Economics 204
Lecture Notes on Measure and Probability Theory

This is a slightly updated version of the Lecture Notes used in 204 in the summer of 2002. The measure-theoretic foundations for probability theory are assumed in courses in econometrics and statistics, as well as in some courses in microeconomic theory and finance. These foundations are not developed in the classes that use them, a situation we regard as very unfortunate. The experience in the summer of 2002 indicated that it is impossible to develop a good understanding of this material in the brief time available for it in 204. Accordingly, this material will not be covered in 204. This handout is being made available in the hope it will be of some help to students as they see measure-theoretic constructions used in other courses.

The Riemann Integral (the integral that is treated in freshman calculus) applies to continuous functions. It can be extended a little beyond the class of continuous functions, but not very far. It can be used to define the lengths, areas, and volumes of sets in $\mathbf{R}$, $\mathbf{R}^2$, and $\mathbf{R}^3$, provided those sets are reasonably nice, in particular not too irregularly shaped. In $\mathbf{R}^2$, the Riemann Integral defines the area under the graph of a function by dividing the x-axis into a collection of small intervals. On each of these small intervals, two rectangles are erected: one lies entirely inside the area under the graph of the function, while the other rectangle lies entirely outside the graph. The function is Riemann integrable (and its integral equals the area under its graph) if, by making the intervals sufficiently small, it is possible to make the sum of the areas of the outside rectangles arbitrarily close to the sum of the areas of the inside rectangles.

Measury theory provides a way to extend our notions of length, area, volume etc. to a much larger class of sets than can be treated using the Riemann Integral. It also provides a way to extend the Riemann Integral to Lebesgue integrable functions, a much larger class of functions than the continuous functions.

The fundamental conceptual difference between the Riemann and Lebesgue integrals is the way in which the partitioning is done. As noted above, the Riemann Integral partitions the domain of the function into small intervals. By contrast, the Lebesgue Integral partitions the *range* of the function into small intervals, then considers the set of points in the domain on which the value of the function falls into one of these intervals. Let $f : [0, 1] \to \mathbf{R}$.

Given an interval $[a,b) \subseteq \mathbf{R}$, $f^{-1}([a,b))$ may be a very messy set. However, as long as we can assign a "length" or "measure" $\mu\left(f^{-1}([a,b))\right)$ to this set, we know that the contribution of this set to the integral of f should be between $a\mu\left(f^{-1}([a,b))\right)$ and $b\mu\left(f^{-1}([a,b])\right)$. By making the partition of the range finer and finer, we can determine the integral of the function.

Clearly, the key to extending the Lebesgue Integral to as wide a class of functions as possible is to define the notion of "measure" on as wide a class of sets as possible. In an ideal world, we would be able to define the measure of *every* set; if we could do this, we could then define the Lebesgue integral of *every* function. Unfortunately, as we shall see, it is not possible to define a measure with nice properties on every subset of $\mathbf{R}$.

Measure theory is thus a *second best* exercise. We try to extend the notion of measure from our intuitive notions of length, area and volume to as large a class of *measurable* subsets of $\mathbf{R}$, $\mathbf{R}^2$, and $\mathbf{R}^3$ as possible. In order to be able to make use of measures and integrals, we need to know that the class of measurable sets is closed under certain types of operations. If we can assign a sensible notion of measure to a set, we ought to be able to assign a sensible notion to its complement. Probability and statistics focus on questions about convergence of sequences of random variables. In order to talk about convergence, we need to be able to assign measures to countable unions and countable intersections of measurable sets. Thus, we would like the collection of measurable sets to be a σ-algebra:

Definition 1 A *measure space* is a triple $(\Omega, \mathcal{B}, \mu)$, where

1. Ω is a set

2. $\mathcal{B}$ is a σ-algebra of subsets of Ω, i.e.

   (a) $\mathcal{B} \subset 2^{\Omega}$, i.e. $\mathcal{B}$ is a collection of subsets of Ω

   (b) $\emptyset, \Omega \in \mathcal{B}$

   (c) $B_n \in \mathcal{B}$, $n \in \mathbf{N} \Rightarrow \cup_{n \in \mathbf{N}} B_n \in \mathcal{B}$

   (d) $B \in \mathcal{B} \Rightarrow \Omega \setminus B \in \mathcal{B}$

3. μ is a nonnegative, countably additive set function on $\mathcal{B}$, i.e.

   (a) $\mu : \mathcal{B} \to \mathbf{R}_+ \cup \{\infty\}$

(b) $B_n \in \mathcal{B}$, $n \in \mathbf{N}$, $B_n \cap B_m = \emptyset$ if $n \neq m \Rightarrow \mu\left(\cup_{n\in\mathbf{N}} B_n\right) = \sum_{n\in\mathbf{N}} \mu(B_n)$

and $\mu(\emptyset) = 0$.

Remark 2 The definition of a σ-algebra is closely related to the properties of open sets in a metric space. Recall that the collection of open sets is closed under (1) arbitrary unions and (2) finite intersections; by contrast, a σ-algebra is closed under (1) countable unions and (2) countable intersections. Notice also that σ-algebras are closed under complements; the complement of an open set is closed, and generally not open, so closure under taking complements is not a property of the collection of open sets. The analogy between the properties of a σ-algebra and the properties of open sets in a metric space will be very useful in developing the Lebesgue integral. Recall that a function $f : X \to Y$ is continuous if and only if $f^{-1}(U)$ is open in X for every open set U in Y. Recall from the earlier discussion that the Lebesgue integral of a function f is defined by partitioning the range of the fuction f into small intervals, and summing up numbers of the form $a\mu\left(f^{-1}([a,b))\right)$; thus, we will need to know that $f^{-1}([a,b)) \in \mathcal{B}$. We will see in a while that a function $f : (\Omega, \mathcal{B}, \mu) \to (\Omega', \mathcal{B}')$ is said to be measurable if $f^{-1}(B') \in \mathcal{B}$ for every $B' \in \mathcal{B}'$. Thus, there is a close analogy between measurable functions and continuous functions. As you know from calculus, continuous functions on a closed interval can be integrated using the so-called Riemann integral; the Lebesgue integral extends the Riemann integral to all bounded measurable functions (and many unbounded measurable functions).

Remark 3 Countable additivity implies $\mu(\emptyset) = 0$ provided there is some set B with $\mu(B) < \infty$; thus, the requirement $\mu(\emptyset) = 0$ is imposed to rule out the pathological case in which $\mu(B) = \infty$ for all $B \in \mathcal{B}$.

Remark 4 If we have a finite collection $B_1, \ldots, B_k \in \mathcal{B}$ with $B_n \cap B_m = \emptyset$ if $n \neq m$, we can write $B_n = \emptyset$ for $n > k$, and obtain $\mu(B_1 \cup \cdots \cup B_k) = \mu(\cup_{n\in\mathbf{N}} B_n) = \sum_{n\in\mathbf{N}} \mu(B_n) = \sum_{n=1}^{k} \mu(B_k) + \sum_{n=k+1}^{\infty} \mu(\emptyset) = \sum_{n=1}^{k} \mu(B_k)$, so the measure is additive over finite collections of disjoint measurable sets.

Example 5 Suppose that we are given a partition $\{\Omega_\lambda : \lambda \in \Lambda\}$ of a set Ω, i.e. $\Omega = \cup_{\lambda\in\Lambda} \Omega_\lambda$ and $\lambda \neq \lambda' \Rightarrow \Omega_\lambda \cap \Omega_{\lambda'} \emptyset$. Then we can form a σ-algebra as follows: Let

$$\mathcal{B}_\Lambda = \{\cup_{\lambda\in C} \Omega_\lambda : C \subseteq \Lambda\}$$

In other words, $\mathcal{B}_\Lambda$ is the collection of all subsets of Ω which can be formed by taking unions of partition sets. $\mathcal{B}_\Lambda$ is closed under complements, as well as arbitrary (not just countable) unions and intersections. Suppose the partition is finite, i.e. Λ is finite, say $\Lambda = \{1, \ldots, n\}$. Then $\mathcal{B}_\Lambda$ is finite; it has exactly 2^n elements, each corresponding to a subset of Λ. Suppose now that Λ is countably infinite; since every subset $C \subseteq \Lambda$ determines a different set $B \in \mathcal{B}_\Lambda$, $\mathcal{B}_\Lambda$ is uncountable. Suppose finally that Λ is uncountable. For concreteness, let $\Omega = \mathbf{R}$, $\Lambda = \mathbf{R}$, and $\Omega_\lambda = \{\lambda\}$, i.e. each set in the partition consists of a single real number. There are many σ-algebras containing this partition. As before, we can take

$$\mathcal{B}_\Lambda = \{\cup_{\lambda \in C} \Omega_\lambda : C \subseteq \Lambda\} = \{C : C \subseteq \mathbf{R}\} = 2^{\mathbf{R}}$$

This is a perfectly good σ-algebra; however, as we indicated above (and as we shall show) it is not possible to define a measure with nice properties on this σ-algebra. The smallest σ-algebra containing the partition is

$$\mathcal{B}_0 = \{C : C \subseteq \mathbf{R},\ C \text{ countable or } \mathbf{R} \setminus C \text{ countable}\}$$

This σ-algebra is too small to allow us to develop probability theory. We want a σ-algebra on $\mathbf{R}$ which lies between $\mathcal{B}_0$ and $\mathcal{B}_\Lambda$.

Definition 6 The *Borel σ-algebra* on $\mathbf{R}$ is the smallest σ-algebra containing all open sets in $\mathbf{R}$. In other words, it is the σ-algebra

$$\mathcal{B} = \cap\{\mathcal{C} : \mathcal{C} \text{ is a } \sigma\text{-algebra, } U \text{ open } \Rightarrow U \in \mathcal{C}\}$$

i.e. it is the intersection of the class of all σ-algebras that contain all the open sets in $\mathbf{R}$. A set is called *Borel* if it belongs to $\mathcal{B}$.

Remark 7 Obviously, if U is an open set in $\mathbf{R}$, then $U \in \mathcal{B}$. If C is a closed set in $\mathbf{R}$, then $\mathbf{R} \setminus C$ is open, so $\mathbf{R} \setminus C \in \mathcal{B}$, so $C \in \mathcal{B}$; thus, every closed set is a Borel set. Every countable intersection of open sets, and every countable union of closed sets, is Borel. Every countable set is Borel (exercise), and every set whose complement is countable is Borel. If $\{U_{nm} : n, m \in \mathbf{N}\}$ is a collection of open sets, then $\cup_n (\cap_m U_{nm})$ is a Borel set. Thus, Borel sets can be quite complicated.

The most important example of a measure space is the Lebesgue measure space, which comes in two flavors. The first flavor is $(\mathbf{R}, \mathcal{B}, \mu)$, where $\mathcal{B}$ is the Borel σ-algebra and μ is Lebesgue measure, the measure defined in the following theorem; the second flavor is $(\mathbf{R}, \mathcal{C}, \mu)$, where $\mathcal{C}$ is the σ-algebra defined in the proof sketch (below) of the following theorem.

Theorem 8 *If $\mathcal{B}$ is the Borel σ-algebra on $\mathbf{R}$, there is a unique measure μ (called Lebesgue measure) defined on $\mathcal{B}$ such that $\mu((a,b)) = b - a$ provided that $b > a$. For every Borel set B,*

$$\begin{aligned} \mu(B) &= \sup\{\mu(K) : K \text{ compact, } K \subseteq B\} \\ &= \inf\{\mu(U) : U \text{ open, } U \supseteq B\} \end{aligned}$$

The proof works by gradually extending μ from the open intervals to the Borel σ-algebra. First, one shows one can extend μ to bounded open sets; it follows that one can extend it to compact sets. Then let $\mathcal{C}_n = \{C \subset [-n,n] : \sup\{\mu(K) : K \text{ compact}, K \subset C\} = \inf\{\mu(U) : U \text{ open}, C \subset U\}\}$, and $\mathcal{C} = \{C \subset \mathbf{R} : C \cap [-n,n] \in \mathcal{C}_n \text{ for all } n\}$. If $C \in \mathcal{C}$, we define $\mu(C) = \sup\{\mu(K) : K \text{ compact}, K \subset C\}$. One can verify that $\mathcal{C}$ is a σ-algebra containing every open set, hence $\mathcal{C} \supset \mathcal{B}$.

Definition 9 A measure space $(\Omega, \mathcal{B}, \mu)$ is *complete* if $B \in \mathcal{B}$, $A \subseteq B$, $\mu(B) = 0 \Rightarrow A \in \mathcal{B}$. The *completion* of a measure space $(\Omega, \mathcal{B}, \mu)$ is the measure space $(\Omega, \bar{\mathcal{B}}, \bar{\mu})$, where

$$\bar{\mathcal{B}} = \{B \subseteq \Omega : \exists_{C,D \in \mathcal{B}}\, C \subseteq B \subseteq D,\ \mu(D \setminus C) = 0\}$$

and

$$\bar{\mu}(B) = \sup\{\mu(C) : C \in \mathcal{B},\ C \subseteq B\} = \inf\{\mu(D) : D \in \mathcal{B},\ B \subseteq D\}$$

for $B \in \bar{\mathcal{B}}$.

It is easy to verify that the Lebesgue measure space $(\mathbf{R}, \mathcal{C}, \mu)$ is complete, and is the completion of $(\mathbf{R}, \mathcal{B}, \mu)$, where $\mathcal{B}$ is the Borel σ-algebra.

Definition 10 Suppose μ is a measure on the Lebesgue σ-algebra $\mathcal{C}$ on $\mathbf{R}$. μ is *translation-invariant* if, for all $x \in \mathbf{R}$ and all $C \in \mathcal{C}$, $\mu(C + x) = \mu(C)$.

Theorem 11 *The Lebesgue measure space $(\mathbf{R}, \mathcal{C}, \mu)$ is translation-invariant.*

The theorem follows readily from the construction of Lebesgue measure, since translation doesn't change the length of intervals.

Observe if $x \in \mathbf{R}$, $\mu(\{x\}) \leq \mu((x-\varepsilon/2, x+\varepsilon/2)) = \varepsilon$ for every positive ε, so $\mu(\{x\}) = 0$.

As we have already indicated, it is not possible to extend Lebesgue measure to every subset of $\mathbf{R}$, at least in the conventional set-theoretic foundations of mathematics.[1]

Theorem 12 *There is a set $D \subset \mathbf{R}$ which is* not *Lebesgue measurable.*

Proof: We actually prove the following stronger statement: there is no translation-invariant measure μ defined on all subsets of $\mathbf{R}$ such that $0 < \mu([0,1]) < \infty$.

Let μ be a translation-invariant measure defined on all subsets of $\mathbf{R}$. Define an equivalence relation $\sim$ on $\mathbf{R}$ by

$$x \sim y \Leftrightarrow x - y \in \mathbf{Q}$$

Let D be formed by choosing exactly one element from each equivalence class, and such that all the chosen elements lie in $[0,1)$. Thus,

$$\forall_{x \in \mathbf{R}}\ \exists_{d \in D}\ d - x \in \mathbf{Q}$$
$$\forall_{d_1, d_2 \in D}\ d_1 - d_2 \notin \mathbf{Q}$$

Given $x, y \in [0,1)$, define

$$x +' y = \begin{cases} x + y & \text{if } \ x + y \in [0,1) \\ x + y - 1 & \text{if } \ x + y \in [1,2) \end{cases}$$

The operation $+'$ is *addition modulo one.* It is easy to check that, given any $C \subseteq [0,1)$ and $y \in [0,1)$,

$$\mu(C +' y) = \mu(C)$$

i.e. μ is translation-invariant with respect to translation using the operation $+'$. Then

$$[0,1) = \cup_{q \in \mathbf{Q} \cap [0,1)} D +' q$$

[1]The crucial axiom needed in the proof of Theorem 12 is the Axiom of Choice; it is possible to construct alternative set theories in which the Axiom of Choice fails, and *every* subset of $\mathbf{R}$ is Lebesgue measurable. This is true, not because Lebesgue measure is extended further, but rather because the class of all sets is restricted.

so

$$\mu([0,1)) = \sum_{q\in\mathbf{Q}\cap[0,1)} \mu(D +' q) = \sum_{q\in\mathbf{Q}\cap[0,1)} \mu(D)$$

Then either $\mu(D) = 0$, in which case $\mu([0,1)) = 0$, so $\mu([0,1]) = 0$, or $\mu(D) > 0$, in which case $\mu([0,1)]) \geq \mu([0,1)) = \infty$. ■

Definition 13 A measure space $(\Omega, \mathcal{B}, \mu)$ is a *probability space* if $\mu(\Omega) = 1$.

From now on, we will restrict attention to probability spaces.

Example 14 By abuse of notation, let $\mathcal{C}$ denote the collection of Lebesgue measurable sets which are subsets of $[0,1]$, and let μ be the restriction of Lebesgue measure to this σ-algebra. $([0,1], \mathcal{C}, \mu)$ is called the Lebesgue probability space.

Theorem 15 *Let $(\Omega, \mathcal{B}, \mu)$ be a probability space. Suppose $B_1 \supseteq B_2 \supseteq \cdots \supseteq B_n \supseteq \cdots$ with $B_n \in \mathcal{B}$ for all n. Then*

$$\lim_{n\to\infty} \mu(B_n) = \mu\left(\cap_{n\in\mathbf{N}} B_n\right)$$

In particular, if $\cap_{n\in\mathbf{N}} B_n = \emptyset$, then

$$\lim_{n\to\infty} \mu(B_n) = 0$$

Proof: Let $C_m = B_m \setminus B_{m+1}$. Then $C_n \cap C_m = \emptyset$ if $n \neq m$, and

$$B_1 = \left(\cap_{n\in\mathbf{N}} B_n\right) \cup \left(\cup_{m\in\mathbf{N}} C_m\right)$$

so

$$\sum_{m\in\mathbf{N}} \mu(C_m) = \mu(B_1) - \mu\left(\cap_{n\in\mathbf{N}} B_n\right) \leq \mu(B_1) < \infty$$

so $\sum_{m\in\mathbf{N}} \mu(C_m)$ converges and hence

$$\sum_{m=n}^{\infty} \mu(C_m) \to 0 \text{ as } n \to \infty$$

$$\begin{aligned} \mu(B_n) &= \mu\left(\cap_{m\in\mathbf{N}} B_m\right) + \sum_{m=n}^{\infty} \mu(C_m) \\ &\to \mu\left(\cap_{m\in\mathbf{N}} B_m\right) \end{aligned}$$

■

We now turn to the definition of a *random variable*. We think of Ω as the set of all possible states of the world tomorrow. Exactly one state will occur tomorrow. Tomorrow, we will be able to observe some function of the state which occurred. We do not know today which state will occur, and hence what value of the function will occur. However, we do know the probability that any given state will occur, and we know the mapping from possible states into values of the function, so we know the probability that the function will take on any given possible value. Thus, viewed from today, the value the function will take on tomorrow is a random variable; we know its probability distribution, but not its exact value. Tomorrow, we will observe the value which is realized.

Definition 16 Let $(\Omega, \mathcal{B}, P)$ be a probability space. A *random variable* on Ω is a function $X : \Omega \to \mathbf{R} \cup \{-\infty, \infty\}$ satisfying

$$P(\{\omega \in \Omega : X(\omega) \in \{-\infty, \infty\}) = 0$$

which is measurable, i.e. for every Borel set $B \subset \mathbf{R}$, $X^{-1}(B) \in \mathcal{B}$.

Observe the analogy beteen the definition of a measurable function and the characterization of a continuous function in terms of open sets: f is continuous if and only if $f^{-1}(U)$ is open for every open set U.

Lemma 17 *Let $(\Omega, \mathcal{B}, P)$ be a probability space. A function $X : \Omega \to \mathbf{R} \cup \{-\infty, \infty\}$ satisfying*

$$P(\{\omega \in \Omega : X(\omega) \in \{-\infty, \infty\}\}) = 0$$

is a random variable if and only if for every open interval (a, b), $X^{-1}((a, b)) \in \mathcal{B}$.

Proof: Since every open interval is a Borel set, if X is a random variable, then $X^{-1}((a, b)) \in \mathcal{B}$.

Now suppose that $X^{-1}((a, b)) \in \mathcal{B}$ for every open interval (a, b). Consider

$$\mathcal{M} = \{B \subset \mathbf{R} : X^{-1}(B) \in \mathcal{B}\}$$

We claim that $\mathcal{M}$ is a σ-algebra. Note that $X^{-1}(\emptyset) = \emptyset \in \mathcal{B}$, so $\emptyset \in \mathcal{M}$. $X^{-1}(\mathbf{R}) = \Omega \setminus X^{-1}(\{-\infty, \infty\})$. Since $P(X^{-1}(\{-\infty, \infty\}))$ is zero (and hence defined), $X^{-1}(\{-\infty,\infty\}) \in \mathcal{B}$; since $\mathcal{B}$ is a σ-algebra, it is closed under complements, so $X^{-1}(\mathbf{R}) \in \mathcal{B}$, and $\mathbf{R} \in \mathcal{M}$.

If $B \in \mathcal{M}$, $X^{-1}(B) \in \mathcal{B}$, so $X^{-1}(\mathbf{R} \setminus B) = X^{-1}(\mathbf{R}) \setminus X^{-1}(B) \in \mathcal{B}$ because $X^{-1}(\mathbf{R}) \in \mathcal{B}$ and $X^{-1}(B) \in \mathcal{B}$. Therefore, $\mathbf{R} \setminus B \in \mathcal{M}$, so $\mathcal{M}$ is closed under complements.

Now suppose $B_n \in \mathcal{M}$, $n \in \mathbf{N}$. Then $X^{-1}(B_n) \in \mathcal{B}$ for all n, so

$$X^{-1}\left(\cup_{n\in\mathbf{N}} B_n\right) = \cup_{n\in\mathbf{N}} X^{-1}(B_n) \in \mathcal{B}$$

and hence $\mathcal{M}$ is closed under countable unions.

Therefore, $\mathcal{M}$ is a σ-algebra. Since every open set is a countable union of open intervals, $\mathcal{M}$ contains every open set; since the Borel σ-algebra is the smallest σ-algebra containing the open sets in $\mathbf{R}$, $\mathcal{M}$ contains every Borel set, so X is measurable, and hence a random variable. ■

Theorem 18 *If $f : [0,1] \to \mathbf{R}$ is continuous, then f is a random variable on the Lebesgue probability space $([0,1], \mathcal{C}, \mu)$.*

Proof: Since f takes values in $\mathbf{R}$, $f^{-1}(\{-\infty,\infty\}) = \emptyset$. If (a,b) is an open interval, then since f is continuous, $f^{-1}((a,b))$ is an open subset of $[0,1]$, so $f^{-1}((a,b)) \in \mathcal{C}$. By Lemma 17, f is a random variable. ■

Example 19 A random variable $f : [0,1] \to \mathbf{R}$ can be discontinuous everywhere. Consider $f(x) = 1$ if x is rational, 0 if x is irrational. f is discontinuous everywhere. Suppose B is a Borel set. There are four cases to consider:

1. If $0,1 \notin B$ $f^{-1}(B) = \emptyset \in \mathcal{C}$.
2. If $0 \in B, 1 \notin B$, $f^{-1}(B) = [0,1] \setminus \mathbf{Q} \in \mathcal{C}$.
3. If $0 \notin B, 1 \in B$, $f^{-1}(B) = [0,1] \cap \mathbf{Q} \in \mathcal{C}$.
4. If $0,1 \in B$, $f^{-1}(B) = [0,1] \in \mathcal{C}$.

Thus, f is measurable.

In elementary probability theory, random variables are not rigorously defined. Usually, they are described only by specifying their cumulative distribution functions. Continuous and discrete distributions often seem like entirely unconnected notions, and the formulation of mixed distributions (which have both continuous and discrete parts) can be problematic. Measure theory gives us a way to deal simultaneously with continuous, discrete, and mixed distibutions in a unified way. We shall first define the cumulative distribution function of a random variable, and establish that it satisfies the defining properties of a cumulative distribution functions, as defined in elementary probability theory. We will then show (first in examples, then in a general theorem) that given any function F satisfying the defining properties of a cumulative distribution function, there is in fact a random variable defined on the Lebesgue probability space whose cumulative distribution function is F.

Definition 20 Given a random variable $X : (\Omega, \mathcal{B}, P) \to \mathbf{R} \cup \{-\infty, \infty\}$, the *cumulative distribution function* of X is the function $F : \mathbf{R} \to [0, 1]$ defined by

$$F(t) = P(\{\omega \in \Omega : X(\omega) \le t\})$$

Theorem 21 *If X is a random variable, its cumulative distribution function F satisfies the following properties:*

1. $$\lim_{t \to -\infty} F(t) = 0$$
   this is often abbreviated as $F(-\infty) = 0$.

2. $$\lim_{t \to \infty} F(t) = 1$$
   this is often abbreviated as $F(\infty) = 1$.

3. *F is increasing, i.e.*
   $$s < t \Rightarrow F(t) \ge F(s)$$

4. *F is right-continuous, i.e. for all t,*
   $$\lim_{s \searrow t} F(s) = F(t)$$

Proof: We prove only right-continuity, leaving the rest as an exercise. Since F is increasing, it is enough to show that (as an exercise, think through carefully why this *is* enough)

$$\lim_{n\to\infty} F\left(t+\frac{1}{n}\right) = F(t)$$

$$\begin{aligned}
& F\left(t+\frac{1}{n}\right) - F(t) \\
&= P\left(\left\{\omega\in\Omega : X(\omega)\le t+\frac{1}{n}\right\}\right) - P\left(\{\omega\in\Omega : X(\omega)\le t\}\right) \\
&= P\left(X^{-1}\left(\left(-\infty, t+\frac{1}{n}\right]\right)\right) - P\left(X^{-1}\left((-\infty,t]\right)\right) \\
&= P\left(X^{-1}\left(\left(t, t+\frac{1}{n}\right]\right)\right) \\
&\to P\left(\bigcap_{n\in\mathbf{N}} X^{-1}\left(\left(t,t+\frac{1}{n}\right]\right)\right) \\
&= P\left(X^{-1}\left(\bigcap_{n\in\mathbf{N}}\left(t,t+\frac{1}{n}\right]\right)\right) \\
&= P\left(X^{-1}(\emptyset)\right) \\
&= P(\emptyset) = 0
\end{aligned}$$

by Theorem 15. ■

Example 22 The *uniform distribution* on $[0,1]$ is the cumulative distribution function

$$F(t) = \begin{cases} 0 & \text{if} \quad t<0 \\ t & \text{if} \quad t\in[0,1] \\ 1 & \text{if} \quad t>1 \end{cases}$$

Consider the random variable X defined on the Lebesgue probability space $X:([0,1],\mathcal{C},\mu)\to\mathbf{R}$, $X(t)=t$. Observe that

$$\mu\left(\{\omega\in[0,1] : X(\omega)\le t\}\right) = \mu\left([0,t]\right) = t$$

Thus, X has the uniform distribution on $[0,1]$. Notice also that F is strictly increasing and hence one-to-one on $[0,1]$. Thus, $F|_{[0,1]}$ has an inverse function $\left(F|_{[0,1]}\right)^{-1} : [0,1]\to[0,1]$. In fact, $X = \left(F|_{[0,1]}\right)^{-1}$.

Example 23 The *standard normal distribution* has the cumulative distribution function

$$F(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} dx$$

Notice that F is strictly increasing, hence one-to-one, and the range of F is $(0,1)$, so F has an inverse function $X : (0,1) \to \mathbf{R}$; extend X to $[0,1]$ by defining $X(0) = -\infty$ and $X(1) = \infty$. One can show that the inverse of a strictly increasing, continuous function is strictly increasing and continuous. Since $X|_{(0,1)}$ is continuous, it is measurable on the Lebesgue probability space. Observe that

$$\mu\left(\{\omega \in [0,1] : X(\omega) \le t\}\right) = \mu\left(X^{-1}((-\infty, t])\right) = F(t)$$

so X has the standard normal distribution.

Example 24 Consider the cumulative distribution function

$$F(t) = \begin{cases} 0 & \text{if} \quad t < -1 \\ \frac{1}{2} & \text{if} \quad t \in [-1,1) \\ 1 & \text{if} \quad t \ge 1 \end{cases}$$

F is the cumulative distribution function of a random variable which takes on the values -1 and 1, each with probability $\frac{1}{2}$. We define a random variable on the Lebesgue probability space by

$$X(\omega) = \begin{cases} -\infty & \text{if} \quad \omega = 0 \\ -1 & \text{if} \quad \omega \in \left(0, \frac{1}{2}\right] \\ 1 & \text{if} \quad \omega \in \left(\frac{1}{2}, 1\right] \end{cases}$$

Clearly, the cumulative distribution function of X is F. Notice that F is weakly but not strictly increasing, so it is not one-to-one and hence does not have an inverse function. However, notice that X satisfies the following weakening of the definition of an inverse:

$$X(\omega) = \inf\{t : F(t) \ge \omega\}$$

Theorem 25 *Let $F : \mathbf{R} \to [0,1]$ be an arbitrary function satisfying conclusions 1-4 of Theorem 21. There is a random variable X defined on the Lebesgue probability space whose cumulative distribution function is F.*

Proof: Let

$$X(\omega) = \inf\{t : F(t) \geq \omega\}$$

Since

$$\begin{aligned}
&\omega' > \omega \\
&\quad \Rightarrow \quad \{t : F(t) \geq \omega'\} \subseteq \{t : F(t) \geq \omega\} \\
&\quad \Rightarrow \quad \inf\{t : F(t) \geq \omega'\} \geq \inf\{t : F(t) \geq \omega\} \\
&\quad \Rightarrow \quad X(\omega') \geq X(\omega)
\end{aligned}$$

X is increasing (not necessarily strictly). Since $\lim_{s\to-\infty} F(s) = 0$ and $\lim_{t\to\infty} F(t) = 1$, if $\omega \in (0,1)$, then there exist $s,t \in \mathbf{R}$ such that $F(s) < \omega < F(t)$; since F is increasing, $-\infty < X(\omega) < \infty$. Thus,

$$\mu\left(\{\omega : X(\omega) \in \{-\infty, \infty\}\}\right) \leq \mu\{0,1\}) = 0$$

Since X is increasing, if (a,b) is an open interval in $\mathbf{R}$, $X^{-1}((a,b))$ is an interval (not necessarily open) in $[0,1]$, and hence $X^{-1}((a,b)) \in \mathcal{C}$. By Lemma 17, X is a random variable.

$$\begin{aligned}
X(\omega) \leq t \quad &\Leftrightarrow \quad \inf\{s : F(s) \geq \omega\} \leq t \\
&\Leftrightarrow \quad \exists_{s\leq t}\ F(s) \geq \omega \\
&\Leftrightarrow \quad F(t) \geq \omega
\end{aligned}$$

so

$$\begin{aligned}
\mu\left(\{\omega : X(\omega) \leq t\}\right) &= \mu(\{\omega : \omega \leq F(t)\}) \\
&= \mu([0, F(t)]) \\
&= F(t)
\end{aligned}$$

■

In probability theory, the choice of a particular probability space is usually considered arbitrary and unimportant; you should choose a probability space which is convenient for the particular problem at hand, in particular one on which you can easily write down random variables with the desired joint distribution. **In what follows, we will allow $(\Omega, \mathcal{B}, P)$ to be an arbitrary probability space, not necessarily the Lebesgue probability space.**

Next, we develop a notion of integration.

Definition 26 A *simple function* is a function of the form

$$f = \sum_{i=1}^{n} \alpha_i \chi_{B_i} \tag{1}$$

where each B_i is a measurable set in Ω and χ_{B_i} is the characteristic function of B_i:

$$\chi_{B_i}(\omega) = \begin{cases} 1 & \text{if} \quad \omega \in B_i \\ 0 & \text{if} \quad \omega \notin B_i \end{cases}$$

We define the integral of f by

$$\int_\Omega f dP = \sum_{i=1}^{n} \alpha_i P(B_i)$$

A given simple function may be expressed in the form of Equation 1 in more than one way; one can show that one gets the same value of the integral from all of these different expressions.

Given a nonnegative random variable $f : (\Omega, \mathcal{B}, P) \to \mathbf{R}_+ \cup \{\infty\}$, we can construct a simple function which closely approximates it from below. Choose constants

$$0 = \alpha_1 < \alpha_2 < \cdots < \alpha_n < \alpha_{n+1} = \infty$$

Let

$$f_n = \sum_{i=1}^{n} \alpha_i \chi_{f^{-1}([\alpha_i, \alpha_{i+1}))}$$

Thus,

$$f_n(\omega) = \alpha_i \Leftrightarrow f(\omega) \in [\alpha_i, \alpha_{i+1})$$

Because f is a random variable, $f^{-1}([\alpha_i, \alpha_{i+1})) \in \mathcal{B}$, so f_n is a simple function.

Definition 27 If $f : (\Omega, \mathcal{B}, P) \to \mathbf{R}_+ \cup \{\infty\}$ is a random variable, define

$$\int_\Omega f dP = \sup \left\{ \int_\Omega g dP : 0 \leq g \leq f, \, g \text{ simple} \right\}$$

The value of the integral may be ∞. We say that f is *integrable* if $\int_\Omega f dP < \infty$.

Example 28 Consider the function $f(x) = \frac{1}{x}$ on the Lebesgue probability space $([0,1], \mathcal{C}, \mu)$. Let

$$f_n = \sum_{i=1}^{n} i\chi_{f^{-1}([i,i+1))} = \sum_{i=1}^{n} i\chi_{\left(\left(\frac{1}{i+1}, \frac{1}{i}\right]\right)}$$

f_n is a simple function and $0 \le f_n \le f$.

$$\begin{aligned}
\int f_n d\mu &= \sum_{i=1}^{n} i\mu\left(\left(\frac{1}{i+1}, \frac{1}{i}\right]\right) \\
&= \sum_{i=1}^{n} i\left(\frac{1}{i} - \frac{1}{i+1}\right) \\
&= \sum_{i=1}^{n} i\left(\frac{i+1-i}{i(i+1)}\right) \\
&= \sum_{i=1}^{n} \frac{1}{i+1} \\
&\to \infty \text{ as } n \to \infty
\end{aligned}$$

Thus, $\int_{[0,1]} f\, d\mu = \infty$.

Definition 29 If $f : (\Omega, \mathcal{B}, P) \to \mathbf{R} \cup \{-\infty, \infty\}$ is a random variable, define $f_+(\omega) = \max\{f(\omega), 0\}$, $f_-(\omega) = -\min\{f(\omega), 0\}$. Note that $f = f_+ - f_-$ and $|f| = f_+ + f_-$. We say f is *integrable* if f_+ and f_- are both integrable, and we define

$$\int_\Omega f dP = \int_\Omega f_+ dP - \int_\Omega f_- dP$$

The following theorem, which we shall not prove, shows that the Lebesgue and Riemann integrals coincide when the Riemann integral is defined:

Theorem 30 *Suppose that $f : [a,b] \to \mathbf{R}$ is Riemann integrable (in particular, this is the case if f is continuous). Then f is Lebesgue integrable and*

$$\int_{[a,b]} f\, d\mu = \int_a^b f(t)dt$$

In other words, the Lebesgue integral of f equals the Riemann integral of f.

Definition 31 We say that a property holds *almost everywhere* (abbreviated a.e.) or *almost surely* (abbreviated a.s.) if the set of ω for which it does *not* hold is a set of measure zero.

Definition 32 Suppose $f : (\Omega, \mathcal{B}, P) \to \mathbf{R}$ is a random variable. If $B \in \mathcal{B}$, we define

$$\int_B f dP = \int_\Omega f \chi_B dP$$

where χ_B is the characteristic function of B.

Theorem 33 *Suppose $f, g : (\Omega, \mathcal{B}, P) \to \mathbf{R}$ are integrable, $A, B \in \mathcal{B}$, $A \cap B = \emptyset$, and $c \in \mathbf{R}$. Then*

1. $\int_B cf dP = c \int_B f dP$
2. $\int_B f + g dP = \int_B f dP + \int_B g dP$
3. $f \leq g$ *a.e. on* $B \Rightarrow \int_B f dP \leq \int_B g dP$
4. $\int_{A \cup B} f dP = \int_A f dP + \int_B f dP$

Definition 34 Suppose $f, g : (\Omega, \mathcal{B}, P) \to \mathbf{R}$ are random variables. Say $f \sim g$ if $f(w) = g(\omega)$ a.e. Let $[f]$ denote the equivalence class of f under the equivalence relation $\sim$, i.e. $[f] = \{g : (\Omega, \mathcal{B}, P) \to \mathbf{R} : g \text{ is measurable}, g \sim f\}$. Define

$$\begin{aligned} L^1(\Omega, \mathcal{B}, P) &= \{[f] : f : \Omega \to \mathbf{R}, f \text{ a random variable, } f \text{ integrable}\} \\ L^2(\Omega, \mathcal{B}, P) &= \left\{[f] : f : \Omega \to \mathbf{R}, f \text{ a random variable, } f^2 \text{ integrable}\right\} \end{aligned}$$

Theorem 35 *If $(\Omega, \mathcal{B}, P)$ is a probability space,*

$$L^2(\Omega, \mathcal{B}, P) \subseteq L^1(\Omega, \mathcal{B}, P)$$

Definition 36 Given a metric space (X, d), the *completion* of (X, d) is a metric space $(\bar{X}, \bar{d})$ such that $(\bar{X}, \bar{d})$ is complete, $\bar{d}|_X = d$, and X is dense in $(\bar{X}, \bar{d})$. We are justified in calling $(\bar{X}, \bar{d})$ *the* completion of (X, d) because if $(\hat{X}, \hat{d})$ is another completion of (X, d), then there is an isomorphism $\phi : (\bar{X}, \bar{d}) \to (\hat{X}, \hat{d})$ such that $\phi(x) = x$ for every $x \in X$, and $\hat{d}(\phi(\bar{x}), \phi(\bar{y})) = \bar{d}(\bar{x}, \bar{y})$ for every $\bar{x}, \bar{y} \in \bar{X}$.

Theorem 37 *$L^1(\Omega, \mathcal{B}, P)$ and $L^2(\Omega, \mathcal{B}, P)$ are Banach spaces under the respective norms*

$$\begin{aligned} \|f\|_1 &= \int_\Omega |f(\omega)| dP \\ \|f\|_2 &= \sqrt{\int_\Omega f^2(\omega) dP} \end{aligned}$$

For the Lebesgue probability space, $L^1([0,1], \mathcal{C}, \mu)$ and $L^2([0,1], \mathcal{C}, \mu)$ are the completions of the normed space $C([0,1])$, with respect to the norms $\|\cdot\|_1$ and $\|\cdot\|_2$.

Example 38 Recall that $C([0,1])$ is a Banach space with the sup norm $\|f\| = \sup\{|f(x)| : x \in [0,1]\}$. $(C[0,1], \|\cdot\|_1)$ and $(C[0,1], \|\cdot\|_2)$ are normed spaces, but they are not complete in these norms, hence are not Banach spaces. To see this, let $f_n(x)$ be the function which is zero for $x \in \left[0, \frac{1}{2} - \frac{1}{n}\right]$, one for $x \in \left[\frac{1}{2} + \frac{1}{n}, 1\right]$ and linear for $x \in \left[\frac{1}{2} - \frac{1}{n}, \frac{1}{2} + \frac{1}{n}\right]$. f_n is not Cauchy with respect to $\|\cdot\|_\infty$. However, f_n *is* Cauchy and has a natural limit with respect to $\|\cdot\|_1$. Indeed, let

$$f(x) = \begin{cases} 0 & \text{if} \quad x < \frac{1}{2} \\ 1 & \text{if} \quad x \geq \frac{1}{2} \end{cases}$$

Then

$$\int_{[0,1]} |f_n - f| \, d\mu \leq \frac{1}{2} \times \frac{2}{n} = \frac{1}{n} \to 0$$

so f_n converges to f in $L^1([0,1], \mathcal{C}, \mu)$, and hence must be Cauchy with respect to $\|\cdot\|_1$. Notice that this limit does not belong to $C([0,1])$.

Definition 39 For $f \in L^1(\Omega, \mathcal{B}, P)$, define the *mean* or *expectation* of f by

$$E(f) = \int_\Omega f dP$$

For $f \in L^2(\Omega, \mathcal{B}, P)$, define the *variance* of f by

$$\mathrm{Var}(f) = \int_\Omega (f(\omega) - E(f))^2 dP$$

Definition 40 If $f, g \in L^2(\Omega, \mathcal{B}, P)$, define the *inner product* of f and g by

$$f \cdot g = \int_\Omega f(\omega)g(\omega)dP$$

The properties of the inner product are closely analogous to those of the dot product of vectors in Euclidean space $\mathbf{R}^n$. In particular, they determine a geometry, including lengths ($\|f\|_2 = \sqrt{f \cdot f}$) and angles. The most basic property of the dot product, the Cauchy-Schwarz inequality, extends to the inner product.

Theorem 41 (Cauchy-Schwarz Inequality) *If $f, g \in L^2(\Omega, \mathcal{B}, P)$,*

$$|f \cdot g| \leq \|f\|_2 \|g\|_2$$

Definition 42 If $f, g \in L^2(\Omega, \mathcal{B}, P)$, define the *covariance* of f and g by

$$\mathrm{Covar}(f, g) = E((f - E(f))(g - E(g)))$$

Proposition 43 *If $f, g \in L^2(\Omega, \mathcal{B}, P)$, then*

$$\mathrm{Covar}(f, g) = E(fg) - E(f)E(g)$$

Observe from the definition that $\mathrm{Covar}(f, g) = \mathrm{Covar}(g, f)$. Thus, given $f_1, \ldots, f_n \in L^2(\Omega, \mathcal{B}, P)$, the covariance matrix C whose (i, j) entry is $c_{ij} = \mathrm{Covar}(f_i, f_j)$ is a symmetric matrix. Hence, there is an orthonormal basis of $\mathbf{R}^n$ composed of eigenvectors of C; expressed in this basis, the matrix becomes diagonal. Observe also that

$$Covar(\sum_{i=1}^n \alpha_i f_i, \sum_{j=1}^n \beta_j f_j) = (\alpha_1, \ldots, \alpha_n)\, C \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}$$

is a quadratic form of the type we studied in the Supplement to Section 3.6.

One very useful consequence of the inner product structure on L^2 is the existence of orthogonal projections. Given a linearly independent family $f_1, \ldots, f_n \in L^2(\Omega, \mathcal{B}, P)$, let V be the vector space spanned by $f_1, \ldots, f_n$. Then any $g \in L^2$ can be written in a unique way as $g = \pi(g) + w$, where $\pi(v) = \sum_{i=1}^n \alpha_i f_i \in V$ and $w \cdot v = 0$ for all $v \in V$ (in particular, $w \cdot f_i = 0$

for each i). $\pi(g)$ is also characterized as the point in V closest to g. The coefficients $\alpha_1, \cdots, \alpha_n$ are the regression coefficients of g on $f_1, \ldots, f_n$. If $f_1, \ldots, f_n$ are orthonormal (i.e. $f_i \cdot f_j = 1$ if $i = j$ and 0 if $i \neq j$), $\alpha_i = g \cdot f_i$ for each i.

One frequently encounters two measures living on a single probability space. For example, Lebesgue measure is one measure on the real line $\mathbf{R}$; any random variable determines another measure on $\mathbf{R}$ by its distribution.

Definition 44 Let μ and ν be two measures on the same measurable space $(X, \mathcal{A})$, i.e. $(X, \mathcal{A}, \mu)$ and $(X, \mathcal{A}, \nu)$ are measure spaces. We say ν is *absolutely continuous with respect to* μ if

$$B \in \mathcal{A},\ \mu(B) = 0 \Rightarrow \nu(B) = 0$$

We say μ is σ-finite if there exist $B_1, B_2, \ldots \in \mathcal{A}$ such that $X = \cup_{n \in \mathbf{N}} B_n$ and $\mu(B_n) < \infty$ for all n.

Theorem 45 (Radon-Nikodym) *Suppose that μ and ν are two σ-finite measures on the same measurable space $(X, \mathcal{A})$. Then ν is absolutely continuous with respect to μ if and only if there exists $f \in L^1(X, \mathcal{A}, \mu)$, $f \geq 0$ almost everywhere, such that*

$$\nu(B) = \int_B f \, d\mu$$

for all $B \in \mathcal{A}$.

The Radon-Nikodym theorem tells us that if a random variable Z has a continuous distribution (in other words, the measure ν determined by its cumulative distribution function F is absolutely continuous with respect to Lebesgues measure), then Z has a density function f, namely f is the Radon-Nikodym derivative of ν with respect to Lebesgue measure.

We now turn to products of probability spaces $(X, \mathcal{A}, P)$ and $(Y, \mathcal{B}, Q)$. If $A \in \mathcal{A}$ and $B \in \mathcal{B}$, it is natural to define

$$(P \times Q)(A \times B) = P(A)Q(B)$$

However, most of the subsets of $X \times Y$ we are interested in are not of the form $A \times B$ with $A \in \mathcal{A}$ and $B \in \mathcal{B}$. For example, if $X = Y = [0, 1]$, the diagonal $\{(x,y) : x = y\}$ is *not* of that form. However, it is possible to write the diagonal in the form

$$\cap_{m \in \mathbf{N}} \cup_{n=1}^{i_m} A_{mn} \times B_{mn}$$

with $A_{mn} \in \mathcal{A}$ and $B_{mn} \in \mathcal{B}$. This suggests we should extend $P \times Q$ to the smallest σ-algebra containing $\{A \times B : A \in \mathcal{A}, B \in \mathcal{B}\}$.

Definition 46 Suppose $(X, \mathcal{A}, P)$ and $(Y, \mathcal{B}, Q)$ are probability spaces. The product probability space is $(X \times Y, \mathcal{A} \times \mathcal{B}, P \times Q)$, where

1. $\mathcal{A} \times' \mathcal{B}$ is the smallest σ-algebra containing all sets of the form $A \times B$, where $A \in \mathcal{A}$ and $B \in \mathcal{B}$
2. $P \times' Q$ is the unique measure on $\mathcal{A} \times' \mathcal{B}$ that takes the values $P(A)Q(B)$ on sets of the form $A \times B$, where $A \in \mathcal{A}$ and $B \in \mathcal{B}$.
3. $(X \times Y, \mathcal{A} \times \mathcal{B}, P \times Q)$ is the completion of $(X \times Y, \mathcal{A} \times' \mathcal{B}, P \times' Q)$

The existence of the product measure $P \times Q$ is proven by extending the definition of $P \times Q$ for sets of the form $A \times B$ to $\mathcal{A} \times \mathcal{B}$, in a manner similar to the process by which Lebesgue measure is obtained from lengths of intervals.

Theorem 47 (Fubini) *Suppose $(X, \mathcal{A}, P)$ and $(Y, \mathcal{B}, Q)$ are complete probability spaces. If $f : X \times Y \to \mathbf{R}$ is $\mathcal{A} \times \mathcal{B}$ measurable and integrable, then*

1. *for almost all $x \in X$, the function defined by $f_x(y) = f(x,y)$ is integrable in Y;*
2. *for almost all $y \in Y$, the function defined by $f_y(x) = f(x,y)$ is integrable in X;*
3. *$\int_Y f(x,y)dQ(y)$ is an integrable function of $x \in X$;*
4. *$\int_X f(x,y)dP(x)$ is an integrable function of $y \in Y$;*
5.
$$\begin{aligned} \int_X \left(\int_Y f(x,y)dQ(y) \right) dP(x) &= \int_Y \left(\int_X f(x,y)dP(x) \right) dQ(y) \\ &= \int_{X \times Y} f(x,y)d(P \times Q)(x,y) \end{aligned}$$

There are many notions of convergence of functions, and results showing that the integral of a limit of functions is the limit of the integrals.

Definition 48 Suppose $f_n, f : (\Omega, \mathcal{B}, P) \to \mathbf{R}$ are random variables.

1. We say f_n converges to f *in probability* or *in measure* if

$$\forall_{\varepsilon>0}\ \exists_N\ n > N \Rightarrow P\left(\{\omega : |f_n(\omega) - f(\omega)| > \varepsilon\}\right) < \varepsilon$$

2. We say f_n converges to f *almost everywhere* (abbreviated *a.e.*) or *almost surely* (abbreviated *a.s.*) if

$$P(\{\omega : f_n(\omega) \to f(\omega)\}) = 1$$

Theorem 49 *If f_n converges to f almost surely, then f_n converges to f in probability.*

In the lecture, I gave an example of a sequence of functions which converges in probability, but not almost surely. The example is much easier to describe with pictures than with words, so I won't give the details here. The idea is to construct a sequence of functions $f_n : [0, 1] \to \mathbf{R}$ so that $\mu(\{x : f_n(x) \neq 0\}) \to 0$; hence f_n converges in probability to the identically zero function f. However, we can move the set on which f_n is not zero around so that, for every $x \in [0, 1]$, $f_n(x) = 1$ for infinitely many n, so almost surely, $f_n(x)$ *fails* to converge to $f(x)$.

Example 50 Although convergence almost surely is a strong notion of convergence of functions, it is not sufficient for convergence of the integrals of the functions to the integral of the limit. For example, consider $f_n : [0, 1] \to \mathbf{R}$ defined by $f_n = n\chi_{[0,1/n]}$, which takes the value n on $[0, 1/n]$ and 0 everywhere else. f_n converges almost surely to f, the function which is identically zero, but if P is Lebesgue measure, $\int_{[0\,1]} f_n dP = n \times \frac{1}{n} = 1$, while $\int_{[0,1]} f dP = 0$. Roughly speaking, f_n is converging to a function which is ∞ with probabilty 0, and in the Lebesgue integral, $0 \times \infty = 0$; some of the mass in f_n is lost in the limit operation. The next result tells us that, for nonnegative functions, this is the only thing that can go wrong, and the limit of the integrals is always at least as big as the integral of the limit function.

Theorem 51 (Fatou's Lemma) *If* $(\Omega, \mathcal{B}, P)$ *is a probability space,* $f_n, f : \Omega \to \mathbf{R}_+$ *are random variables, and* f_n *converges to* f *almost surely, then*

$$\int_\Omega f\, dP \leq \liminf_{n\to\infty} \int_\Omega f_n\, dP$$

In the next theorem, the functions f_n converge to f from below; this guarantees that no mass which is present in the f_n can suddenly disappear at the limit.

Theorem 52 (Monotone Convergence Theorem) *If* $(\Omega, \mathcal{B}, P)$ *is a probability space,* $f_n : \Omega \to \mathbf{R}_+$ *are random variables,* f_n *converges to* f *almost surely, and for all* n *and* ω, $f_n(\omega) \leq f(\omega)$. *Then*

$$\int_\Omega f\, dP = \lim_{n\to\infty} \int_\Omega f_n\, dP$$

Theorem 53 (Lebesgue's Dominated Convergence Theorem) *Suppose* $(\Omega, \mathcal{B}, P)$ *is a probability space,* $g : \Omega \to \mathbf{R}$ *is integrable,* f_n *random variables,* $|f_n| \leq g$ *a.s., and* $f_n \to f$ *a.s. Then* f *is integrable and*

$$\int_\Omega f_n dP \to \int_\Omega f dP$$

Definition 54 Suppose $(\Omega, \mathcal{B}, P)$ is a probability space. We say $\{f_n\}$ is *uniformly integrable* if

$$\forall_{\varepsilon>0}\ \exists_{\delta>0}\ \forall_{n\in\mathbf{N}}\ P(B) < \delta \Rightarrow \left|\int_B f_n\, dP\right| < \varepsilon$$

The following theorem is a useful generalization the Dominated Convergence Theorem.

Theorem 55 *Suppose* $(\Omega, \mathcal{B}, P)$ *is a probability space. If* $\{f_n\}$ *is uniformly integrable and* $f_n \to f$ *a.s., then* f *is integrable and*

$$\int_\Omega f_n\, dP \to \int_\Omega f\, dP$$