM 240. (PORTMANTEAU THEOREM) *The following are all equivalent if $\mathcal{X}$ is a metric space:*

1. $\lim_{n\to\infty} \mathrm{E}[f(X_n)] = \mathrm{E}[f(X)]$, *for all bounded continuous* f,
2. *For each closed* $C \subseteq \mathcal{X}$, $\limsup_{n\to\infty} \mu_{X_n}(C) \le \mu_X(C)$.
3. *For each open* $A \subseteq \mathcal{X}$, $\liminf_{n\to\infty} \mu_{X_n}(A) \ge \mu_X(A)$.
4. *For each* $B \in \mathcal{B}$ *such that* $\mu_X(\partial B) = 0$, $\lim_{n\to\infty} \mu_{X_n}(B) = \mu_X(B)$.

We will not prove this whole theorem, but we will look a bit more at the four conditions. If $\mathcal{X} = \mathbb{R}$, then the fourth condition is a lot like the familiar convergence of cdf's in places where the limit is continuous. An interval $B = (-\infty, b]$ has $\mu_X(\partial B) = 0$ if and only if there is no mass at b, hence if and only if the cdf is continuous at b. The second condition says that we don't want any mass from the distributions of the X_n's to be able to escape from a closed set, although it could happen that mass from outside of a closed set approaches the boundary. That is why the inequality goes the way it does. Similarly, for the third condition, mass can escape from an open set but nothing should be allowed to "jump" into the open set. The first condition is related to the often overlooked fact that the distribution of a random quantity is equivalent to the means of all bounded continuous functions. The first condition is also a version of what mathematicians call *weak* convergence*, a concept that arises in the theory of normed linear spaces. Many statisticians and probabilists call convergence in distribution "weak convergence," but convergence in distribution is not quite the same as weak convergence in normed linear spaces.

PROOF.Theorem 240 First, notice that the second and third conditions are equivalent since closed sets are complements of open sets. Together the second and third conditions imply the fourth one. We will prove that the fourth condition implies the first one. We will waive hands over the proof that the first condition implies the second one.

Assume the fourth condition. Let f be bounded and continuous, $|f(x)| \leq K$ for all x. Let $\epsilon > 0$. Let $v_0 < v_1 < \cdots < v_M$ be real numbers such that $v_0 < -K < K < v_M$, $v_j - v_{j-1} < \epsilon$ for all $j = 1, \ldots, M$, and $\mu_X(\{x : f(x) = v_j\}) = 0$ for all j. Let $F_j = \{x : v_{j-1} < f(x) \leq v_j\}$. The continuity of f and the fact that $\partial F_j \subseteq \{x : f(x) \in \{v_j, v_{j-1}\}\}$ imply that

$$\{x : v_{j-1} < f(x) < v_j\} \subseteq \mathrm{int}(F_j) \subseteq \overline{F}_j \subseteq \{x : v_{j-1} \leq f(x) \leq v_j\}.$$

By construction

$$\begin{aligned} \left| \sum_{j=1}^{M} v_j \mu_{X_n}(F_j) - \mathrm{E}[f(X_n)] \right| &\leq \epsilon, \\ \left| \sum_{j=1}^{M} v_j \mu_X(F_j) - \mathrm{E}[f(X)] \right| &\leq \epsilon. \end{aligned}$$

By assumption $\mu_X(\partial F_j) = 0$ for all j and

$$\lim_{n\to\infty} \sum_{j=1}^{M} v_j \mu_{X_n}(F_j) = \sum_{j=1}^{M} v_j \mu_X(F_j).$$

Combining these yields $|\lim_{n\to\infty} \mathrm{E}[f(X_n)] - \mathrm{E}[f(X)]| < 2\epsilon$, hence the first condition holds.

To see why the first condition implies the second one, let C be a closed set. For each m, let C_m be the set of points that are at most $1/m$ away from C. The function $f_m(x) = \max\{0, 1 - md(x, C)\}$ is bounded and continuous, equals 0 on C_m^C, equals 1 on C, and lies between 0 and 1 everywhere. We know that $\lim_{n\to\infty} \mathrm{E}(f_m(X_n)) = \mathrm{E}(f_m(X))$ for all m. Also, $\mu_{X_n}(C) \leq \mathrm{E}(f_m(X_n)) \leq \mu_{X_n}(C_m)$ for all n and m. So

$$\limsup_{n\to\infty} \mu_{X_n}(C) \leq \mathrm{E}(f_m(X)) \leq \mu_X(C_m), \tag{241}$$

for all m. Since $\{C_m\}_{m=1}^{\infty}$ is a decreasing sequence of sets whose intersection is C, we have $\lim_{m\to\infty} \mu_X(C_m) = \mu_X(C)$. Since the left side of (241) doesn't depend on m, we have the result. □

Because convergence in distribution depends only on the distributions of the random quantities involved, we do not actually need random quantities in order to discuss convergence in distribution. Hence, we might also use notation like $\mu_n \xrightarrow{\mathcal{D}} \mu$, where μ_n and μ are probability measures on the same space. If $\mathcal{X} = \mathbb{R}$, we might refer to the cdf's and say $F_n \xrightarrow{\mathcal{D}} F$. We might even refer to the names of distributions and say that X_n converges in distribution to a standard normal distribution or some other distribution. Even if we do have random quantities, they don't even have to be defined on the same probability spaces. They do have to take values in the same space, however. For example, for each n, let $(\Omega_n, \mathcal{F}_n, P_n)$ be a probability space, and let $(\Omega, \mathcal{F}, P)$ be another one. Let $(\mathcal{X}, \mathcal{B})$ be a topological space with Borel σ-field. Let $X_n : \Omega_n \to \mathcal{X}$ and $X : \Omega \to \mathcal{X}$ be random quantities. We could then ask whether or not $X_n \xrightarrow{\mathcal{D}} X$. We won't use this last bit of added generality.

Example 242. Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of iid standard normal random variables. Then X_n converges in distribution to standard normal, but does not converge in probability to anything.

Some authors use the expression *converges in law* to mean "converges in distribution". They might write this $X_n \xrightarrow{\mathcal{L}} X$. Others use the expression *converges weakly* and might write it $X_n \xrightarrow{w} X$.[2]

Skorohod proved a result that simplifies some proofs of convergence in distribution when $\mathcal{X} = \mathbb{R}$.

Lemma 243. (Skorohod Theorem) *Let $(\mathcal{X}, \mathcal{B}) = (\mathbb{R}, \mathcal{B}^1)$. Suppose that $X_n \xrightarrow{\mathcal{D}} X$. Then there exist $\{Y_n\}_{n=1}^{\infty}$ and Y defined on $((0,1), \mathcal{B}^1, \lambda)$ (λ being Lebesgue measure) such that Y_n has the same distribution as X_n for all n, Y has the same distribution as X, and $Y_n(\omega) \to Y(\omega)$ for all ω.*

Proof. Let F_n be the cdf of X_n and let F be the cdf of X. Then $\lim_{n\to\infty} F_n(x) = F(x)$ for all x at which F is continuous by part 4 of Theorem 240. Define $Y_n(\omega) = F_n^{-1}(\omega)$ and $Y(\omega) = F^{-1}(\omega)$. Here, the inverse of a general cdf G is defined by $G^{-1}(p) = \inf\{x : G(x) \geq p\}$. It is easy to see that Y_n has the same distribution as X_n and Y has the same distribution as X. For example,

$$\Pr(Y \leq y) = \Pr(F^{-1}(\omega) \leq y) = \Pr(\omega \leq F(y)) = F(y).$$

To see that $Y_n(\omega) \to Y(\omega)$, let $\epsilon > 0$ and let $Y(\omega) - \epsilon < x < Y(\omega)$ be such that F is continuous at x. Then $F(x) < \omega$, so eventually $F_n(x) < \omega$ and eventually $Y(\omega) - \epsilon < x < Y_n(\omega)$, so $\liminf_n Y_n(\omega) \geq Y(\omega)$. A similar argument shows $\limsup_n Y_n(\omega) \leq Y(\omega)$. $\Box$

The following result says that the usual definition of convergence in distribution in one dimension is equivalent to what we have stated above.

Lemma 244. *Let $(\mathcal{X}, \mathcal{B}) = (\mathbb{R}, \mathcal{B}^1)$. Let F_n be the cdf of X_n and let F be the cdf of X. Then $X_n \xrightarrow{\mathcal{D}} X$ if and only if $\lim_{n\to\infty} F_n(x) = F(x)$ for all x at which F is continuous.*

Proof. The proof of the "only if" direction is direct from Theorem 240 because F is continuous at x if and only if $\mu_X(\{x\}) = 0$ and $\{x\}$ is the boundary of $(-\infty, x]$. For the "if" part, construct Y_n and Y as in the proof of Lemma 243. It then follows from the dominated convergence theorem that $\mathrm{E}(f(Y_n)) \to \mathrm{E}(f(Y))$ for all bounded continuous f. $\Box$

Example 245. Let Φ be the standard normal cdf, and let

$$F_n(x) = \begin{cases} 0 & \text{if } x < -n, \\ \frac{\Phi(x) - \Phi(-n)}{\Phi(n) - \Phi(-n)} & \text{if } -n \leq x < n, \\ 1 & \text{if } x \geq n. \end{cases}$$

Then, we see that $\lim_{n\to\infty} F_n(x) = \Phi(x)$ for all x. Each F_n gives probability 1 to a bounded set, but the limit distribution does not.

[2]Convergence in distribution is not the same as weak convergence of continuous linear functionals in functional analysis. It is the same as *weak* convergence*, but we will not go into that distinction here.

Example 246. Let Φ be the standard normal cdf, and let

$$F_n(x) = \begin{cases} 0 & \text{if } x < -n, \\ \Phi(x) & \text{if } -n \le x < n, \\ 1 & \text{if } x \ge n. \end{cases}$$

Then, we see that $\lim_{n\to\infty} F_n(x) = \Phi(x)$ for all x. Each F_n is neither discrete nor continuous, but the limit is continuous.

Example 247. Enumerate the dyadic rationals in this sequence: 1/2, 1/4, 3/4, 1/8, 3/8, 5/8, 7/8, 1/16, 3/16, ... Let μ_n be the measure that puts mass $1/n$ on each of the first n in the list. Then the subsequence $\{\mu_{2^n-1}\}_{n=1}^\infty$ converges in distribution to the uniform distribution on $[0,1]$, but the whole sequence does not converge. Consider the subsequence $\{\mu_{2^{n+2}-2^n-1}\}_{n=1}^\infty$, which converges to a distribution with twice as much probability on $[0,1/2]$ as on $(1/2,1]$.

Example 248. Let F_n be the cdf of the uniform distribution on $[-n,n]$. No subsequence of F_n converges in distribution even though each cdf gives probability 1 to a bounded set.

Examples 245 and 248 illustrate a necessary and sufficient condition for a sequence of distributions to have a convergent (in distribution) subsequence. Even though the F_n in both examples assign probability to 1 to the same intervals, the probability moves out to infinity at different rates in the two examples. In **??**, we will see a condition on how fast probability can move out to infinity and still allow subsequences to converge in distribution.

Convergence in distribution is weaker than convergence in probability, hence it is also weaker than convergence a.s. and L^p convergence.

Proposition 249. *Let $(\mathcal{X},\mathcal{B})$ be a metric space (having metric d) and its Borel σ-field. Let $\{X_n\}_{n=1}^\infty$ be a sequence of random quantities taking values in $\mathcal{X}$ and let X be another random quantity taking values in $\mathcal{X}$.*

1. *If $\lim_{n\to\infty} X_n = X$ a.s., then $X_n \xrightarrow{P} X$.*
2. *If $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{\mathcal{D}} X$.*
3. *If X is degenerate and $X_n \xrightarrow{\mathcal{D}} X$, then $X_n \xrightarrow{P} X$.*
4. *If $X_n \xrightarrow{P} X$, then there is a subsequence $\{n_k\}_{k=1}^\infty$ such that $\lim_{k\to\infty} X_{n_k} = X$, a.s.*

Proof. The first and last claims were proven earlier and are only included for completeness. For the second claim, let C be a closed set and let $C_m = \{x : d(x,C) \le 1/m\}$ for each integer $m > 0$. Then

$$\mu_{X_n}(C) \le \mu_X(C_m) + \Pr(d(X,X_n) > 1/m).$$

It follows that $\limsup_n \mu_{X_n}(C) \le \mu_X(C_m)$. Since $\lim_{m\to\infty} \mu_X(C_m) = \mu_X(C)$, we have that $X_n \xrightarrow{\mathcal{D}} X$ by Theorem 240. The third claim follows by approximating $I_{[c-\epsilon,c+\epsilon]}$ by a bounded continuous function, where $\Pr(X = c) = 1$. $\square$

36-752: Lecture 20

If f is a continuous function and $X_n \xrightarrow{\mathcal{D}} X$, then $f(X_n) \xrightarrow{\mathcal{D}} f(X)$. Indeed, even if f is not continuous, so long as μ_X assigns 0 probability to the set of discontinuities, the result still holds.

THEOREM 250. (CONTINUOUS MAPPING THEOREM) *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random quantities, and let X be another random quantity all taking values in the same metric space $\mathcal{X}$. Suppose that $X_n \xrightarrow{\mathcal{D}} X$. Let $\mathcal{Y}$ be a metric space and let $g : \mathcal{X} \to \mathcal{Y}$. Define*

$$C_g = \{x : \ g \text{ is continuous at } x\}.$$

Suppose that $\Pr(X \in C_g) = 1$. *Then* $g(X_n) \xrightarrow{\mathcal{D}} g(X)$.

The proof of Theorem 250 together with the proof of Theorem 252 are in another course document. They both rely on the second part of Theorem 240, and they resemble the part of the proof of Proposition 249 that we already did.

EXAMPLE 251. If $(S_n - n\mu)/[\sqrt{n}\sigma]$ converges in distribution to standard normal, then $(S_n - n\mu)^2/(n\sigma^2)$ converges in distribution to χ^2 with one degree of freedom.

THEOREM 252. *Let $\{X_n\}_{n=1}^{\infty}$, X, and $\{Y_n\}_{n=1}^{\infty}$ be random quantities taking values in a metric space with metric d. Suppose that $X_n \xrightarrow{\mathcal{D}} X$ and $d(X_n, Y_n) \xrightarrow{P} 0$, then $Y_n \xrightarrow{\mathcal{D}} X$.*

The most common use of this theorem is the following. If the difference between two sequences converges to 0 in probability and if one of the two sequences converges in distribution to X, then so does the other one. A related result is the following.

THEOREM 253. *Let X_n take values in a metric space and let Y_n take values in a metric space. Suppose that $X_n \xrightarrow{\mathcal{D}} X$ and $Y_n \xrightarrow{P} c$, then $(X_n, Y_n) \xrightarrow{\mathcal{D}} (X, c)$.*

PROOF. Let d_1 be the metric in the space where X_n takes values and let d_2 be the metric in the space where Y_n takes values. Then

$$d((x_1, y_1), (x_2, y_2)) = d_1(x_1, x_2) + d_2(y_1, y_2),$$

defines a metric in the product space and the product σ-field is the Borel σ-field. First note that $(X_n, c) \xrightarrow{\mathcal{D}} (X, c)$ since every bounded continuous function of (X_n, c) is a bounded continuous function of X_n alone. Next, note that $d((X_n, Y_n), (X_n, c)) = d_2(Y_n, c)$ and $P_n(d_2(Y_n, c) > \epsilon) \to 0$ for all $\epsilon > 0$, so $d((X_n, Y_n), (X_n, c)) \xrightarrow{P} 0$. By Theorem 252, $(X_n, Y_n) \xrightarrow{\mathcal{D}} (X, c)$. □

EXAMPLE 254. Suppose that $U_n = (S_n - n\mu)/(\sqrt{n}\sigma)$ converges in distribution to standard normal. Suppose also, that $T_n \xrightarrow{P} \sigma$. Then $(U_n, T_n) \xrightarrow{\mathcal{D}} (Z, \sigma)$, where $Z \sim N(0, 1)$. Consider the continuous function $g(z, s) = z\sigma/s$. It follows that

$$g(U_n, T_n) = \frac{S_n - n\mu}{\sqrt{n}T_n} \xrightarrow{\mathcal{D}} Z.$$

EXAMPLE 255. (DELTA METHOD) Suppose that $\lim_{n\to\infty} r_n = \infty$ and $r_n(X_n - a) \xrightarrow{\mathcal{D}} Y$. Then $X_n \xrightarrow{P} a$. Suppose that g is a function that has a derivative $g'(a)$ at a. Define

$$h(x) = \frac{g(x) - g(a)}{x - a} - g'(a).$$

We know that $\lim_{x\to a} h(x) = 0$, so we can make h continuous at a by setting $h(a) = 0$. Also $g(x) - g(a) = (x - a)g'(a) + (x - a)h(x)$. So,

$$r_n[g(X_n) - g(a)] = r_n(X_n - a)g'(a) + r_n(X_n - a)h(X_n).$$

It follows from Theorems 250 and 249 that $h(X_n) \xrightarrow{P} 0$. By Theorem 253, $r_n(X_n - a)h(X_n) \xrightarrow{P} 0$ and $r_n(X_n - a)g'(a) \xrightarrow{\mathcal{D}} g'(a)Y$. By Theorem 252, $r_n[g(X_n) - g(a)] \xrightarrow{\mathcal{D}} g'(a)Y$. After we see the central limit theorem, there will be many examples of the use of this result.

If $g'(a) = 0$ in the above example, there may still be hope if a higher derivative is nonzero.

EXAMPLE 256. Let $\{X_n\}_{n=1}^{\infty}$ be iid with exponential distribution with parameter 2. That is, the density is $2\exp(-2x)$ for $x > 0$. Let $Y_n = \min\{X_1, \ldots, X_n\}$. Then Y_n has an exponential distribution with parameter $2n$. So $n(Y_n - 0) \xrightarrow{\mathcal{D}} X_1$. Let $g(y) = \cos(y)$ so that $g'(y) = -\sin(y)$. Then $n[\cos(Y_n) - 1] \xrightarrow{\mathcal{D}} 0$. But $g(y) - 1 = 0 - y^2/2 + o(y^2)$ as $y \to 0$. So,

$$n^2[g(Y_n) - 1] = \frac{n^2}{2} Y_n^2 + Z_n \xrightarrow{\mathcal{D}} \frac{1}{2} X_1^2,$$

where $Z_n \xrightarrow{P} 0$.

Continuous Mapping and Related Theorems

LEMMA 257. *Let $\mathcal{X}$ and $\mathcal{Y}$ be metric spaces. Let B be a closed subset of $\mathcal{X}$. Let $g : \mathcal{X} \to \mathcal{Y}$. If $x \in \overline{g^{-1}(B)}$ and g is continuous at x, then $x \in g^{-1}(B)$.*

PROOF. If $x \in \overline{g^{-1}(B)}$ then there exists a sequence $\{x_n\}_{n=1}^{\infty}$ of elements of $g^{-1}(B)$ such that $x_n \to x$. If g is continuous at x then $g(x_n) \to g(x)$. Since all $g(x_n) \in B$ and B is closed, $g(x) \in B$. □

THEOREM 250. (CONTINUOUS MAPPING THEOREM) *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random quantities, and let X be another random quantity all taking values in the same metric space $\mathcal{X}$. Suppose that $X_n \xrightarrow{\mathcal{D}} X$. Let $\mathcal{Y}$ be a metric space and let $g : \mathcal{X} \to \mathcal{Y}$. Define*

$$C_g = \{x : \ g \text{ is continuous at } x\}.$$

Suppose that $\Pr(X \in C_g) = 1$. *Then $g(X_n) \xrightarrow{\mathcal{D}} g(X)$.*

PROOF. Let Q_n be the distribution of $g(X_n)$ and let Q be the distribution of $g(X)$. Let R_n be the distribution of X_n and let R be the distribution of X. Let B be a closed subset of $\mathcal{Y}$. If $x \in \overline{g^{-1}(B)}$ but $x \notin g^{-1}(B)$, then g is not continuous at x by Lemma 257. It follows that $\overline{g^{-1}(B)} \subseteq g^{-1}(B) \cup C_g^C$. Now write

$$\begin{aligned}
\limsup_{n\to\infty} Q_n(B) &= \limsup_{n\to\infty} R_n(g^{-1}(B)) \le \limsup_{n\to\infty} R_n(\overline{g^{-1}(B)}) \\
&\le R(\overline{g^{-1}(B)}) \le R(g^{-1}(B)) + R(C_g^C) \\
&= R(g^{-1}(B)) = Q(B),
\end{aligned}$$

and the result now follows from the Theorem 240. □

THEOREM 252. *Let $\{X_n\}_{n=1}^{\infty}$, X, and $\{Y_n\}_{n=1}^{\infty}$ be random quantities taking values in a metric space with metric d. Suppose that $X_n \xrightarrow{\mathcal{D}} X$ and $d(X_n, Y_n) \xrightarrow{P} 0$, then $Y_n \xrightarrow{\mathcal{D}} X$.*

PROOF. Let Q_n be the distribution of Y_n, let R_n be the distribution of X_n and let R be the distribution of X. Let B be an arbitrary closed set. According to Theorem 240, it suffices to show that $\limsup Q_n(B) \le R(B)$. Then

$$\{Y_n \in B\} \subseteq \{d(X_n, B) \le \epsilon\} \cup \{d(X_n, Y_n) > \epsilon\}.$$

Define $C_\epsilon = \{x : d(x, B) \le \epsilon\}$, which is a closed set. So,

$$\begin{aligned}
Q_n(B) &= P_n(Y_n \in B) \\
&\le P_n(d(X_n, B) \le \epsilon) + P_n(d(X_n, Y_n) > \epsilon) \\
&= R_n(C_\epsilon) + P_n(d(X_n, Y_n) > \epsilon).
\end{aligned}$$

We have assumed that $\lim_{n\to\infty} P_n(d(X_n, Y_n) > \epsilon) = 0$ and that $X_n \xrightarrow{\mathcal{D}} X$, so we conclude $\limsup_{n\to\infty} Q_n(B) \le \limsup_{n\to\infty} R_n(C_\epsilon) \le R(C_\epsilon)$. Since B is closed, $\lim_{\epsilon\to 0} R(C_\epsilon) = R(B)$. It follows then that

$$\limsup_{n\to\infty} Q_n(B) \le R(B),$$

hence $Y_n \xrightarrow{\mathcal{D}} X$. □

Here is the convergence in probability version of Theorem 250.

THEOREM 258. *Let* $\{X_n\}_{n=1}^{\infty}$ *be a sequence of random quantities, and let* X *be another random quantity all taking values in the same metric space* $\mathcal{X}$ *with metric* d_1*. Suppose that* $X_n \xrightarrow{P} X$*. Let* $\mathcal{Y}$ *be a metric space with metric* d_2 *and let* $g : \mathcal{X} \to \mathcal{Y}$*. Define*

$$C_g = \{x : \ g \textit{ is continuous at } x\}.$$

Suppose that $\Pr(X \in C_g) = 1$*. Then* $g(X_n) \xrightarrow{P} g(X)$*.*

PROOF. For each $x \in \mathcal{X}$ and $\epsilon > 0$, there exists $\delta(x, \epsilon)$ such that $d_2(g(x), g(y)) < \epsilon$ whenever $d_1(x, y) < \delta(x, \epsilon)$. For every set A,

$$\Pr(A \cap \{d_2(g(X), g(X_n)) < \epsilon\}) \geq \Pr(A \cap \{d_1(X, X_n) < \delta(X, \epsilon)\}).$$

Define $A_m = C_g \cap \{X : \delta(X, \epsilon) \geq 1/m\}$. We know that $\lim_{m\to\infty} \Pr(A_m) = 1$. Also, for every m,

$$\begin{aligned} \Pr(d_2(g(X), g(X_n)) < \epsilon) &\geq \Pr(A_m \cap \{d_2(g(X), g(X_n)) < \epsilon\}) \\ &\geq \Pr(A_m \cap \{d_1(X, X_n) < \delta(X, \epsilon)\}) \\ &\geq \Pr(A_m \cap \{d_1(X, X_n) < 1/m\}). \end{aligned}$$

This last converges to $\Pr(A_m)$ as $n \to \infty$ because $X_n \xrightarrow{P} X$. Hence

$$\liminf_n \Pr(d_2(g(X), g(X_n)) < \epsilon) \geq \Pr(A_m).$$

Now, take limits as $m \to \infty$ on both sides to get that $\liminf_n \Pr(d_2(g(X), g(X_n)) < \epsilon) \geq 1$. So, $g(X_n) \xrightarrow{P} g(X)$. $\square$

36-752: Lecture 21

Characteristic Functions. For the special case in which $(\mathcal{X}, \mathcal{B}) = (\mathrm{I\!R}^p, \mathcal{B}^p)$, there is a useful technique for determining if a sequence of random vectors convergences in distribution. It is based on a characterization of distributions by something simpler than the means of all bounded continuous functions. The means of a special collection of bounded continuous functions, namely $\{\exp(it^\top x) : t \in \mathrm{I\!R}^p\}$, are enough to characterize a distribution. From here on in the notes, i is one of the complex square-roots of -1.[3]

Definition 259. The function $\phi_X(t) = \mathrm{E}\exp(it^\top X)$ is called the *characteristic function* (cf) of X.

(Mathematicians will recognize the cf as the Fourier transform of f.) Every distribution on $\mathrm{I\!R}^p$ has a cf regardless of whether moments exist. Recall from complex analysis that $\exp(iu) = \cos(u) + i\sin(u)$. So, we see that $\exp(it^\top x)$ is indeed bounded as a function of x for each t.

Example 260. (Cauchy distribution) Let $f_X(x) = [\pi(1 + x^2)]^{-1}$. Then $\phi_X(t) = \exp(-|t|)$. To prove this requires contour integration.

The remaining theorems about convergence in distribution are

- the inversion/uniqueness theorem that says that each cf corresponds to a unique distribution,
- the continuity theorem that says that $X_n \xrightarrow{\mathcal{D}} X$ if and only if $\phi_{X_n}(t) \to \phi_X(t)$ for all t (the "only if" direction being trivial), and
- the central limit theorem that says that certain normalized sums of independent (not necessarily identically distributed) random variables with finite variance converge in distribution to a standard normal distribution.

Proposition 261. *If X and Y are independent, then $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$.*

Example 262. (Normal distribution) Let $f_X(x) = \exp(-x^2/2)/\sqrt{2\pi}$ be the density of X. Then

$$\begin{aligned}
\phi_X(t) &= \frac{1}{\sqrt{2\pi}} \int \exp(itx - x^2/2)dx \\
&= \frac{1}{\sqrt{2\pi}} \int \exp\left(-\frac{1}{2}[x - it]^2 - \frac{t^2}{2}\right) dx \\
&= \exp(-t^2/2).
\end{aligned}$$

[3] If I ever use i again to mean something else, stop me so that I can fix it.

EXAMPLE 263. (UNIFORM DISTRIBUTION) Let $f(x) = 1/2$ for $-1 < x < 1$. Then

$$\phi(t) = \frac{1}{2}\int_{-1}^{1} \exp(itx)dx = \frac{\exp(it) - \exp(-it)}{2it} = \frac{\sin(t)}{t}.$$

The smoothness of the cf is related to the existence of moments. Of course all cf's are continuous by the dominated convergence theorem. Since $|\exp(it^\top x) - \exp(iu^\top x)| \le 2$ for all t, u, x, we can pass the limit as $u \to t$ under the integral in $\int[\exp(it^\top x) - \exp(iu^\top x)]d\mu_X(x)$ to get 0 for the limit. Now, suppose that X is a random variable with finite mean. We can write

$$|\exp(ix) - 1|^2 = |\cos(x) + i\sin(x) - 1|^2 = 2 - 2\cos(x) = 2\int_0^x \sin(t)dt \le 2\int_0^x tdt = x^2.$$

This implies that $|\exp(ix) - 1| \le |x|$ for all x. Clearly, $|\exp(ix) - 1| \le 2$ for all x also. So

$$|\exp(ix) - 1| \le \min\{2, |x|\}. \tag{264}$$

This implies that $[\exp(ixt) - 1]/t$ is bounded by a μ_X-integrable function $|x|$. By the dominated convergence theorem, we can pass the limit as $t \to 0$ under the integral to get that $\phi'(0)$ exists and equals $i\mathrm{E}(X)$. With a bit more effort similar results hold if higher moments exist. That is, higher order derivatives of ϕ exist and equal powers of i times the moments times real constants.

THEOREM 265. (INVERSION AND UNIQUENESS) *Let ϕ be the cf for the probability P on $(\mathbb{R}^p, \mathcal{B}^p)$. Let A be a rectangular region of the form*

$$A = \{(x_1, \ldots, x_p) : a_j \le x_j \le b_j \text{ for all } j\},$$

where $a_j < b_j$ for all j and $P(\partial A) = 0$. For each $T > 0$, let

$$B_T = \{(t_1, \ldots, t_p) : -T \le t_j \le T \text{ for all } j\}.$$

Then

$$P(A) = \lim_{T\to\infty} \frac{1}{(2\pi)^p}\int_{B_T} \prod_{j=1}^{p} \left[\frac{\exp(-it_j a_j) - \exp(-it_j b_j)}{it_j}\right] \phi(t)dt_1 \cdots dt_p.$$

Distinct probability measures have distinct cf's.

The proof of Theorem 265 is provided in another course document. The proof relies on the following interesting result.

$$\lim_{T\to\infty} \int_{-T}^{T} \frac{\sin(ct)}{t}dt = \begin{cases} \pi & \text{if } c > 0, \\ 0 & \text{if } c = 0, \\ -\pi & \text{if } c < 0. \end{cases}$$

(Because dt/t is invariant measure with respect to scale changes on $(0, \infty)$, the integral doesn't depend on $|c|$ for $c \ne 0$.) Basically, replace $\phi(t)$ by $\int \prod_{j=1}^{p} \exp(it_j x_j)dP(x)$, change the order of integration, pass the limit inside the integral over x, combine the two products into one, rewrite $\exp(-it_jc_j)$ in terms of sines and cosines (for $c_j \in \{x_j - a_j, x_j - b_j\}$), notice that the cosine terms integrate to 0 over t_j, and apply the above formula to the sine terms. When x_j is between a_j and b_j, the limit of the integral over t_j yields $\pi - (-\pi) = 2\pi$. When x_j is outside of $[a_j, b_j]$, the limit yields either $\pi - \pi$ or $-\pi - (-\pi)$, both 0.

The following theorem allows us to simplify some future proofs by doing only the $p = 1$ case.

LEMMA 266. (CRAMÉR-WOLD) *Let X and Y be p-dimensional random vectors. Then X and Y have the same distribution if and only if $\alpha^\top X$ and $\alpha^\top Y$ have the same distribution for every $\alpha \in I\!R^p$.*

PROOF. We know that X and Y have the same distribution if and only if $\phi_X(t) = \phi_Y(t)$ for every $t \in \mathbb{R}^p$. This is true if and only if $\phi_X(s\alpha) = \phi_Y(s\alpha)$ for all $\alpha \in \mathbb{R}^p$ and all $s \in \mathbb{R}$. But $\phi_X(s\alpha)$ is the cf of $\alpha^\top X$ (as a function of s) and $\phi_Y(s\alpha)$ is the cf of $\alpha^\top Y$. So, $\phi_X(s\alpha) = \phi_Y(s\alpha)$ for all $\alpha \in \mathbb{R}^p$ and all $s \in \mathbb{R}$ if and only if $\alpha^\top X$ and $\alpha^\top Y$ have the same distribution for every $\alpha \in \mathbb{R}^p$. □

If the characteristic function is integrable, a continuous density exists. We will not prove this result.

PROPOSITION 267. *If ϕ is the cf of the cdf F on $(I\!R, \mathcal{B}^1)$ and if ϕ is integrable, then F has a density*

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-itx)\phi(t)dt, \tag{268}$$

which is continuous.

The connection between characteristic functions and convergence in distribution is the following.

THEOREM 269. (CONTINUITY THEOREM) *Let $\{P_n\}_{n=1}^\infty$ be a sequence of probabilities on $(I\!R^p, \mathcal{B}^p)$, and let P be another probability. Let ϕ_n be the cf for P_n, and let ϕ be the cf for P. Then $P_n \xrightarrow{\mathcal{D}} P$ if and only if $\lim_{n\to\infty} \phi_n(t) = \phi(t)$ for all $t \in I\!R^p$.*

See Section 7.2 in Ash for the proof and underlying theory for the case $p = 1$.

EXAMPLE 270. For each j, let Y_j have a uniform distribution on the interval $[-1, 1]$ and let $X_n = \sqrt{\frac{3}{n}}\sum_{j=1}^n Y_j$. Then the cf of X_n is

$$\phi_n(t) = \left(\frac{\sin\left(t\sqrt{3/n}\right)}{t\sqrt{3/n}}\right)^n.$$

We can write $\sin(t) = t - t^3/6 + o(t^3)$ so that, for each t,

$$\frac{\sin\left(t\sqrt{3/n}\right)}{t\sqrt{3/n}} = 1 - \frac{t^2}{2n} + o(1/n),$$

as $n \to \infty$. It follows easily that $\lim_{n\to\infty} \phi_n(t) = \exp(-t^2/2)$. This is the cf of the standard normal distribution.

An interesting corollary to the continuity theorem is that if $\lim_{n\to\infty} \phi_n(t)$ exists for all t and is continuous at 0, then the limit is a cf, and the distributions converge to the distribution with that cf. Another interesting corollary (thanks to Cramér and Wold) is that if $\{X_n\}_{n=1}^{\infty}$ is a sequence of p-dimensional random vectors and X is a random vector, then $X_n \xrightarrow{\mathcal{D}} X$ if and only if $\alpha^\top X_n \xrightarrow{\mathcal{D}} \alpha^\top X$ for all $\alpha \in \mathbb{R}^p$.

Inversion/Uniqueness Theorem

LEMMA 271.

$$\lim_{T\to\infty}\int_{-T}^{T}\frac{\sin(ct)}{t}dt = \begin{cases}\pi & \text{if } c>0,\\ 0 & \text{if } c=0,\\ -\pi & \text{if } c<0.\end{cases} \tag{272}$$

PROOF. Since $\sin(-ct) = -\sin(ct)$ and $\sin(0t) = 0$, it suffices to prove the result for $c > 0$. First, we do an auxiliary integral. For fixed u and c, consider

$$f(x) = \frac{1}{1+u^2}[1-\exp(-ucx)(u\sin(cx)+\cos(cx))].$$

Then $f(0) = 0$ and $f'(x) = c\exp(-ucx)\sin(cx)$. It follows that

$$c\int_0^T \exp(-utc)\sin(ct)dt = f(T).$$

The integrand in (272) is symmetric around 0, so we will work with the integral from 0 to T. Assume that $c > 0$. Since

$$\begin{aligned} c\int_0^\infty \exp(-uct)du &= 1/t,\\ \sin(ct) &= \sin(ct),\end{aligned}$$

we have

$$\begin{aligned}\int_0^T \frac{\sin(ct)}{t}dt &= c\int_0^T\int_0^\infty \sin(ct)\exp(-uct)dudt\\ &= c\int_0^\infty\int_0^T \exp(-uct)\sin(ct)dtdu\\ &= \int_0^\infty \frac{du}{1+u^2} - \int_0^\infty \frac{\exp(-ucT)}{1+u^2}(u\sin(cT)+\cos(cT))du.\end{aligned}$$

The first integral in the last equation equals $\pi/2$ and the last integral goes to 0 as $T\to\infty$.
□

THEOREM 265. (INVERSION AND UNIQUENESS) *Let ϕ be the cf for the probability P on $(\mathbb{R}^p, \mathcal{B}^p)$. Let A be a rectangular region of the form*

$$A = \{(x_1,\ldots,x_p) : a_j \le x_j \le b_j \text{ for all } j\},$$

where $a_j < b_j$ for all j and $P(\partial A) = 0$. For each $T > 0$, let

$$B_T = \{(t_1,\ldots,t_p) : -T \le t_j \le T \text{ for all } j\}.$$

Then

$$P(A) = \lim_{T\to\infty}\frac{1}{(2\pi)^p}\int_{B_T}\prod_{j=1}^p\left[\frac{\exp(-it_ja_j)-\exp(-it_jb_j)}{it_j}\right]\phi(t)dt_1\cdots dt_p.$$

Distinct probability measures have distinct cf's.

PROOF. Apply Fubini's theorem to write

$$\int_{B_T}\prod_{j=1}^{p}\left[\frac{\exp(-it_ja_j)-\exp(-it_jb_j)}{it_j}\right]\phi(t)dt_1\cdots dt_p \tag{273}$$
$$= \int_{\mathbb{R}^p}\int_{B_T}\prod_{j=1}^{p}\left[\frac{\exp(it_j[x_j-a_j])-\exp(it_j[x_j-b_j])}{it_j}\right]dt_1\cdots dt_jd\mu(x).$$

We can do this beacuse the integrand is bounded by $\prod_{j=1}^{p}|b_j-a_j|$ according to (264) and the set over which we are integrating has finite product measure. Rewrite the jth factor in the integrand on the right-side of (273) as

$$\frac{\cos(t_j[x_j-a_j])-\cos(t_j[x_j-b_j])+i\sin(t_j[x_j-a_j])-i\sin(t_j[x_j-b_j])}{it_j}.$$

Since the integration over t_j is from $-T$ to T and $\{\cos(t_j[x_j-a_j])-\cos(t_j[x_j-b_j])\}/t_j$ is bounded and an odd function, its integral is 0. We rewrite the right side of (273) as

$$\int_{\mathbb{R}^p}\int_{B_T}\prod_{j=1}^{p}\left[\frac{\sin(t_j[x_j-a_j])}{t_j}-\frac{\sin(t_j[x_j-b_j])}{t_j}\right]dt_1\cdots dt_jd\mu(x). \tag{274}$$

Define

$$\begin{aligned} g_T(x) &= \int_{B_T}\prod_{j=1}^{p}\left[\frac{\sin(t_j[x_j-a_j])}{t_j}-\frac{\sin(t_j[x_j-b_j])}{t_j}\right]dt_1\cdots dt_p \\ &= \prod_{j=1}^{p}\int_{-T}^{T}\frac{\sin(t_j[x_j-a_j])}{t_j}-\frac{\sin(t_j[x_j-b_j])}{t_j}dt_j. \end{aligned}$$

This function is uniformly bounded for all T and x, hence the limit as $T\to\infty$ of the integral in (274) equals $\int\lim_{T\to\infty}g_T(x)d\mu(x)$. If we define

$$\psi_{a,b}(x)=\begin{cases} 0 & \text{if } x<a, \\ \pi & \text{if } x=a, \\ 2\pi & \text{if } a<x<b, \\ \pi & \text{if } x=b, \\ 0 & \text{if } x>b, \end{cases}$$

then Lemma 271 says that $\lim_{T\to\infty}g_T(x)=\prod_{j=1}^{p}\psi_{a_j,b_j}(x_j)$, which equals $(2\pi)^p$ for $x\in\operatorname{int}(A)$ and equals 0 for $x\in\overline{A}^C$. Since $\mu(\partial A)=0$, we have

$$\frac{1}{(2\pi)^p}\int_{\mathbb{R}^p}\lim_{T\to\infty}g_T(x)d\mu(x)=\mu(A).$$

At most countably many hyperplanes perpendicular to the coordinate axes can have positive μ probability. So, the rectangular regions A with $\mu(\partial A)=0$ form a π-system that generate $\mathcal{B}^p$. It follows from the inversion formula that $\phi_1=\phi_2$ implies $\mu_1=\mu_2$. That is, the characteristic function determines the distribution. □

36-752: Lecture 22

Central Limit Theorem. Suppose that S_n is the sum of n independent random variables with mean 0. Under some conditions to be given soon, there will exist a constant c_n such that S_n/c_n has a cf that is close to the cf of the standard normal distribution. If we can show that the cf of S_n/c_n converges to the cf of the standard normal distribution then we have that S_n/c_n converges in distribution to the standard normal distribution. To achieve this goal, we will need to be able to approximate arbitrary characteristic functions.

By various integrations by parts and reasoning similar to that which achieved (264), we can obtain the following bound:

$$\left|\exp(ix) - \left[1 + ix - \frac{x^2}{2}\right]\right| \leq \min\{|x|^3, x^2\}.$$

In terms of the cf of a random variable X with mean 0 and variance σ^2, this equation says that

$$\left|\phi_X(t) - \left[1 - \frac{t^2\sigma^2}{2}\right]\right| \leq \mathrm{E}\left[\min\{|Xt|^3, (Xt)^2\}\right]. \tag{275}$$

Notice that only a second moment is required in order for the mean on the far right to exist. In order to apply a bound like this to a sum like S_n, we need to approximate a product of cf's by a product of approximations. The following simple results are useful. Their proofs are contained in another course document.

Proposition 276. *Let $z_1, \ldots, z_m$ and $w_1, \ldots, w_m$ be complex numbers with modulus at most 1. Then*

$$\left|\prod_{k=1}^m z_k - \prod_{k=1}^m w_k\right| \leq \sum_{k=1}^m |z_k - w_k|$$

Proposition 277. *For complex z, $|\exp(z) - 1 - z| \leq |z|^2 \exp(|z|)$.*

We are now in position to state and prove a central limit theorem.

Theorem 278. (Lindeberg-Feller central limit theorem) *Let $\{r_n\}_{n=1}^\infty$ be a sequence of integers. For each $n = 1, 2, \ldots$, let $X_{n,1}, \ldots, X_{n,r_n}$ be independent random variables with $X_{n,k}$ having mean 0 and finite nonzero variance $\sigma_{n,k}^2$. Define $s_n^2 = \sum_{k=1}^{r_n} \sigma_{n,k}^2$ and $S_n = \sum_{k=1}^{r_n} X_{n,k}$. Assume that, for every $\epsilon > 0$,*

$$\lim_{n\to\infty} \sum_{k=1}^{r_n} \frac{1}{s_n^2} \int_{\{|X_{n,k}| \geq \epsilon s_n\}} X_{n,k}^2(\omega) dP(\omega) = 0. \tag{279}$$

Then S_n/s_n converges in distribution to the standard normal distribution.

The usual iid central limit theorem is a special case. If $X_1, X_2, \ldots$, are iid with mean 0 and variance σ^2, then let $r_n = n$ and $X_{n,k} = X_k$ for all n and all $k \le n$. Then $s_n^2 = n\sigma^2$ and

$$\sum_{k=1}^{n} \frac{1}{s_n^2} \int_{\{|X_{n,k}|>\epsilon s_n\}} |X_{n,k}(\omega)|^2 dP(\omega) = \frac{1}{\sigma^2} \int_{\{|X_1|>\epsilon\sigma\sqrt{n}\}} |X_1(\omega)|^2 dP(\omega),$$

which goes to 0 as $n \to \infty$, because $\{\omega : |X_1(\omega)| > \epsilon\sigma\sqrt{n}\}$ decreases to the empty set as $n \to \infty$.

The Lyapounov central limit theorem is another special case. In this theorem, instead of assuming (279), we assume that there exists $\delta > 0$ such that $\mathrm{E}[|X_{n,k}|^{2+\delta}] < \infty$ and that

$$\lim_{n\to\infty} \sum_{k=1}^{r_n} \frac{1}{s_n^{2+\delta}} \mathrm{E}\left[|X_{n,k}|^{2+\delta}\right] = 0. \tag{280}$$

Since $|X_{n,k}|^2 \le |X_{n,k}|^{2+\delta}/[\epsilon^\delta s_n^\delta]$ when $|X_{n,k}| > \epsilon s_n$, we have that the sum in (279) is bounded by

$$\frac{1}{\epsilon^\delta} \sum_{k=1}^{r_n} \frac{1}{s_n^{2+\delta}} \int_{\{|X_{n,k}|>\epsilon s_n\}} |X_{n,k}^{2+\delta}(\omega)| d\mu(\omega) \le \frac{1}{\epsilon^\delta} \sum_{k=1}^{r_n} \frac{1}{s_n^{2+\delta}} \mathrm{E}\left[|X_{n,k}|^{2+\delta}\right].$$

Hence, if (280) holds, so does (279).

Example 281. Let $Y_1, Y_2, \ldots$ be independent Poisson random variables with the parameter of Y_k being $1/k$. Then let $X_{n,k} = Y_k - 1/k$ for all n and all $k \le n$. Now, $s_n^2 = L_n = \sum_{k=1}^n 1/k$. For $\delta = 1$, $\mathrm{E}(X_{n,k}^3) = 1/k$ also. Hence

$$\mathrm{E}|X_{n,k}|^3 \le \mathrm{E}\left(\left[X_{n,k} + \frac{1}{k}\right]^3\right) = \frac{1}{k} + \frac{3}{k^2} + \frac{1}{k^3} \le \frac{5}{k}.$$

The sum on the left of (280) is bounded by $5/\sqrt{L_n}$, which goes to 0. So, $[\sum_{k=1}^n Y_k - L_n]/\sqrt{L_n}$ converges in distribution to standard normal. Notice that $L_n = \log(n) + c_n$ where c_n is bounded. By Theorem 252, $[\sum_{k=1}^n Y_k - \log(n)]/\sqrt{\log(n)}$ converges in distribution to standard normal also.

The proof of Theorem 278 works by applying the continuity theorem 269. We must show that the cf of S_n/s_n converges to $\exp(-t^2/2)$ for all t. The proof has two (lengthy) steps. One is to approximate the cf $\phi_{n,k}$ of each $X_{n,k}/s_n$ by $1 - t^2\sigma_{n,k}^2/(2s_n^2)$. The other is to approximate $\exp(-t^2/2)$ by $\prod_{k=1}^{r_n}[1 - t^2\sigma_{n,k}^2/(2s_n^2)]$.

Also, notice that $X_{n,k}$ is divided by s_n in all formulas in the statement of the theorem. Hence, without loss of generality, we can assume that $s_n = 1$ for all n. We do this in the proof, given in a separate document.

Proposition 282. *If the $X_{n,k}$ are uniformly bounded and if $\lim_{n\to\infty} s_n^2 = \infty$, then (279) will hold.*

Example 283. (Bernoulli distribution) If $X_{n,k}$ has a Bernoulli distribution with parameter $1/k$ and $r_n = n$, the condition holds. The theorem does not apply, however, if the Bernoulli parameter is $1/k^2$. Indeed, if the Bernoulli parameter is $1/k^2$, $\sum_{k=1}^n X_{n,k}$ converges almost surely according to Theorem 202. As another example, if $r_n = n$ and the Bernoulli parameter is $k/(n+1)$ for $k = 1, \ldots, n$, then $s_n^2 = n(n+2)/[6(n+1)]$. In fact, r_n could be as small as $n^{1/2+\epsilon}$ for $0 < \epsilon \leq 1/2$, and the theorem would still apply. This example cannot be described as a single sequence as all of the distributions of $X_{n,k}$ change as n changes.

Example 284. (Delta method) Suppose that $Y_1, Y_2, \ldots$ are iid with common mean η and common variance σ^2. Let $X_n = \frac{1}{n}\sum_{j=1}^n Y_j$. Then $\sqrt{n}(X_n - \eta) \xrightarrow{\mathcal{D}} Z$, where Z has a normal distribution with mean 0 and variance σ^2. If g is a function with derivative g' at η, then $\sqrt{n}[g(X_n) - g(\eta)]$ converges in distribution to a normal distribution with mean 0 and variance $[g'(\eta)]^2\sigma^2$.

A multivariate central limit theorem exists for iid sequences, and the proof combines the univariate central limit theorem together with the method of the Cramér-Wold lemma 266 and the Continuity theorem 269.

Theorem 285. (Multivariate Central Limit Theorem) *Let $\{X_n\}_{n=1}^\infty$ be a sequence of iid random vectors with common mean vector η and common covariance matrix Σ. Let $\overline{X}_n$ be the average of the first n of these vectors. Then $Z_n = \sqrt{n}(\overline{X}_n - \eta)$ converges in distribution to multivariate normal with zero mean vector and covariance matrix Σ.*

Proof. By Lemma 266 and its application to convergence in distribution, all we need to show is that, for all α, $\alpha^\top Z_n \xrightarrow{\mathcal{D}} N(0, \alpha^\top \Sigma\alpha)$. For every vector α, let $Y_k = \alpha^\top X_k$ which are iid with common mean $\alpha^\top\eta$ and common variance $\alpha^\top\Sigma\alpha$. Let $s_n^2 = n\alpha^\top\Sigma\alpha$. If $\alpha^\top\Sigma\alpha = 0$, then $\Pr(Y_k = \alpha^\top\eta) = 1$ and $\Pr(\alpha^\top Z_n = 0) = 1$ for all n, which means that $\alpha^\top Z_n \xrightarrow{\mathcal{D}} N(0, \alpha^\top\Sigma\alpha)$. For the rest of the proof, assume that $\alpha^\top\Sigma\alpha > 0$. Theorem 278 says that

$$\frac{n\alpha^\top\overline{X}_n - n\alpha^\top\eta}{s_n} = \frac{\alpha^\top Z_n}{\sqrt{\alpha^\top\Sigma\alpha}} \xrightarrow{\mathcal{D}} N(0,1).$$

Multiply by $\sqrt{\alpha^\top\Sigma\alpha}$ to get that $\alpha^\top Z_n \xrightarrow{\mathcal{D}} N(0, \alpha^\top\Sigma\alpha)$. □

A multivariate central limit theorem also exists for general independent sequences, but it is very cumbersome to state. (Imagine replacing all of the σ^2's and s_n^2's in Theorem 278 by matrices.)

Some Results About Complex Numbers

Proposition 276. *Let $z_1, \ldots, z_m$ and $w_1, \ldots, w_m$ be complex numbers with modulus at most 1. Then*

$$\left|\prod_{k=1}^{m} z_k - \prod_{k=1}^{m} w_k\right| \leq \sum_{k=1}^{m} |z_k - w_k|$$

Proof. We shall use induction. The result is trivially true when $m = 1$. Assume that it is true for $m = m_0$. For $m = m_0 + 1$, we have

$$\begin{aligned}
\left|\prod_{k=1}^{m_0+1} z_k - \prod_{k=1}^{m_0+1} w_k\right| &= \left|\prod_{k=1}^{m_0+1} z_k - w_{m_0+1}\prod_{k=1}^{m} z_k + w_{m_0+1}\prod_{k=1}^{m} z_k - \prod_{k=1}^{m_0+1} w_k\right| \\
&\leq \left|\prod_{k=1}^{m} z_k\right| |z_{m_0+1} - w_{m_0+1}| + \left|\prod_{k=1}^{m} z_k - \prod_{k=1}^{m} w_k\right| |w_{m_0+1}| \\
&\leq \sum_{k=1}^{m} |z_k - w_k| + |z_{m_0+1} - w_{m_0+1}|. \square
\end{aligned}$$

Proposition 277. *For complex z, $|\exp(z) - 1 - z| \leq |z|^2 \exp(|z|)$.*

Proof. Write $\exp(z) - 1 - z = \sum_{k=2}^{\infty} z^k/k!$. Since $k! < (k+2)!$ for $k \geq 0$, we have

$$\left|\sum_{k=2}^{\infty} \frac{z^k}{k!}\right| \leq |z|^2 \sum_{k=0}^{\infty} \frac{|z|^k}{(k+2)!} \leq |z|^2 \exp(|z|). \square$$

Central Limit Theorem

PROOF OF THEOREM 278. Without loss of generality, we assume that $s_n = 1$ for all n. The cf of S_n is

$$\phi_n(t) = \prod_{k=1}^{r_n} \phi_{n,k}(t).$$

According to (275), for each n, k, and t,

$$\begin{aligned}
\left|\phi_{n,k}(t) - \left[1 - \frac{t^2\sigma_{n,k}^2}{2}\right]\right| &\leq \mathrm{E}\left[\min\{|X_{n,k}t|^3, (X_{n,k}t)^2\}\right] \\
&\leq \int_{\{|X_{n,k}|<\epsilon\}} |tX_{n,k}(\omega)|^3 dP(\omega) + \int_{\{|X_{n,k}|\geq\epsilon\}} |tX_{n,k}(\omega)|^2 dP(\omega) \\
&\leq \epsilon|t|^3\sigma_{n,k}^2 + t^2 \int_{\{|X_{n,k}|\geq\epsilon\}} X_{n,k}^2(\omega) dP(\omega).
\end{aligned}$$

It follows that

$$\sum_{k=1}^{r_n} \left|\phi_{n,k}(t) - \left[1 - \frac{t^2\sigma_{n,k}^2}{2}\right]\right| \leq \epsilon|t|^3 + t^2 \sum_{k=1}^{r_n} \int_{\{|X_{n,k}|\geq\epsilon\}} X_{n,k}^2(\omega) dP(\omega).$$

The last sum goes to 0 as $n \to \infty$ according to (279). Since ϵ is arbitrary, we have

$$\lim_{n\to\infty} \sum_{k=1}^{r_n} \left|\phi_{n,k}(t) - \left[1 - \frac{t^2\sigma_{n,k}^2}{2}\right]\right| = 0. \tag{286}$$

In order to apply Proposition 276, we need $\sigma_{n,k}^2$ to all be small. For each $\epsilon > 0$, we have

$$\begin{aligned}
\sigma_{n,k}^2 &= \int_{\{|X_{n,k}|\leq\epsilon\}} X_{n,k}^2(\omega) dP(\omega) + \int_{\{|X_{n,k}|>\epsilon\}} X_{n,k}^2(\omega) dP(\omega) \\
&\leq \epsilon^2 + \int_{\{|X_{n,k}|<\epsilon\}} X_{n,k}^2(\omega) dP(\omega).
\end{aligned}$$

It follows from (279) that

$$\lim_{n\to\infty} \max_k \sigma_{n,k}^2 = 0. \tag{287}$$

Next, fix $t \neq 0$ and notice that for n sufficiently large $0 < t^2\sigma_{n,k}^2/2 < 1$ for all k simultaneously. It follows from Proposition 276 and (286) that

$$\lim_{n\to\infty} \left|\phi_n(t) - \prod_{k=1}^{r_n} \left[1 - \frac{t^2\sigma_{n,k}^2}{2}\right]\right| = 0. \tag{288}$$

Since $s_n^2 = 1$, we have that $\exp(-t^2/2) = \prod_{k=1}^{r_n} \exp(-t^2\sigma_{n,k}^2/2)$. For n large enough so that $t^2\sigma_{n,k}^2/2 < 1$ for all k write

$$\left|\exp\left(-\frac{t^2}{2}\right) - \prod_{k=1}^{r_n}\left[1 - \frac{t^2\sigma_{n,k}^2}{2}\right]\right| \leq \sum_{k=1}^{r_n}\left|\exp\left(-\frac{t^2\sigma_{n,k}^2}{2}\right) - 1 + \frac{t^2\sigma_{n,k}^2}{2}\right|$$

$$
\begin{aligned}
&\le \frac{t^4}{4}\sum_{k=1}^{r_n}\sigma_{n,k}^4 \exp\left(\frac{t^2}{2}\right) \\
&\le \frac{t^4}{4}\max_k \sigma_{n,k}^2 \exp\left(\frac{t^2}{2}\right), \qquad (289)
\end{aligned}
$$

where the first inequality follows from Proposition 276, the second follows from Proposition 277, and the third follows from the fact that $s_n^2 = 1$. Finally, the last term in (289) goes to 0 according to (287). Combining this with (288) says that $\lim_{n\to\infty}\phi_n(t) = \exp(-t^2/2)$. $\square$

Existence of rcd's

This document contains more details about the proof of **??**.

For each rational number q, let $\mu_{X|\mathcal{C}}((-\infty,q])$ be a version of $\Pr(X \le q|\mathcal{C})$. Define

$$\begin{aligned}
C_1 &= \left\{\omega : \mu_{X|\mathcal{C}}((-\infty,q])(\omega) = \inf_{\text{rational } r>q} \mu_{X|\mathcal{C}}((-\infty,r])(\omega), \text{ for all rational } q\right\}, \\
C_2 &= \left\{\omega : \lim_{x\to-\infty,\ x \text{ rational}} \mu_{X|\mathcal{C}}((-\infty,x])(\omega) = 0\right\}, \\
C_3 &= \left\{\omega : \lim_{x\to\infty,\ x \text{ rational}} \mu_{X|\mathcal{C}}((-\infty,x])(\omega) = 1\right\}.
\end{aligned}$$

(Notice that C_2 and C_3 are defined slightly differently than in the original class notes.)

Define

$$M_{q,r} = \{\omega : \mu_{X|\mathcal{C}}((-\infty,q](\omega) < \mu_{X|\mathcal{C}}((-\infty,r])(\omega))\}, \quad M = \bigcup_{q>r} M_{q,r}.$$

If $P(M_{q,r}) > 0$, for some $q > r$ then

$$\begin{aligned}
\Pr(M_{q,r} \cap \{X \le q\}) &= \int_{M_{q,r}} \mu_{X|\mathcal{C}}((-\infty,q])dP < \int_{M_{q,r}} \mu_{X|\mathcal{C}}((-\infty,q])dP \\
&= \Pr(M_{q,r} \cap \{X \le r\}),
\end{aligned}$$

which is a contradiction. Hence, $P(M) = 0$. Next, define

$$N_q = \{\omega \in M^C : \lim_{r\downarrow q,\ r \text{ rational}} \mu_{X|\mathcal{C}}((-\infty,r](\omega) \ne \mu_{X|\mathcal{C}}((-\infty,q])(\omega), \quad N = \bigcup_{\text{All q}} N_q.$$

If $P(N_q) > 0$ for some q, then

$$\begin{aligned}
\Pr(N_q \cap \{X \le q\}) &= \int_{N_q} \mu_{X|\mathcal{C}}((-\infty,q])dP < \int_{N_q} \lim_{r\downarrow q,\ r \text{ rational}} \mu_{X|\mathcal{C}}((-\infty,r])dP \\
&= \lim_{r\downarrow q,\ r \text{ rational}} \int \mu_{X|\mathcal{C}}((-\infty,r])dP = \lim_{r\downarrow q,\ r \text{ rational}} \Pr(N_q \cap \{X \le r\}),
\end{aligned}$$

which is a contradiction. We can use Example 235 once again to prove that $P(N) = 0$. Notice that $C_1 = N^C$, so $P(C_1) = 1$.

Next, notice that

$$\begin{aligned}
0 = P\left(C_1 \cap C_2^C \cap \bigcap_{\text{rational } x} \{X \le x\}\right) &= \lim_{x\to-\infty,\ x \text{ rational}} \int_{C_1\cap C_2^C} \mu_{X|\mathcal{C}}((-\infty,x])dP \\
&= \int_{C_1\cap C_2^C} \lim_{x\to-\infty,\ x \text{ rational}} \mu_{X|\mathcal{C}}((-\infty,x])dP.
\end{aligned}$$

If $P(C_1 \cap C_2) < 1$, then the last integral above is strictly positive, a contradiction. The interchange of limit and integral is justified by the fact that, for $\omega \in C_1$, $\mu_{X|\mathcal{C}}((-\infty,x])$ is nondecreasing in x. A similar contradiction arises if $P(C_1 \cap C_3) < 1$.

Martingales

Martingales. Let $(\Omega, \mathcal{F}, P)$ be a probability space.

DEFINITION 290. Let $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots$ be a sequence of sub-σ-field's of $\mathcal{F}$. We call $\{\mathcal{F}_n\}_{n=1}^\infty$ a *filtration* . If $X_n : \Omega \to \mathrm{I\!R}$ is $\mathcal{F}_n$-measurable for every n, we say that $\{X_n\}_{n=1}^\infty$ is *adapted* to the filtration. If $\{X_n\}_{n=1}^\infty$ is adapted to a filtration $\{\mathcal{F}_n\}_{n=1}^\infty$, and if $\mathrm{E}|X_n| < \infty$ for all n and $\mathrm{E}(X_{n+1}|\mathcal{F}_n) = X_n$ for all n, then we say that $\{X_n\}_{n=1}^\infty$ is a *martingale* relative to the filtration. Alternatively we say that $\{(X_n, \mathcal{F}_n)\}_{n=1}^\infty$ is a martingale. If $X_n \le \mathrm{E}(X_{n+1}|\mathcal{F}_n)$ for all n, we say that $\{(X_n, \mathcal{F}_n)\}_{n=1}^\infty$ is a *submartingale*. If the inequality goes the other way, it is a *supermartingale*.

PROPOSITION 291. *A martingale is both a submartingale and a supermartingale.* $\{X_n\}_{n=1}^\infty$ *is a submartingale if and only if* $\{-X_n\}_{n=1}^\infty$ *is a supermartingale.*

EXAMPLE 292. Let $\{Y_n\}_{n=1}^\infty$ be a sequence of independent random variables with finite mean. Let $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$ and $X_n = \sum_{j=1}^n Y_j$. If $\mathrm{E}(Y_n) = 0$ for every n, then $\{(X_n, \mathcal{F}_n)\}_{n=1}^\infty$ is a martingale. If $\mathrm{E}(Y_n) \ge 0$ for every n, then we have a submartingale, and if $\mathrm{E}(Y_n) \le 0$ for every n, we have a supermartingale.

EXAMPLE 293. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $\{\mathcal{F}_n\}_{n=1}^\infty$ be a filtration. Let ν be a finite measure on $(\Omega, \mathcal{F})$ such that for every n, ν has a density X_n with respect to P when both are restricted to $(\Omega, \mathcal{F}_n)$. Then $\{X_n\}_{n=1}^\infty$ is adapted to the filtration. To see that we have a martingale, we need to show that for every n and $A \in \mathcal{F}_n$

$$\int_A X_{n+1}(\omega) dP(\omega) = \int_A X_n(\omega) dP(\omega). \tag{294}$$

Since $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$, each $A \in \mathcal{F}_n$ is also in $\mathcal{F}_{n+1}$. Hence both sides of (294) equal $\nu(A)$.

EXAMPLE 295. As a more specific example of Example 293, let $\Omega = \mathrm{I\!R}^\infty$ and let $\mathcal{F}_n = \{B \times \mathrm{I\!R}^\infty : B \in \mathcal{B}^n\}$. That is, $\mathcal{F}_n$ is the collection of cylinder sets corresponding to the first n coordinates (the σ-field generated by the first n coordinate projection functions). Let P be the joint distribution of an infinite sequence of iid standard normal random variables. Let ν be the joint distribution of an infinite sequence of iid exponential random variables with parameter 1. For each n, when we restrict both P and ν to $\mathcal{F}_n$, ν has the density

$$X_n(\omega) = \begin{cases} (2\pi)^{n/2} \exp\left(\sum_{j=1}^n [\omega_j^2/2 - \omega_j]\right) & \text{for } \omega_1, \dots, \omega_n > 0, \\ 0 & \text{otherwise,} \end{cases}$$

with respect to P. It is easy to see that

$$\mathrm{E}(X_{n+1}|\mathcal{F}_n) = X_n \mathrm{E}(\sqrt{2\pi} \exp(\omega_{n+1}^2/2 - \omega_{n+1}) I_{(0,\infty)}(\omega_{n+1})) = X_n.$$

EXAMPLE 296. Let $\{(X_n, \mathcal{F}_n)\}_{n=1}^\infty$ be a martingale. Let $\phi : \mathbb{R} \to \mathbb{R}$ be a convex function such that $\mathrm{E}[\phi(X_n)]$ is finite for all n. Define $Y_n = \phi(X_n)$. Then

$$\begin{aligned} \mathrm{E}(Y_{n+1}|\mathcal{F}_n) &= \mathrm{E}[\phi(X_{n+1})|\mathcal{F}_n] \\ &\geq \phi\left[\mathrm{E}(X_{n+1}|\mathcal{F}_n)\right] \\ &= \phi(X_n) = Y_n, \end{aligned}$$

where the inequality follows from the conditional version of Jensen's inequality. So $\{(Y_n, \mathcal{F}_n)\}_{n=1}^\infty$ is a submartingale.

EXAMPLE 297. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $\{Y_n\}_{n=1}^\infty$ be a sequence of random variables and $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$. Suppose that, for each n, $\mu_{Y_1,\ldots,Y_n}$ has a strictly positive density p_n with respect to Lebesgue measure λ^n. Let Q be another probability on $(\Omega, \mathcal{F})$ such that $Q((Y_1, \ldots, Y_n)^{-1}(\cdot))$ has a density q_n with respect to λ^n for each n. Define

$$X_n = \frac{q_n(Y_1, \ldots, Y_n)}{p_n(Y_1, \ldots, Y_n)}.$$

It is easy to check that, for each n and $H \in \mathcal{B}^n$

$$\begin{aligned} \mathrm{E}(I_H(Y_1, \ldots, Y_n)X_n) &= \int_H \frac{q_n(y_1, \ldots, y_n)}{p_n(y_1, \ldots, y_n)} p_n(y_1, \ldots, y_n) d\lambda^n(y_1, \ldots, y_n) \\ &= Q((Y_1, \ldots, Y_n)^{-1}(H)). \end{aligned}$$

This makes X_n a density for Q with respect to P on the σ-field $\mathcal{F}_n$. Hence we have a martingale according to the construction in Example 293. This is an example of likelihood ratios, and it is a generalization of Example 295.

EXAMPLE 298. Let $\{\mathcal{F}_n\}_{n=1}^\infty$ be a filtration and let X be a random variable with finite mean. Define $X_n = \mathrm{E}(X|\mathcal{F}_n)$. By the law of total probability we have a martingale. Such a martingale is sometimes called a *Lévy martingale.*

EXAMPLE 299. Consider Example 292 again. Think of Y_n as being the amount that a gambler wins per unit of currency bet on the nth play in a sequence of games. Let Y_0 denote the gambler's initial fortune which we can assume is a known value, and let $\mathcal{F}_0$ be the trivial σ-field. (We could let Y_0 be a random variable and let $\mathcal{F}_0 = \sigma(Y_0)$, but then we would also have to expand $\mathcal{F}_n$ to $\sigma(Y_0, \ldots, Y_n)$.) Suppose that the gambler devises a system for determining how much $W_n \geq 0$ to bet on the nth play. We assume that W_n is $\mathcal{F}_{n-1}$ measurable for each n. This forces the gambler to choose the amount to bet before knowing what will happen. Now, define $Z_n = Y_0 + \sum_{j=1}^n W_j Y_j$. Since

$$\mathrm{E}(W_{n+1}Y_{n+1}|\mathcal{F}_n) = W_{n+1}\mathrm{E}(Y_{n+1}|\mathcal{F}_n) = W_{n+1}\mathrm{E}(Y_{n+1}),$$

and $W_{n+1} \geq 0$, we have that $\mathrm{E}(W_{n+1}Y_{n+1}|\mathcal{F}_n)$ is ≥ 0, $= 0$, or ≤ 0 according as $\mathrm{E}(Y_{n+1}) \geq 0$, $= 0$, or ≤ 0. That is, $\{(Z_n, \mathcal{F}_n)\}_{n=1}^\infty$ is a submartingale, a martingale, or a supermartingale according as $\{(X_n, \mathcal{F}_n)\}_{n=1}^\infty$ is a submartingale, a martingale, or a supermartingale. This result is often described by saying that gambling systems cannot change whether a game is favorable, fair, or unfavorable to a gambler.

Definition 300. A sequence $\{W_n\}_{n=1}^{\infty}$ of random variables such that W_n is $\mathcal{F}_{n-1}/\mathcal{B}^1$-measurable is called *previsible* . (If there is no $\mathcal{F}_0$, then require that W_1 be constant.)

Example 301. Let $\{(X_n, \mathcal{F}_n)\}_{n=1}^{\infty}$ be a martingale, and let $\{W_n\}_{n=1}^{\infty}$ be previsible. Define $Z_1 = X_1$ and $Z_{n+1} = Z_n + W_{n+1}(X_{n+1} - X_n)$ for $n \geq 1$. Then Z_n is $\mathcal{F}_n/\mathcal{B}^1$-measurable and

$$\mathrm{E}(Z_{n+1}|\mathcal{F}_n) = Z_n + W_{n+1}\mathrm{E}(X_{n+1} - X_n|\mathcal{F}_n) = Z_n,$$

for all $n \geq 1$. This makes $\{(Z_n, \mathcal{F}_n)\}_{n=1}^{\infty}$ a martingale. This is called a *martingale transform.* Example 299 is an example of this.

Theorem 302. (Doob decomposition) *$\{(X_n, \mathcal{F}_n)\}_{n=1}^{\infty}$ is a submartingale if and only if there is a martingale $\{(Z_n, \mathcal{F}_n)\}_{n=1}^{\infty}$ and a nondecreasing previsible process $\{A_n\}_{n=1}^{\infty}$ with $A_1 = 0$ such that $X_n = Z_n + A_n$ for all n. The decomposition is unique (a.s.).*

Proof. For the "if" direction, notice that

$$\mathrm{E}(Z_{n+1} + A_{n+1}|\mathcal{F}_n) = Z_n + A_{n+1} \geq Z_n + A_n = X_n.$$

For the "only if" dirction, Define $A_1 = 0$ and

$$A_n = \sum_{k=2}^{n} \left(\mathrm{E}(X_k|\mathcal{F}_{k-1}) - X_{k-1}\right),$$

for $n > 1$. Also, define $Z_n = X_n - A_n$. Because $\mathrm{E}(X_k|\mathcal{F}_{k-1}) \geq X_{k-1}$ for all $k > 1$, we have $A_n \geq A_{n-1}$ for all $k > 1$, so $\{A_n\}_{n=1}^{\infty}$ is nondecreasing. Also, $\mathrm{E}(X_k|\mathcal{F}_{k-1})$ is $\mathcal{F}_{n-1}/\mathcal{B}^1$-measurable for all $1 < k \leq n$, so $\{A_n\}_{n=1}^{\infty}$ is previsible. Finally, notice that

$$\begin{aligned}
\mathrm{E}(Z_{n+1}|\mathcal{F}_n) &= \mathrm{E}(X_{n+1}|\mathcal{F}_n) - A_{n+1} \\
&= \mathrm{E}(X_{n+1}|\mathcal{F}_n) - \sum_{k=2}^{n+1} [\mathrm{E}(X_k|\mathcal{F}_{k-1}) - X_{k-1}] \\
&= X_n - \sum_{k=2}^{n} [\mathrm{E}(X_k|\mathcal{F}_{k-1}) - X_{k-1}] = Z_n,
\end{aligned}$$

so Z_n is a martingale.

For uniqueness, suppose that $X_n = Y_n + W_n$ is another decomposition so that Y_n is a martingale and W_n is previsible. Then write

$$\begin{aligned}
\sum_{k=2}^{n} [\mathrm{E}(X_k|\mathcal{F}_{k-1}) - X_{k-1}] &= \sum_{k=2}^{n} [\mathrm{E}(Y_k + W_k|\mathcal{F}_{k-1}) - X_{k-1}] \\
&= \sum_{k=2}^{n} (Y_{k-1} + W_k - X_{k-1}) \\
&= \sum_{k=2}^{n} (W_k - W_{k-1}) = W_k.
\end{aligned}$$

It follows that $W_k = A_k$ a.s., hence $Y_k = Z_k$ a.s.$\square$

The previsible process in Theorem 302 is called the *compensator* for the submartingale.

Stopping Times. Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $\{\mathcal{F}_n\}_{n=1}^{\infty}$ be a filtration.

DEFINITION 303. A positive[4] extended integer valued random variable τ is called a *stopping time* with respect to the filtration if $\{\tau = n\} \in \mathcal{F}_n$ for all finite n. A special σ-field, $\mathcal{F}_\tau$ is defined by

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \leq k\} \in \mathcal{F}_k, \text{ for all finite } k\}.$$

If $\{X_n\}_{n=1}^{\infty}$ is adapted to the filtration and if $\tau < \infty$ a.s., then X_τ is defined as $X_{\tau(\omega)}(\omega)$. (Define X_τ equal to some arbitrary random variable X_∞ for $\tau = \infty$.)

EXAMPLE 304. Let $\{X_n\}_{n=1}^{\infty}$ be adapted to the filtration and let $\tau = k_0$, a constant. Then $\{\tau = n\}$ is either Ω or $\varnothing$ and it is in every $\mathcal{F}_n$, so τ is a stopping time. Also,

$$A \cap \{\tau \leq k\} = \begin{cases} A & \text{if } k_0 \leq k, \\ \varnothing & \text{if } k_0 > k. \end{cases}$$

So $A \cap \{\tau \leq k\} \in \mathcal{F}_k$ for all finite k if and only if $A \in \mathcal{F}_{k_0}$. So $\mathcal{F}_\tau = \mathcal{F}_{k_0}$.

EXAMPLE 305. Let $\{X_n\}_{n=1}^{\infty}$ be adapted to the filtration. Let B be a Borel set and let $\tau = \inf\{n : X_n \in B\}$. As usual, $\inf \varnothing = \infty$. For each finite n,

$$\{\tau = n\} = \{X_n \in B\} \bigcap_{k<n} \{X_k \in B^C\} \in \mathcal{F}_n.$$

So, τ is a stopping time.

We can show that τ and X_τ are both $\mathcal{F}_\tau$ measurable. For example, to show that X_τ is $\mathcal{F}_\tau$ measurable, we must show that, for all $B \in \mathcal{B}^1$ $X_\tau^{-1}(B) \in \mathcal{F}$ and for all $1 \leq k < \infty$, $\{\tau \leq k\} \cap X_\tau^{-1}(B) \in \mathcal{F}_k$. Now,

$$X_\tau^{-1}(B) = \bigcup_{k=1}^{\infty} \left(\{\tau = k\} \cap [X_k^{-1}(B)]\right) \cup \left(\{\tau = \infty\} \cap X_\infty^{-1}(B)\right) \in \mathcal{F}.$$

This shows that X_τ is $\mathcal{F}$-measurable. Next, fix k and write

$$\{\tau \leq k\} \cap X_\tau^{-1}(B) = \bigcup_{j=1}^{k} \left[X_j^{-1}(B) \cap \{\tau = j\}\right] \in \mathcal{F}_k.$$

This proves that X_τ is $\mathcal{F}_\tau$ measurable. Suppose that τ_1 and τ_2 are two stopping times such that $\tau_1 \leq \tau_2$. Let $A \in \mathcal{F}_{\tau_1}$. Since $A \cap \{\tau_2 \leq k\} = A \cap \{\tau_1 \leq k\} \cap \{\tau_2 \leq k\}$ for every event A, it follows that $A \cap \{\tau_2 \leq k\} \in \mathcal{F}_k$ and $A \in \mathcal{F}_{\tau_2}$. Hence $\mathcal{F}_{\tau_1} \subseteq \mathcal{F}_{\tau_2}$. As an example, let τ be an arbitrary stopping time (not necessarily finite a.s.) and define $\tau_k = \min\{k, \tau\}$ for finite k. Then τ_k is a finite stopping time with $\tau_k \leq \tau$. Hence X_{τ_k} is $\mathcal{F}_{\tau_k}$ measurable for each k and so X_{τ_k} is $\mathcal{F}_\tau$ measurable. Similarly, $\tau_k \leq k$ so that $\mathcal{F}_{\tau_k} \subseteq \mathcal{F}_k$ and X_{τ_k} is $\mathcal{F}_k$ measurable.

[4]If your filtration starts at $n = 0$, you can allow stopping times to be nonnegative valued. Indeed, if your filtration starts at an arbitrary integer k, then a stopping time can take any value from k on up. There is a trivial extension of every filtration to one lower subscript. For example, if we start at $n = 1$, we can extend to $n = 0$ by defining $\mathcal{F}_0 = \{\Omega, \varnothing\}$. Every martingale can also be extended by defining $X_0 = \mathrm{E}(X_1)$. For the rest of the course, we will assume that the lowest possible value for a stopping time is 1.

Example 306. The gambler in Example 299 can try to build a stopping time into a gambling system. For example, let $\tau = \min\{n : Z_n \geq Y_0 + x\}$ for some integer $x > 0$. This would seem to guarantee winning at least x. There are two possible drawbacks. One is that there may be positive probability that $\tau = \infty$. Even if $\tau < \infty$ a.s., it might require infinite resources to guarantee that we can survive until τ. For example, let $Y_0 = 0$ and let Y_n have equal probability of being 1 or -1 all n. So, we stop as soon as we have won x more than we have lost. If we modify the problem so that we have only finite resources (say k units) then this becomes the classic *gambler's ruin problem*. The probability of achieving $Z_n = x$ before $Z_n = -k$ is $k/(k + x)$, which goes to 1 as $k \to \infty$. So, if we have infinite resources, the probability is 1 that $\tau < \infty$, otherwise, we may never achieve the goal. If the probability of winning on each game is less than 1/2, then $P(\tau = \infty) > 0$.

Suppose that we start with a martingale $\{(X_n, \mathcal{F}_n)\}_{n=1}^{\infty}$ and a stopping time τ. We can define

$$X_n^* = \begin{cases} X_n & \text{if } n \leq \tau, \\ X_\tau & \text{if } n > \tau \end{cases} = X_{\min\{\tau,n\}}.$$

We can call this the *stopped martingale*. It turns out that $\{X_n^*\}_{n=1}^{\infty}$ is also a martingale relative to the filtration. First, note that $X_{\min\{\tau,n\}}$ is $\mathcal{F}_n$ measurable. Next, notice that

$$\begin{aligned} \mathrm{E}(|X_n^*|) &= \sum_{k=1}^{n-1} \int_{\{\tau=k\}} |X_k| dP + \int_{\{\tau \geq n\}} |X_n| dP \\ &\leq \sum_{k=1}^{n} \mathrm{E}(|X_k|) < \infty. \end{aligned}$$

Finally, let $A \in \mathcal{F}_n$. Then $A \cap \{\tau > n\} \in \mathcal{F}_n$, so

$$\int_{A \cap \{\tau > n\}} X_{n+1} dP = \int_{A \cap \{\tau > n\}} X_n dP,$$

because $X_n = \mathrm{E}(X_{n+1}|\mathcal{F}_n)$. It now follows that

$$\begin{aligned} \int_A X_{n+1}^* dP &= \int_{A \cap \{\tau > n\}} X_{n+1} dP + \int_{A \cap \{\tau \leq n\}} X_\tau dP \\ &= \int_{A \cap \{\tau > n\}} X_n dP + \int_{A \cap \{\tau \leq n\}} X_\tau dP \\ &= \int_A X_n^* dP. \end{aligned}$$

It follows that $X_n^* = \mathrm{E}(X_{n+1}^*|\mathcal{F}_n)$ and the stopped martingale is also a martingale. Notice that $\lim_{n\to\infty} X_n^* = X_\tau$ a.s., if $\tau < \infty$ a.s.

Example 307. Consider the stopping time in Example 306 with $x = 1$. That is τ is the first time that a gambler, betting on iid fair coin flips, wins 1 more than he/she has lost. This $\tau < \infty$ a.s. It follows that $\lim_{n\to\infty} X_n^* = X_\tau$ a.s. However, $\mathrm{E}(X_n^*) = 0$ for all n while $\mathrm{E}(X_\tau) = 1$ because $X_\tau = 1$ a.s.

Optional Sampling. Let $\{(X_n, \mathcal{F}_n)\}_{n=1}$ be a martingale. Consider a sequence of a.s. finite stopping times $\{\tau_n\}_{n=1}^{\infty}$ such that $1 \le \tau_j \le \tau_{j+1}$ for all j. Then we can construct $\{(X_{\tau_n}, \mathcal{F}_{\tau_n})\}_{n=1}^{\infty}$ and ask whether or not it is a martingale. In general, an unpleasant integrability condition is needed to prove this. We shall do a simplified case.

THEOREM 308. (OPTIONAL SAMPLING THEOREM) *Let $\{(X_n, \mathcal{F}_n)\}_{n=1}^{\infty}$ be a (sub)martingale. Suppose that for each n, there is a finite constant M_n such that $\tau_n \le M_n$ a.s. Then $\{(X_{\tau_n}, \mathcal{F}_{\tau_n})\}_{n=1}^{\infty}$ is a (sub)martingale.*

The proof is given in a separate document.

The unpleasant integrability condition that can replace $P(\tau_n \le M_n) = 1$ is the following: For every n,

- $P(\tau_n < \infty) = 1$,
- $\mathrm{E}(|X_{\tau_n}|) < \infty$, and
- $\liminf_{m\to\infty} \mathrm{E}(|X_m| I_{(m,\infty)}(\tau_n)) = 0$.

Because we can use a constant stopping time to stop a martingale, it follows that martingale theorems will apply to finite sequences of random variables as well as infinite sequences.

Martingale Convergence. The upcrossing lemma says that a submartingale cannot cross a fixed nondegenerate interval very often with high probability. If the submartingale were to cross an interval infinitely often, then its lim sup and lim inf would have to be different and it couldn't converge.

LEMMA 309. (UPCROSSING LEMMA) *Let $\{(X_k, \mathcal{F}_k)\}_{k=1}^{n}$ be a submartingale. Let $r < q$, and define V to be the number of times that the sequence $X_1, \ldots, X_n$ crosses from below r to above q. Then*

$$\mathrm{E}(V) \le \frac{1}{q-r}\left(\mathrm{E}|X_n| + |r|\right). \tag{310}$$

We will only give an outline of the proof of Lemma 309. Let $Y_k = \max\{0, X_k - r\}$. Then V is the number of times that Y_k moves from 0 to above $q - r$, and $\{(Y_k, \mathcal{F}_k)\}_{k=1}^{\infty}$ is a submartingale. Figure 1 shows an example of the path of Y_k indicating those times that it is

FIG. 1. *Step in the proof of Lemma 309.*

crossing up. It is easy to see that V is at most the sum of the upcrossing increments divided by $q - r$. That is,

$$V \le \frac{1}{q-r}\sum_{k=2}^{n}(Y_k - Y_{k-1}) I_{E_k},$$

where E_k is the event that the path is crossing up at time k. Notice that $E_k \in \mathcal{F}_{k-1}$ for all k. Hence, for each $k \geq 2$,

$$\mathrm{E}([Y_k - Y_{k-1}]I_{E_k}) = \int_{E_k} (Y_k - Y_{k-1})dP = \int_{E_k} [\mathrm{E}(Y_k|\mathcal{F}_{k-1}) - Y_{k-1}]dP.$$

Because $\mathrm{E}(Y_k|\mathcal{F}_{k-1}) - Y_{k-1} \geq 0$ a.s. by the submartingale property, we can expand the integral from E_k to all of Ω to get

$$\mathrm{E}([Y_k - Y_{k-1}]I_{E_k}) \leq \int [\mathrm{E}(Y_k|\mathcal{F}_{k-1}) - Y_{k-1}]dP = \mathrm{E}(Y_k - Y_{k-1}).$$

It follows that $\mathrm{E}(V) \leq \mathrm{E}(Y_n) - \mathrm{E}(Y_1) \leq \mathrm{E}(Y_n)$ because $Y_1 \geq 0$. Because $\max\{0, x\}$ is a convex function of x, $\mathrm{E}(Y_n) \leq \mathrm{E}(|X_n|) + r$. The full proof is in another course document.

THEOREM 311. (MARTINGALE CONVERGENCE THEOREM) *Let* $\{(X_n, \mathcal{F}_n)\}_{n=1}^{\infty}$ *be a submartingale such that* $\sup_n \mathrm{E}|X_n| < \infty$. *Then* $X = \lim_{n\to\infty} X_n$ *exists a.s. and* $\mathrm{E}|X| < \infty$.

PROOF. Let $X^* = \limsup_{n\to\infty} X_n$ and $X_* = \liminf_{n\to\infty} X_n$. Let $B = \{\omega : X_*(\omega) < X^*(\omega)\}$. We will prove that $P(B) = 0$. We can write

$$B = \bigcup_{r<q,\ r,q \text{ rational}} \{\omega : X^*(\omega) \geq q > r \geq X_*(\omega)\}.$$

Now, $X^*(\omega) > q > r \geq X_*(\omega)$ if and only if the values of $X_n(\omega)$ cross from being below r to being above q infinitely often. For fixed r and q, we now prove that this has probability 0; hence $P(B) = 0$. Let V_n equal the number of times that $X_1, \ldots, X_n$ cross from below r to above q. According to Lemma 309,

$$\sup_n \mathrm{E}(V_n) \leq \frac{1}{q-r}\left(\sup_n \mathrm{E}(|X_n|) + |r|\right) < \infty.$$

The number of times the values of $\{X_n(\omega)\}_{n=1}^{\infty}$ cross from below r to above q equals $\lim_{n\to\infty} V_n(\omega)$. By the monotone convergence theorem,

$$\infty > \sup_n \mathrm{E}(V_n) = \mathrm{E}(\lim_{n\to\infty} V_n).$$

It follows that $P(\{\omega : \lim_{n\to\infty} V_n(\omega) = \infty\}) = 0$.

Since $P(B) = 0$, we have that $X = \lim_{n\to\infty} X_n$ exists a.s. Fatou's lemma says $\mathrm{E}(|X|) \leq \liminf_{n\to\infty} \mathrm{E}(|X_n|) \leq \sup_n \mathrm{E}(|X_n|) < \infty$. $\square$

EXAMPLE 312. For the Lévy martingale of Example 298,

$$\mathrm{E}(|X_n|) = \mathrm{E}(|\mathrm{E}[X|\mathcal{F}_n]|) \leq \mathrm{E}\mathrm{E}(|X||\mathcal{F}_n) = \mathrm{E}(|X|) < \infty,$$

for all n, so the martingale converges. In Theorem 316, we can say even more about the limit.

Example 313. For the random walk martingale of Example 292, if the Y_n's are iid with finite variance σ^2, then $X_n/\sqrt{n}$ converges in distribution so X_n can't converge a.s. Indeed, the Markov inequality says that

$$\frac{\mathrm{E}(|X_n|)}{\sqrt{n}c} \geq P(|X_n| > c\sqrt{n}) \to 2\left[1 - \Phi\left(\frac{c}{\sigma}\right)\right],$$

for each positive c. So, eventually $\mathrm{E}(|X_n|) \geq \sqrt{n}[1 - \Phi(c/\sigma)]$ and $\lim_{n\to\infty} \mathrm{E}(|X_n|) = \infty$.

Example 314. For the martingale of Example 292, if $\sum_{n=1}^{\infty} \mathrm{Var}(Y_n) < \infty$, then Theorem 202 already told us that the sum converges a.s.

We need the following result before we can identify the limit of a Lévy martingale. The proof is given in another course document.

Lemma 315. *Let $\{\mathcal{F}_n\}_{n=1}^{\infty}$ be a sequence of σ-fields. Let $\mathrm{E}(|X|) < \infty$. Define $X_n = \mathrm{E}(X|\mathcal{F}_n)$. Then $\{X_n\}_{n=1}^{\infty}$ is a uniformly integrable sequence.*

Theorem 316. (Lévy's theorem) *Let $\{\mathcal{F}_n\}_{n=1}^{\infty}$ be an increasing sequence of σ-fields. Let $\mathcal{F}_\infty$ be the smallest σ-field containing all of the $\mathcal{F}_n$'s. Let $\mathrm{E}(|X|) < \infty$. Define $X_n = \mathrm{E}(X|\mathcal{F}_n)$ and $X_\infty = \mathrm{E}(X|\mathcal{F}_\infty)$. Then $\lim_{n\to\infty} X_n = X_\infty$, a.s.*

The proof of Theorem 316 is in another course document.

Lemma 317. *Let $\{(X_n, \mathcal{F}_n)\}_{n=1}^{\infty}$ be a nonnegative supermartingale. Then X_n converges a.s. to a random variable with finite mean.*

Proof. Let $Y_n = -X_n$. Then $\{(Y_n, \mathcal{F}_n)\}_{n=1}^{\infty}$ is a submartingale.

$$\mathrm{E}(|Y_n|) = \mathrm{E}(X_n) = \mathrm{E}[\mathrm{E}(X_n|\mathcal{F}_{n-1})] \leq \mathrm{E}(X_{n-1}).$$

It follows that $\mathrm{E}(|Y_n|) \leq \mathrm{E}(X_1) < \infty$ for all n, so Theorem 311 applies and Y_n converges a.s. Trivially $-Y_n = X_n$ also converges. □

Reversed Martingales.

Definition 318. For $n = -1, -2, \ldots$, let sub-σ-field's $\mathcal{F}_{n-1} \subseteq \mathcal{F}_n$, suppose that X_n is $\mathcal{F}_n$ measurable, $\mathrm{E}(|X_n|) < \infty$, and $\mathrm{E}(X_n|\mathcal{F}_{n-1}) = X_{n-1}$. Then $\{(X_n, \mathcal{F}_n)\}_{n=-1}^{-\infty}$ is a *reversed martingale.*

An equivalent way to think about reversed martingales is through a decreasing sequence of σ-field's $\{\mathcal{F}_n\}_{n=1}^{\infty}$ such that $\mathcal{F}_{n+1} \subseteq \mathcal{F}_n$ for $n \geq 1$. The proofs of the next two theorems are similar to the corresponding theorems for forward martingales.

Theorem 319. (Reversed martingale convergence theorem) *If $\{(X_n, \mathcal{F}_n)\}_{n=-1}^{-\infty}$ is a reversed martingale, then $X = \lim_{n\to-\infty} X_n$ exists a.s. and $\mathrm{E}(X) = \mathrm{E}(X_{-1})$.*

PROOF. Just as in the proof of Theorem 311, we let V_n be the number of times that the finite sequence $X_n, X_{n+1}, \ldots, X_{-1}$ crosses from below a rational r to above another rational q (for $n < 0$). Lemma 309 says that

$$\mathrm{E}(V_n) \leq \frac{1}{q-r}\left(\mathrm{E}(|X_{-1}|) + |r|\right) < \infty.$$

As in the proof of Theorem 311, it follows that $X = \lim_{n\to-\infty} X_n$ exists with probability 1. Since $X_n = \mathrm{E}(X_{-1}|\mathcal{F}_n)$ for each $n < -1$, Lemma 315 says that

$$\mathrm{E}(X) = \lim_{n\to-\infty} \mathrm{E}(X_n) = \mathrm{E}(X_{-1}). \quad \square$$

Notice that reversed martingales are all of the Lévy type. Not surprisingly, there is a version of Lévy's theorem 316 for reversed martingales. We state it in terms of decreasing σ-field's.

THEOREM 320. *Let $\{\mathcal{F}_n\}_{n=1}^{\infty}$ be a decreasing sequence of σ-fields. Let $\mathcal{F}_\infty = \bigcap_{n=1}^{\infty} \mathcal{F}_n$. Let $\mathrm{E}(|X|) < \infty$. Define $X_n = \mathrm{E}(X|\mathcal{F}_n)$ and $X_\infty = \mathrm{E}(X|\mathcal{F}_\infty)$. Then $\lim_{n\to\infty} X_n = X_\infty$ a.s.*

PROOF. It is easy to see that $\{(X_{-n}, \mathcal{F}_{-n})\}_{n=-1}^{-\infty}$ is a reversed martingale and that $\mathrm{E}(|X_1|) < \infty$. By Theorem 319, it follows that $\lim_{n\to-\infty} X_{-n} = Y$ exists and is finite a.s. To prove that $Y = X_\infty$ a.s., note that $X_\infty = \mathrm{E}(X_1|\mathcal{F}_\infty)$ since $\mathcal{F}_\infty \subseteq \mathcal{F}_1$. So, we must show that $Y = \mathrm{E}(X_1|\mathcal{F}_\infty)$. Let $A \in \mathcal{F}_\infty$. Then

$$\int_A X_n(\omega)dP(\omega) = \int_A X_1(\omega)dP(\omega),$$

since $A \in \mathcal{F}_n$ and $X_n = \mathrm{E}(X_1|\mathcal{F}_n)$. Once again, using Lemma 315, it follows that

$$\lim_{n\to\infty} \int_A X_n(\omega)dP(\omega) = \int_A Y(\omega)dP(\omega) = \int_A X_1(\omega)dP(\omega),$$

hence $Y = \mathrm{E}(X_1|\mathcal{F}_\infty)$. $\square$

Theorem 320 allows us to prove a strong law of large numbers that is even more general than the usual version. The greater generality comes from the fact that it applies to sequences that are not necessarily independent.

Exchangeability. A sequence of random quantities $\{X_n\}_{n=1}^{\infty}$ is *exchangeable* if, for every n and all distinct $j_1, \ldots, j_n$, the joint distribution of $(X_{j_1}, \ldots, X_{j_n})$ is the same as the joint distribution of $(X_1, \ldots, X_n)$.

EXAMPLE 321. (CONDITIONALLY IID RANDOM QUANTITIES) Let $\{X_n\}_{n=1}^{\infty}$ be conditionally iid given a σ-field $\mathcal{C}$. Then $\{X_n\}_{n=1}^{\infty}$ is an exchangeable sequence. The result follows easily from the fact that

$$\mu_{X_{j_1},\ldots,X_{j_n}|\mathcal{C}} = \mu_{X_1,\ldots,X_n|\mathcal{C}}, \quad \text{a.s.}$$

EXAMPLE 322. Let $\{X_n\}_{n=1}^{\infty}$ be Bernoulli random variables such that

$$P(X_1 = x_1, \ldots, X_n = x_n) = \frac{1}{(n+1)\binom{n}{y}},$$

where $y = \sum_{j=1}^{n} x_j$. One can show that this specifies consistent joint distributions. One can also check that the X_n's are not independent.

$$\begin{aligned} P(X_1 = 1) &= \frac{1}{2}, \\ P(X_1 = 1, X_2 = 1) &= \frac{1}{3} \neq \left(\frac{1}{2}\right)^2. \end{aligned}$$

THEOREM 323. (STRONG LAW OF LARGE NUMBERS) *Let $\{X_n\}_{n=1}^{\infty}$ be an exchangeable sequence of random variables with finite mean. Then $\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} X_j$ exists a.s. and has mean equal to $\mathrm{E}(X_1)$. If, the X_j's are independent, then the limit equals $\mathrm{E}(X_1)$ a.s.*

PROOF. Define $Y_n = \frac{1}{n}\sum_{j=1}^{n} X_j$ and let $\mathcal{F}_n$ be the σ-field generated by all function of $(X_1, X_2, \ldots)$ that are invariant under permutations of the first n coordinates. (For example, Y_n is such a function.) Let $Z_n = \mathrm{E}(X_1|\mathcal{F}_n)$. Theorem 320 says that Z_n converges a.s. to $\mathrm{E}(X_1|\mathcal{F}_\infty)$, where $\mathcal{F}_\infty = \bigcap_{n=1}^{\infty}\mathcal{F}_n$. We prove next that $Z_n = Y_n$, a.s. Since Y_n is $\mathcal{F}_n$ measurable, we need only prove that, for all $A \in \mathcal{F}_n$, $\mathrm{E}(I_A Y_n) = \mathrm{E}(I_A X_1)$. Notice that I_A is a function of $X_1, X_2, \ldots$ that is invariant under permutations of $X_1, \ldots, X_n$ since it depends on $X_1, \ldots, X_n$ only through Y_n. That is, there are functions g and h such that

$$\begin{aligned} I_A(\omega) &= h(X_1(\omega), X_2(\omega), \ldots) \\ &= g(X_1(\omega) + \cdots + X_n(\omega), X_{n+1}(\omega), X_{n+2}(\omega), \ldots). \end{aligned} \tag{324}$$

For all $j = 1, \ldots, n$, $X_1 h(X_1, X_2, \ldots)$ has the same distribution as $X_j h(X_j, X_2, \ldots, X_{j-1}, X_1, X_{j+1}, \ldots)$. But

$$h(X_j, X_2, \ldots, X_{j-1}, X_1, X_{j+11}, \ldots) = h(X_1, X_2, \ldots) = I_A,$$

according to (324). Hence, for all $j = 1, \ldots, n$, $I_A X_j$ has the same distribution as $I_A X_1$. It follows that

$$\mathrm{E}(I_A X_1) = \frac{1}{n}\sum_{j=1}^{n} \mathrm{E}(I_A X_j) = \mathrm{E}(I_A Y_n).$$

Clearly $\mathrm{E}(X_1|\mathcal{F}_\infty)$ has mean $\mathrm{E}(X_1)$. If the X_n's are independent, then the limit, being measurable with respect to the tail σ-field, must be constant a.s., by Theorem 168 (Kolmogorov 0-1 law). The constant must equal the mean of the random variable, which is $\mathrm{E}(X_1)$. □

EXAMPLE 325. In Example 322, we know that Y_n converges a.s., hence it converges in distribution. We can compute the distribution of Y_n exactly: $P(Y_n = k/n) = 1/(n+1)$ for $k = 0, \ldots, n$. Hence, Y_n converges in distribution to uniform on the interval $[0, 1]$, which must be the distribution of the limit. The limit is not a.s. constant.

The rest of this material was not covered in class, but is left here for your information.

There is a very useful theorem due to deFinetti about exchangeable random quantities that relies upon the strong law of large numbers. To state the theorem, we need to recall the concept of "random probability measure" that was introduced in Example 173. Let $(\mathcal{X}, \mathcal{B})$ be a Borel space, and let $\mathcal{P}$ be the set of all probability measures on $(\mathcal{X}, \mathcal{B})$. We can think of $\mathcal{P}$ as a subset of the function space $[0,1]^{\mathcal{B}}$, hence it has a product σ-field. Recall that the product σ-field is the smallest σ-field such that for all $B \in \mathcal{B}$, the function $f_B : \mathcal{P} \to [0,1]$ defined by $f_B(Q) = Q(B)$ is measurable. These are the coordinate projection functions.

Example 326. (Empirical probability measure) Let $X_1, \ldots, X_n$ be random quantities taking values in $\mathcal{X}$. For each $B \in \mathcal{B}$, define $\mathbf{P}_n(\omega)(B) = \frac{1}{n}\sum_{j=1}^{n} I_B(X_j(\omega))$. For each ω, $\mathbf{P}_n(\omega)(\cdot)$ is clearly a probability measure, so $\mathbf{P}_n : \Omega \to \mathcal{P}$ is a function that we could prove was measurable, but that proof will not be given here. Theorem 323 says that $\mathbf{P}_n(\omega)(B)$ converges to $\mathrm{E}(I_B(X_1)|\mathcal{F}_\infty)(\omega)$ for all B and almost all ω. If we assume that the X_n's take values in a Borel space, then $\mathrm{E}(I_B(X_1)|\mathcal{F}_\infty) = \Pr(X_1 \in B|\mathcal{F}_\infty)$ is part of an rcd. This rcd plays an important roll in deFinetti's theorem.

DeFinetti's theorem says that a sequence of random quantities is exchangeable if and only if it is conditionally iid given a random probability measure, and that random probability measure is the limit of the empirical probability measures of $X_1, \ldots, X_n$. That is, Example 321 is essentially the only example of exchangeable sequences. The proof is given in another course document.

Theorem 327. (DeFinetti's theorem) *A sequence $\{X_n\}_{n=1}^{\infty}$ of random quantities is exchangeable if and only if $\mathbf{P}_n$ (the empirical probability measure of $X_1, \ldots, X_n$) converges a.s. to a random probability measure $\mathbf{P}$ and the X_n's are conditionally iid with distribution Q given $\mathbf{P} = Q$.*

Example 328. In Example 322, the empirical probability measure is equivalent to $Y_n = \sum_{k=1}^{n} X_k/n$, since Y_n is one minus the proportion of the observations less than or equal to 0. So $\mathbf{P}$ is equivalent to the limit of Y_n, the limit of relative frequency of 1's in the sequence. Conditional on the limit of relative frequency of 1's being x, the X_k's are iid with Bernoulli distribution with parameter x.

Optional Sampling Theorem

PROOF OF THEOREM 308. Without loss of generality, assume that $M_n \le M_{n+1}$ for every n. Since $\tau_n \le M_n$,

$$\mathrm{E}(|X_{\tau_n}|) = \sum_{k=1}^{M_n} \int_{\{\tau_n = k\}} |X_k| dP \le \sum_{k=1}^{M_n} \mathrm{E}(|X_k|) < \infty.$$

We already know that X_{τ_n} is $\mathcal{F}_{\tau_n}$ measurable. Let $A \in \mathcal{F}_{\tau_n}$. We need to show that $\int_A X_{\tau_{n+1}} dP (\ge) = \int_A X_{\tau_n} dP$. Write

$$\int_A [X_{\tau_{n+1}} - X_{\tau_n}] dP = \int_{A \cap \{\tau_{n+1} > \tau_n\}} [X_{\tau_{n+1}} - X_{\tau_n}] dP.$$

Next, for each $\omega \in \{\tau_{n+1} > \tau_n\}$, write

$$X_{\tau_{n+1}}(\omega) - X_{\tau_n}(\omega) = \sum_{\text{All } k \text{ such that } \tau_n(\omega) < k \le \tau_{n+1}(\omega)} [X_k(\omega) - X_{k-1}(\omega)].$$

The smallest k such that $\tau_n < k$ is $k = 2$, So,

$$\int_A [X_{\tau_{n+1}} - X_{\tau_n}] dP = \int_A \sum_{k=2}^{M_{n+1}} I_{\{\tau_n < k \le \tau_{n+1}\}} (X_k - X_{k-1}) dP.$$

Since $A \in \mathcal{F}_{\tau_n}$ and $\{\tau_n < k \le \tau_{n+1}\} = \{\tau_n \le k-1\} \cap \{\tau_{n+1} \le k-1\}^C$, it follows that

$$B_k = A \cap \{\tau_n < k \le \tau_{n+1}\} \in \mathcal{F}_{k-1},$$

for each k. So

$$\begin{aligned} \int_A [X_{\tau_{n+1}} - X_{\tau_n}] dP &= \sum_{k=2}^{M_{n+1}} \int_{B_k} (X_k - X_{k-1}) dP \\ (\ge) &= \sum_{k=2}^{M_{n+1}} \int_{B_k} [X_k - \mathrm{E}(X_k | \mathcal{F}_{k-1})] dP = 0. \end{aligned}$$

because $X_{k-1} (\le) = \mathrm{E}(X_k | \mathcal{F}_{k-1})$ and $B_k \in \mathcal{F}_{k-1}$. $\square$

Upcrossing Lemma

LEMMA 309. (UPCROSSING LEMMA) *Let $\{(X_k, \mathcal{F}_k)\}_{k=1}^n$ be a submartingale. Let $r < q$, and define V to be the number of times that the sequence $X_1, \ldots, X_n$ crosses from below r to above q. Then*

$$\mathrm{E}(V) \leq \frac{1}{q-r}\left(\mathrm{E}|X_n| + |r|\right). \tag{310}$$

PROOF. Let $Y_k = \max\{0, X_k - r\}$ for every k so that $\{(Y_k, \mathcal{F}_k)\}_{k=1}^n$ is a submartingale. Note that a consecutive set of $X_k(\omega)$ cross from below r to above q if and only if the corresponding consecutive set of $Y_k(\omega)$ cross from 0 to above $q-r$. Let $T_0(\omega) = 0$ and define T_m for $m = 1, 2, \ldots$ as

$$\begin{aligned} T_m(\omega) &= \inf\{k \leq n : k > T_{m-1}(\omega), Y_k(\omega) = 0\}, \text{ if } m \text{ is odd}, \\ T_m(\omega) &= \inf\{k \leq n : k > T_{m-1}(\omega), Y_k(\omega) \geq q - r\}, \text{ if } m \text{ is even}, \\ T_m(\omega) &= n+1, \text{ if the corresponding set above is empty.} \end{aligned}$$

Now $V(\omega)$ is one-half of the largest even m such that $T_m(\omega) \leq n$. Define, for $k = 1, \ldots, n$,

$$R_k(\omega) = \begin{cases} 1 & \text{if } T_m(\omega) < k \leq T_{m+1}(\omega) \text{ for } m \text{ odd}, \\ 0 & \text{otherwise.} \end{cases}$$

Then $(q-r)V(\omega) \leq \sum_{k=1}^n R_k(\omega)[Y_k(\omega) - Y_{k-1}(\omega)] = \hat{X}$, where $Y_0 \equiv 0$ for convenience. First, note that for all m and k, $\{T_m(\omega) \leq k\} \in \mathcal{F}_k$. Next, note that for every k,

$$\{\omega : R_k(\omega) = 1\} = \bigcup_{m \text{ odd}} \left(\{T_m \leq k-1\} \cap \{T_{m+1} \leq k-1\}^C\right) \in \mathcal{F}_{k-1}. \tag{329}$$

$$\begin{aligned} \mathrm{E}(\hat{X}) &= \sum_{k=1}^n \int_{\{\omega : R_k(\omega)=1\}} [Y_k(\omega) - Y_{k-1}(\omega)] dP(\omega) \\ &= \sum_{k=1}^n \int_{\{\omega : R_k(\omega)=1\}} [\mathrm{E}(Y_k|\mathcal{F}_{k-1})(\omega) - Y_{k-1}(\omega)] dP(\omega) \\ &\leq \sum_{k=1}^n \int [\mathrm{E}(Y_k|\mathcal{F}_{k-1})(\omega) - Y_{k-1}(\omega)] dP(\omega) \\ &= \sum_{k=1}^n [\mathrm{E}(Y_k) - \mathrm{E}(Y_{k-1})] = \mathrm{E}(Y_n), \end{aligned}$$

where the second equality follows from (329) and the inequality follows from the fact that $\{(Y_k, \mathcal{F}_k)\}_{k=1}^n$ is a submartingale. It follows that $(q-r)\mathrm{E}(V) \leq \mathrm{E}(Y_n)$. Since $\mathrm{E}(Y_n) \leq |r| + \mathrm{E}(|X_n|)$, it follows that (310) holds. $\square$

Lévy's Theorem

Lemma 315. *Let* $\{\mathcal{F}_n\}_{n=1}^{\infty}$ *be a sequence of* σ*-fields. Let* $\mathrm{E}(|X|) < \infty$. *Define* $X_n = \mathrm{E}(X|\mathcal{F}_n)$. *Then* $\{X_n\}_{n=1}^{\infty}$ *is a uniformly integrable sequence.*

Proof. Since $\mathrm{E}(X|\mathcal{F}_n) = \mathrm{E}(X^+|\mathcal{F}_n) - \mathrm{E}(X^-|\mathcal{F}_n)$, and the sum of uniformly integrable sequences is uniformly integrable, we will prove the result for nonnegative X. Let $A_{c,n} = \{X_n \geq c\} \in \mathcal{F}_n$. So $\int_{A_{c,n}} X_n(\omega)dP(\omega) = \int_{A_{c,n}} X(\omega)dP(\omega)$. If we can find, for every $\epsilon > 0$, a C such that $\int_{A_{c,n}} X(\omega)dP(\omega) < \epsilon$ for all n and all $c \geq C$, we are done. Define $\eta(A) = \int_A X(\omega)dP(\omega)$. We have $\eta \ll P$ and η is finite. Absolute continuity implies that for every $\epsilon > 0$ there exists δ such that $P(A) < \delta$ implies $\eta(A) < \epsilon$. By the Markov inequality,

$$P(A_{c,n}) \leq \frac{1}{c}\mathrm{E}(X_n) = \frac{1}{c}\mathrm{E}(X),$$

for all n. Let $C = 2\mathrm{E}(X)/\delta$. Then $c \geq C$ implies $P(A_{c,n}) < \delta$ for all n, so $\eta(A_{c,n}) < \epsilon$ for all n. $\square$

Theorem 316. (Lévy's theorem) *Let* $\{\mathcal{F}_n\}_{n=1}^{\infty}$ *be an increasing sequence of* σ*-fields. Let* $\mathcal{F}_\infty$ *be the smallest* σ*-field containing all of the* $\mathcal{F}_n$*'s. Let* $\mathrm{E}(|X|) < \infty$. *Define* $X_n = \mathrm{E}(X|\mathcal{F}_n)$ *and* $X_\infty = \mathrm{E}(X|\mathcal{F}_\infty)$. *Then* $\lim_{n\to\infty} X_n = X_\infty$, *a.s.*

Proof. By Lemma 315, $\{X_n\}_{n=1}^{\infty}$ is a uniformly integrable sequence. Let Y be the limit of the martingale guaranteed by Theorem 311. Since Y is a limit of functions of the X_n, it is measurable with respect to $\mathcal{F}_\infty$. It follows from uniform integrability that for every event A, $\lim_{n\to\infty} \mathrm{E}(X_n I_A) = \mathrm{E}(Y I_A)$. Next, note that, for every m and $A \in \mathcal{F}_m$,

$$\begin{aligned}
\int_A Y dP &= \lim_{n\to\infty} \int_A \mathrm{E}(X|\mathcal{F}_n)dP \\
&= \lim_{n\to\infty} \int_A X_n dP \\
&= \int_A X dP,
\end{aligned}$$

where the last equality follows from the fact that $A \in \mathcal{F}_n$ for all $n \geq m$ so $\int_A X_n dP = \int_A X dP$ because $X_n = \mathrm{E}(X|\mathcal{F}_n)$. Since $\int_A Y dP = \int_A X dP$ for all $A \in \mathcal{F}_m$ for all m, it holds for all A in the field $\mathcal{F} = \bigcup_{n=1}^{\infty} \mathcal{F}_n$. Since $|X|$ is integrable and $\mathcal{F}$ is a field, we can conclude that the equality holds for all $A \in \mathcal{F}_\infty$, the smallest σ-field containing $\mathcal{F}$. The equality $\mathrm{E}(X I_A) = \mathrm{E}(Y I_A)$ for all $A \in \mathcal{F}_\infty$ together with the fact that Y is $\mathcal{F}_\infty$ measurable is precisely what it means to say that $Y = \mathrm{E}(X|\mathcal{F}_\infty) = X_\infty$. $\square$

Solutions to Exercises

Exercise 13: For notational convenience, define $\mathcal{F} \equiv \cup_{n=1}^{\infty} \mathcal{F}_n$.

For the first part, it is clear that $\Omega \in \mathcal{F}$ and that $\mathcal{F}$ is closed under complements. If $A \in \mathcal{F}$ and $B \in \mathcal{F}$, then for some n, $A \in \mathcal{F}_n$ and for some m, $B \in \mathcal{F}_m$. Thus, $A \cup B \in \mathcal{F}_k \subset \mathcal{F}$ for $k \geq \max\{n, m\}$.

For the second part, consider the given example and define the set A_n to be $\{n\}$ if n is even, and to be $\varnothing$ if n is odd. Clearly, $A_n \in \mathcal{F}_n \subset \mathcal{F}$ for each n. But, $\cup_{n=1}^{\infty} A_n \notin \mathcal{F}$. To see this, consider that $\cup_{n=1}^{\infty} A_n$ is the set of all positive, even integers. The class $\mathcal{F}_n$ includes the set of positive, even integers less than or equal to n. While it is true that $n \to \infty$, **there is no $\mathcal{F}_n$ which has $\cup_{n=1}^{\infty} A_n$ as a member,** so it cannot be in $\mathcal{F}$.

Exercise 17: Examples include:

1. The set of all closed intervals,
2. All intervals $[a, b)$ where $a, b \in \mathbb{R}$,
3. All intervals (a, ∞) where $a \in \mathbb{R}$,
4. All intervals $[a, \infty)$ where $a \in \mathbb{R}$,
5. All intervals $(-\infty, b)$ where $b \in \mathbb{R}$,
6. All intervals $(-\infty, b]$ where $b \in \mathbb{R}$.

Exercise 18: Yes. See, for example, page 45 in Billingsley or Exercise 6 on page 34 of Ash and Doleans-Dade.

Exercise 22: If $\Omega = \mathbb{Z}$, $\mathcal{F} = 2^{\Omega}$ and μ is counting measure, then let $A_1 = \{0\}$ and $A_k = \{-(k-1), k-1\}$ for $k > 1$. It is impossible to construct a countable sequence of sets, each with a finite number of elements, whose union is an uncountable set.

Exercise 27: Not here yet...

Exercise 30: First, recall that

$$\limsup_{n \to \infty} x_n = \inf_n \sup_{k \geq n} x_k$$

and

$$\liminf_{n \to \infty} x_n = \sup_n \inf_{k \geq n} x_k.$$

Then we can write

$$I_{\limsup_{n \to \infty} A_n} = \limsup_{n \to \infty} \, I_{A_n}$$

and

$$I_{\liminf_{n\to\infty} A_n} = \liminf_{n\to\infty} I_{A_n}.$$

EXERCISE 31:

$$\limsup_{n\to\infty} A_n = (-1,1] \quad \text{and} \quad \liminf_{n\to\infty} A_n = \{0\}$$

EXERCISE 33: Not here yet...

EXERCISE 35: Let $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B}^1)$, let μ be Lebesgue measure, and let $A_n = (n, \infty)$. Then $\mu(A_n) = \infty$ for all n but $\lim_{n\to\infty} A_n = \varnothing$.

EXERCISE 41: For Proposition 39, it is clear that $\Omega \in \mathcal{C}$ and that $\mathcal{C}$ is closed under complements since $\mathcal{C}$ is a λ-system. Further, if $A_1, A_2, \ldots \in \mathcal{C}$, then

$$\bigcup_{i=1}^{n} A_i = A_1 \cup \left(A_2 \cap A_1^C\right) \cup \left(A_3 \cap A_1^C \cap A_2^C\right) \cup \cdots,$$

and $A_1^C \in \mathcal{C}$ since $\mathcal{C}$ is a λ-system, $A_2 \cap A_1^C \in \mathcal{C}$ since $\mathcal{C}$ is a π-system, $A_1 \cup \left(A_2 \cap A_1^C\right) \in \mathcal{C}$ since $\mathcal{C}$ is a λ-system, and so forth.

For Proposition 40, note that $A \cap B^C = (A^C \cup (A \cap B))^C$.

EXERCISE 44: If we define $\mathcal{C}$ to be the class of subsets of $\Omega = \mathbb{R}$ of the form $(-\infty, a]$ with $a \in \mathbb{R}$, then $\mathcal{C}$ is a π-system and $\sigma(\mathcal{C}) = \mathcal{B}^1$. Also note that P is σ-finite since it is a probability measure. Thus, there is a unique extension of P from $\mathcal{C}$ to $\mathcal{B}^1$.

EXERCISE 45: If we define the value of $P(\{b\})$ we will define the measure on the σ-field generated by the sets $\{a, b\}$ and $\{b, c\}$. This is the case since the class of sets over which P is defined is now a π-system.

EXERCISE 51:

First, we show that μ^* extends μ. We only need to show that $\mu^*(A) = \mu(A)$ for all $A \in \mathcal{C}$. Clearly, $\mu^*(A) \le \mu(A)$ for all $A \in \mathcal{C}$. To show the reverse inequality, let $\{A_i\}_{i=1}^{\infty}$ be disjoint elements of $\mathcal{C}$ such that $A \subseteq \bigcup_{i=1}^{\infty} A_i$, and let $B_i = A_i \cap A$ so that $A = \bigcup_{i=1}^{\infty} B_i$. Countable additivity of μ gives us

$$\mu(A) = \sum_{i=1}^{\infty} \mu(B_i) \le \sum_{i=1}^{\infty} \mu(A_i).$$

Since every sum in (52) is a case of the rightmost sum in the equation immediately above, we have $\mu(A) \le \mu^*(A)$.

Next, we show that μ^* is monotone and subadditive. Clearly, $B_1 \subseteq B_2$ implies $\mu^*(B_1) \le \mu^*(B_2)$. It is also easy to see that $\mu^*(B_1 \cup B_2) \le \mu^*(B_1) + \mu^*(B_2)$ for all $B_1, B_2 \in 2^{\Omega}$. In fact, if $\{B_n\}_{n=1}^{\infty} \in 2^{\Omega}$, then $\mu^*(\bigcup_{i=1}^{\infty} B_i) \le \sum_{i=1}^{\infty} \mu^*(B_i)$. To see this, first choose $\epsilon > 0$. Note that there must exist a sequence of sets $A_{i1}, A_{i2}, \ldots \in \mathcal{C}$ such that $\mu^*(B_i) > \sum_{j=1}^{\infty} \mu(A_{ij}) - \epsilon/2^i$ and $B_i \subseteq \cup_{j=1}^{\infty} A_{ij}$ for each i. Thus,

$$\sum_{i=1}^{\infty} \mu^*(B_i) > \left[\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \mu(A_{ij})\right] - \epsilon \ge \mu^*\left(\bigcup_{i=1}^{\infty} B_i\right) - \epsilon$$

since the $\{A_{ij}\}$ cover $\cup_{i=1}^{\infty} B_i$. Thus, $\mu^*(\cup_{i=1}^{\infty} B_i) < \sum_{i=1}^{\infty} \mu^*(B_i) + \epsilon$ for any $\epsilon > 0$, the desired inequality holds.

Next, we show that $\mathcal{C} \subseteq \mathcal{A}$. Let $A \in \mathcal{C}$ and $C \in 2^{\Omega}$. Since μ^* is subadditive, we only need to show that $\mu^*(C) \ge \mu^*(C \cap A) + \mu^*(C \cap A^C)$. If $\mu^*(C) = \infty$, this is clearly true. So let $\mu^*(C) < \infty$. From the definition of μ^*, for every $\epsilon > 0$, there exists a collection $\{A_i\}_{i=1}^{\infty}$ of elements of $\mathcal{C}$ such that $\sum_{i=1}^{\infty} \mu(A_i) < \mu^*(C) + \epsilon$. Since $\mu(A_i) = \mu(A_i \cap A) + \mu(A_i \cap A^C)$ for every i, we have

$$\begin{aligned}
\mu^*(C) + \epsilon &> \sum_{i=1}^{\infty} \mu(A_i \cap A) + \sum_{i=1}^{\infty} \mu(A_i \cap A^C) \\
&\ge \mu^*(C \cap A) + \mu^*(C \cap A^C).
\end{aligned}$$

Since this is true for every $\epsilon > 0$, it must be that $\mu^*(C) \ge \mu^*(C \cap A) + \mu^*(C \cap A^C)$, hence $A \in \mathcal{A}$.

Next, we show that $\mathcal{A}$ is a field. It is clear that $\Omega \in \mathcal{A}$ and $A \in \mathcal{A}$ implies $A^C \in \mathcal{A}$ by the symmetry in the definition of $\mathcal{A}$. Let $A_1, A_2 \in \mathcal{A}$ and $C \in 2^{\Omega}$. We can write

$$\begin{aligned}
\mu^*(C) &= \mu^*(C \cap A_1) + \mu^*(C \cap A_1^C) \\
&= \mu^*(C \cap A_1) + \mu^*(C \cap A_1^C \cap A_2) + \mu^*(C \cap A_1^C \cap A_2^C) \\
&\ge \mu^*(C \cap [A_1 \cup A_2]) + \mu^*(C \cap [A_1 \cup A_2]^C),
\end{aligned}$$

where the two equalities follow from $A_1, A_2 \in \mathcal{A}$, and the inequality follows from the subadditivity of μ^*. Another application of subadditivity shows that $A_1 \cup A_2 \in \mathcal{A}$.

Next, we prove that μ^* is finitely additive on $\mathcal{A}$. If A_1, A_2 are disjoint elements of $\mathcal{A}$, then $A_1 = (A_1 \cup A_2) \cap A_1$ and $A_2 = (A_1 \cup A_2) \cap A_1^C$. It follows that

$$\mu^*(A_1 \cup A_2) = \mu^*(A_1) + \mu^*(A_2).$$

By induction, μ^* is finitely additive on $\mathcal{A}$.

Next, we prove that $\mathcal{A}$ is a σ-field. (We have already shown that $\mathcal{A}$ is a field.) Let $\{A_n\}_{n=1}^{\infty} \in \mathcal{A}$; then we can write $A = \bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$, where each $B_i \in \mathcal{A}$ and the B_i are disjoint. (This just makes use of complements and finite unions of elements of $\mathcal{A}$ being in $\mathcal{A}$.) Let $D_n = \bigcup_{i=1}^{n} B_i$ and $C \in 2^{\Omega}$. By the same argument we used in proving that μ^* is finitely additive, we have

$$\mu^*(C \cap [B_1 \cup B_2]) = \mu^*(C \cap B_1) + \mu^*(C \cap B_2).$$

A simple induction argument extends this to

$$\mu^*(C \cap D_n) = \sum_{i=1}^{n} \mu^*(C \cap B_i).$$

Since $A^C \subseteq D_n^C$ and $D_n \in \mathcal{A}$ for each n, we have

$$\begin{aligned} \mu^*(C) &= \mu^*(C \cap D_n) + \mu^*(C \cap D_n^C) \\ &\geq \mu^*(C \cap D_n) + \mu^*(C \cap A^C) \\ &= \sum_{i=1}^{n} \mu^*(C \cap B_i) + \mu^*(C \cap A^C). \end{aligned}$$

Since this is true for every n,

$$\begin{aligned} \mu^*(C) &\geq \sum_{i=1}^{\infty} \mu^*(C \cap B_i) + \mu^*(C \cap A^C) \\ &\geq \mu^*(C \cap A) + \mu^*(C \cap A^C), \end{aligned}$$

where the last inequality follows from subadditivity. The reverse inequality holds by subadditivity, so, $A \in \mathcal{A}$, and $\mathcal{A}$ is a σ-field.

Next, we show that μ^* is countably additive when restricted to $\mathcal{A}$. (We already proved that μ^* is finitely additive.) Let $A = \bigcup_{i=1}^{\infty} A_i$, where each $A_i \in \mathcal{A}$ and the A_i are disjoint. Since $\bigcup_{i=1}^{n} A_i \subseteq A$, we have, for every n, $\mu^*(A) \geq \sum_{i=1}^{n} \mu^*(A_i)$, which implies $\mu^*(A) \geq \sum_{i=1}^{\infty} \mu^*(A_i)$. By subadditivity, we get the reverse inequality, hence μ^* is countably additive on $\mathcal{A}$.

Next, we prove uniqueness. Suppose that μ' also extends μ to $\mathcal{A}$. Since $\mathcal{C}$ is a π-system and μ is σ-finite on $\mathcal{C}$, Theorem 43 implies that $\mu' = \mu$ on $\sigma(\mathcal{C})$. It is straightforward to prove that $\mu' = \mu$ on the completion of $\sigma(\mathcal{C})$. (See Theorem 1.3.8 in Ash and Dade.)

Finally, we prove completeness. Since we already proved that μ^* is monotone, we need only prove that $\mathcal{A}$ contains all B such that $\mu^*(B) = 0$. Let $\mu^*(B) = 0$. Then $\mu^*(B \cap C) = 0$ for all $C \in 2^{\Omega}$. By subadditivity and monotonicity, we have

$$\mu^*(C) \leq \mu^*(C \cap B) + \mu^*(C \cap B^C) = \mu^*(C \cap B^C) \leq \mu^*(C).$$

It follows that $B \in \mathcal{A}$.

Exercise 57: Nothing yet...

Exercise 61: The result is true if we can prove that $\mathcal{D} = \{A \in \mathcal{A} : f^{-1}(A) \in \mathcal{F}\}$ is a σ-field. Clearly $S \in \mathcal{D}$. Since inverse image commutes with complement, $A \in \mathcal{D}$ implies $A^C \in \mathcal{D}$. Since inverse image commutes with union, $A_n \in \mathcal{D}$ for all n implies $\bigcup_{n=1}^{\infty} A_n \in \mathcal{D}$. So, $\mathcal{D}$ is a σ-field.

Exercise 63: Nothing yet...

Exercise 64: Let f be a monotone increasing function. Then, for each a, there exists b such that $f^{-1}((-\infty, a])$ is an interval of the form $(-\infty, b)$ or $(-\infty, b]$. A similar result holds for decreasing functions. Hence, monotone functions are measurable.

Exercise 73: For part 1,

$$\{\omega : \limsup_n f_n \geq a\} = \bigcap_{m=1}^{\infty} \{f_n \geq a - 1/m, \text{ i.o.}\},$$

a measurable set. Similarly for part 2. For part 3, the set in question is the set where the difference of two measurable functions is 0, a measurable set. Part 4 is clear from the first three parts, but you can fill in the details in a homework problem. The functions in parts 1, 2, and 4 might be extended real-valued.

Exercise 78: Let $\Omega = \mathbb{Z}^2$ with the σ-field 2^{Ω}, while $S = \mathbb{Z}$ with σ-field 2^S. Let $f(x, y) = x$. Let μ be counting measure. Then, for each integer x, $f^{-1}(\{x\}) = \{(x, y) : y \in \mathbb{Z}\}$ and $\nu(\{x\}) = \infty$. Even though μ is σ-finite, ν is not.

Exercise 90: Let $g = I_{[a,b]} f$. Every lower Riemann sum is the integral of a simple function $\phi_1 \leq g$ and every upper Riemann sum is the integral of a simple function $\phi_2 \geq g$. It follows that

$$\int \phi_1 d\mu \leq \int g d\mu \leq \int \phi_2 d\mu.$$

But for each $\epsilon > 0$, ϕ_1 and ϕ_2 can be chosen so that $\int \phi_2 d\mu - \int \phi_1 d\mu < \epsilon$. So, limits of the upper and lower sums both equal $\int g d\mu$.

Exercise 95: Let X put probability 1/2 on c and $-c$.

Exercise 109: Clearly, ν is nonnegative and $\nu(\varnothing) = 0$, since $f I_{\varnothing} = 0$, a.e. $[\mu]$. Let $\{A_n\}_{n=1}^{\infty}$ be disjoint. For each n, define $g_n = f I_{A_n}$ and $f_n = \sum_{i=1}^{n} g_i$. Define $A = \bigcup_{n=1}^{\infty} A_n$. Then $0 \leq f_n \leq f I_A$, a.e. $[\mu]$ and f_n converges to $f I_A$, a.e. $[\mu]$. So, the monotone convergence theorem says that

$$\lim_{n \to \infty} \int f_n d\mu = \nu(A). \tag{330}$$

Also, $\nu(A_i) = \int g_i d\mu$, for each i. It follows from Theorem 100 that

$$\nu\left(\bigcup_{i=1}^{n} A_i\right) = \int f_n d\mu = \sum_{i=1}^{n} \int g_i d\mu = \sum_{i=1}^{n} \nu(A_i). \tag{331}$$

Take the limit as $n \to \infty$ of the second and last terms in (331) and compare to (330) to see that ν is countably additive.