Sample Space and Collections of Sets

A sample space is a set. It is the "universe of discourse" in a given problem. It is often denoted by Ω.

We usually denote collections of sets with upper-case caligraphic letters, e.g., $\mathcal{A}$, $\mathcal{F}$, etc.

The usual set operators and set relations are used with collections of sets, and generally have the same meaning.

An important collection of sets is the collection of all open intervals of $\mathbb{R}$.

σ-Fields

A collection of subsets, $\mathcal{F}$, of a given sample space, Ω, such that

1. $\emptyset \in \mathcal{F}$

2. $A \in \mathcal{F} \Rightarrow A^{\mathsf{c}} \in \mathcal{F}$

3. $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \cup_i A_i \in \mathcal{F}$.

Notice that a σ-field is defined with respect to some given sample space.

A σ-field is also called a σ-algebra. The definition of a σ-field sometimes includes the condition that $\Omega \in \mathcal{F}$ rather than $\emptyset \in \mathcal{F}$.

Given any sample space Ω and any collection $\mathcal{C}$ of subsets of Ω, the "smallest" σ-field of which $\mathcal{C}$ is a subset is called the σ-field generated by $\mathcal{C}$, and is denoted by $\sigma(\mathcal{C})$.

Example: If $\mathcal{A} = \{A\}$, with respect to the sample space Ω, $\sigma(\mathcal{A}) = \{\emptyset, A, A^{\mathsf{c}}, \Omega\}$.

σ-fields grow very rapidly.

If k is the maximum number of sets that partition Ω that can be formed by operations on the sets in $\mathcal{A}$, then the number of sets in the σ-field is 2^k.

Measurable Space: The Structure $(\Omega, \mathcal{F})$

If Ω is a sample space, and $\mathcal{F}$ is a σ-field over Ω, the double $(\Omega, \mathcal{F})$ is called a *measurable space*.

Notice that no "measure" is required.

We have not yet defined *measure*.

Functions and Images

A function is a set of ordered pairs such that no two pairs have the same first element.

We use "function" and "mapping" synonymously, although the latter term is sometimes interpreted more generally.

To say that f is a mapping from Ω to Λ, written

$$f : \Omega \mapsto \Lambda,$$

means that for every $\omega \in \Omega$ there is a pair in f whose first member is ω.

We use the notation $f(\omega)$ to represent the second member of the pair in f whose first member is ω.

Types include functions that are *onto*, meaning that for every $\lambda \in \Lambda$ there is a pair in f whose second member is λ; and functions that are *one-to-one*, written $1:1$, meaning that no two pairs have the same second member.

A function that is one-to-one and onto is called a *bijection*.

A function f that is one-to-one has an inverse, written f^{-1}.

If $(a, b) \in f$, we may write $a = f^{-1}(b)$, although sometimes this notation is restricted to the cases in which f is one-to-one.

(There are some subtleties here; if f is not one-to-one, if the members of the pairs in f are reversed, the resulting set is not a function. We say f^{-1} does not exist; yet we may write $a = f^{-1}(b)$, with the meaning above.)

If $A \subset \Omega$, the *image* of A, denoted by $f[A]$, or just by $f(A)$, is the set of all $\lambda \in \Lambda$ for which $\lambda = f(\omega)$ for some $\omega \in \Omega$.

(The notation $f[A]$ is preferable, but we will often just use $f(A)$ in this course.) Similarly, if $\mathcal{C}$ is a collection of sets, the notation $f[\mathcal{C}]$ denotes the collection of sets $\{f[C] : C \in \mathcal{C}\}$.

For the function f that maps from Ω to Λ, Ω is called the *domain* of the function and and $f[\Omega]$ is called the *range* of the function.

For a subset B of Λ, the *inverse image* or the *preimage* of B, denoted by $f^{-1}[B]$, or just by $f^{-1}(B)$, is the set of all $\omega \in \Omega$ such that $f(\omega) \in B$.

The notation f^{-1} used in this sense must not be confused with the inverse function f^{-1} (if the latter exists).

We use this notation for the inverse image whether or not the inverse of the function exists.

Notice that the inverse image of a set may not generate the set; that is, $f[f^{-1}[B]] \subset B$.

We also write $f[f^{-1}[B]]$ as $f \circ f^{-1}[B]$. The set $f[f^{-1}[B]]$ may be a proper subset of B; that is, there may be an element λ in B for which there is no $\omega \in \Omega$ such that $f(\omega) = \lambda$. If f is bijective, then $f[f^{-1}[B]] = B$.

The following are useful facts about a function f that maps Ω to Λ.

(i) $f(\emptyset) = \emptyset$ and $f^{-1}(\emptyset) = \emptyset$

This is the only way that makes sense!

(ii) For $B \subset \Lambda$, $f^{-1}(B^{\mathsf{c}}) = (f^{-1}(B))^{\mathsf{c}}$

(where $B^{\mathsf{c}} = \Lambda - B$, and $(f^{-1}(B))^{\mathsf{c}} = \Omega - f^{-1}(B)$).

We see this in the standard way by showing that each is a subset of the other.

Let ω be an arbitrary element of Ω.

Suppose $\omega \in f^{-1}(B^{\mathsf{c}})$. Then $f(\omega) \in B^{\mathsf{c}}$, so $f(\omega) \notin B$, hence $\omega \notin f^{-1}(B)$, and so $\omega \in (f^{-1}(B))^{\mathsf{c}}$. We have $f^{-1}(B^{\mathsf{c}}) \subset (f^{-1}(B))^{\mathsf{c}}$.

Now suppose $\omega \in (f^{-1}(B))^{\mathsf{c}}$. Then $\omega \notin f^{-1}(B)$, so $f(\omega) \notin B$, hence $f(\omega) \in B^{\mathsf{c}}$, and so $\omega \in f^{-1}(B^{\mathsf{c}})$. We have $(f^{-1}(B))^{\mathsf{c}} \subset f^{-1}(B^{\mathsf{c}})$.

(iii) Let $A_1, A_2, \ldots \subset \Lambda$ and suppose $(\cup_{i=1}^{\infty} A_i) \subset \Lambda$, then $f^{-1}(\cup_{i=1}^{\infty} A_i) = \cup_{i=1}^{\infty} f^{-1}(A_i)$.

Again, let λ be an arbitrary element of Λ.

Suppose $\lambda \in f^{-1}(\cup_{i=1}^{\infty} A_i)$. Then $f(\lambda) \in \cup_{i=1}^{\infty} A_i$, so for some j, $f(\lambda) \in A_j$ and $\lambda \in f^{-1}(A_j)$; hence $\lambda \in \cup_{i=1}^{\infty} f^{-1}(A_i)$. We have $f^{-1}(\cup_{i=1}^{\infty} A_i) \subset \cup_{i=1}^{\infty} f^{-1}(A_i)$.

Now suppose $\lambda \in \cup_{i=1}^{\infty} f^{-1}(A_i)$. Then for some j, $\lambda \in f^{-1}(A_j)$, so $f(\lambda) \in A_j$ and $f(\lambda) \in \cup_{i=1}^{\infty} A_i$; hence $\lambda \in f^{-1}(\cup_{i=1}^{\infty} A_i)$. We have $\cup_{i=1}^{\infty} f^{-1}(A_i) \subset f^{-1}(\cup_{i=1}^{\infty} A_i)$.

Measurable Functions

If $(\Omega, \mathcal{F})$ and $(\Lambda, \mathcal{G})$ are measurable spaces, and f is a mapping from Ω to Λ, with the property that $\forall A \in \mathcal{G}, f^{-1}(A) \in \mathcal{F}$, then f is a *measurable* function with respect to $\mathcal{F}$ and $\mathcal{G}$. It is also said to be measurable $\mathcal{F}/\mathcal{G}$

Note that a measurable function $f(\cdot)$ does not depend on a measure.

The domain of $f(\cdot)$ has no relationship to $\mathcal{F}$, except through the range of $f(\cdot)$ that happens to be in the subsets in $\mathcal{G}$.

Product Spaces

Two measurable spaces $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ can be used to form a *cartesian product measurable space* with sample space $\Omega_1 \times \Omega_2$. The product of the σ-fields is not necessarily a σ-field.

Given $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$, we define the cartesian product measurable space as
$(\Omega_1 \times \Omega_2, \sigma(\mathcal{F}_1 \times \mathcal{F}_2))$.

This provides us the basis for developing a probability theory for vectors and multivariate distributions.

The Borel σ-field

Consider the sample space $\mathbb{R}$, and let $\mathcal{C}$ be the collection of all open intervals in $\mathbb{R}$. The σ-field $\sigma(\mathcal{C})$ is called the Borel σ-field, and is usually denoted by $\mathcal{B}$.

Borel Sets

Any set in $\mathcal{B}$ is called a Borel set.
The following are Borel sets:

- $\mathbb{R}$
- $\emptyset$
- any countable set; in particular, any finite set, $\mathbb{Z}$, $\mathbb{Z}_+$ (the *natural numbers*), and the set of all rational numbers
- hence, from the foregoing, the set of all irrational numbers (which is uncountable)
- any interval, open, closed, or neither

- the Cantor set
  We see this by writing the Cantor set as $\cap_{i=1}^{\infty} C_i$, where

  $$C_1 = [0, 1/3] \cup [2/3, 1], \quad C_2 = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1],$$

  and realizing that each of these is Borel.

- any union of any of the above

So, are all subsets of $\mathbb{R}$ Borel sets?
No. Interestingly enough, the cardinality of $\mathcal{B}$ can be shown to be the same as that of $\mathbb{R}$, and the cardinality of the collection of all subsets of $\mathbb{R}$, that is, the cardinality of the power set, $2^{\mathbb{R}}$, is much larger – which means there are *many* subsets of $\mathbb{R}$ that are not Borel sets. Construction of a non-Borel set uses the Axiom of Choice, which can be used to construct some truly weird sets.

Product Borel σ-fields

For the product measurable space generated by $\mathbb{R}^k$, a σ-field is $\sigma(\mathcal{B}^k)$. It can be shown that this is the same as the σ-field generated by all open intervals (or "hyperrectangles") in $\mathbb{R}^k$. We denote this product measurable space as $(\mathbb{R}^k, \mathcal{B}^k)$.

The σ-field $\mathcal{B}_{[0,1]}$

This is the σ-field generated by all open intervals on $[0, 1]$.

Borel Function

A measurable function from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-field, is said to be *Borel measurable* with respect to $\mathcal{F}$. A function that is Borel measurable is called a *Borel function*.

Random Variables

A measurable function from any measurable space $(\Omega, \mathcal{F})$ to the measurable space $(\mathbb{R}, \mathcal{B})$ is called a *random variable*. That is, “Borel function” and “random variable” are synonymous.

This is the definition of the phrase *random variable*; the words “random” and “variable” do not carry any separate meaning.

Measures

A measure is a real-valued nonnegative set function whose domain is a σ-field and with the property that the measure of the union of any collection of disjoint sets is the sum of the measures of the sets.

From this simple definition, several properties derive. For example, if ν is a measure with domain $\mathcal{F}$ then

- $\nu(\emptyset) = 0$

- if $A_1 \subset A_2 \in \mathcal{F}$, then $\nu(A_1) \leq \nu(A_2)$ (this is "monotonicity")

- if $A_1, A_2, \ldots \in \mathcal{F}$, then $\nu(\cup_i A_i) \leq \sum_i \nu(A_i)$ (this is "subadditivity")

- if $A_1 \subset A_2 \subset \ldots \in \mathcal{F}$, then $\nu(\cup_{i=1}^{\infty} A_i) = \lim_{i\to\infty} \nu(A_i)$ (this is "continuity from below"; think of the A_is as nested intervals).

  We can see this by defining a sequence of disjoint sets, $B_1, B_2, \ldots$ as $B_j = A_{j+1} - A_j$, so $\cup_{j=1}^{i} B_j = A_i$, and

$$\cup_{i=1}^{\infty} B_i = \cup_{i=1}^{\infty} A_i$$

  Hence,

$$\begin{aligned}
\nu\left(\cup_{i=1}^{\infty} A_i\right) &= \nu\left(\cup_{i=1}^{\infty} B_i\right) \\
&= \sum_{i=1}^{\infty} \nu(B_i) \\
&= \lim_{i\to\infty} \sum_{j=1}^{i} \nu(B_j) \\
&= \lim_{i\to\infty} \nu\left(\cup_{j=1}^{i} B_j\right) \\
&= \lim_{i\to\infty} \nu(A_i).
\end{aligned}$$

Sequences of nested intervals are important. We denote a sequence $A_1 \subset A_2 \subset \ldots$ with $A = \cup_{i=1}^{\infty} A_i$, as $A_i \nearrow A$. (This same notation is used for a sequence of real numbers x_i such that $x_1 \leq x_2 \cdots$ and $\lim x_i = x$; that is, in that case, we write $x_i \nearrow x$.)

Continuity from below is actually a little stronger than what is stated above, because the sequence of values of the measure is also monotonic: for
$A_i \in \mathcal{F}, A_i \nearrow A \Rightarrow \nu(A_i) \nearrow \nu(A)$

A similar sequence is $A_1 \supset A_2 \supset \ldots$ with $A = \cup_{i=1}^{\infty} A_i$. We denote this as $A_i \searrow A$. Continuity from above is the fact that for $A_i \in \mathcal{F}$, if $A_i \searrow A$ and $\nu(A_1) < \infty$, then $\nu(A_i) \searrow \nu(A)$.

To evaluate $\nu(\cup_i A_i)$ we form disjoint sets by intersections. For example, we have $\nu(A_1 \cup A_2) = \nu(A_1) + \nu(A_2) - \nu(A_1 \cap A_2)$. This is the simplest form of the inclusion-exclusion formula.

If there are three sets, we take out all pairwise intersections and then add back in the triple intersection.

We can easily extend this (the proof is by induction) so that, in general, we have

$$\begin{aligned}
\nu(\cup_i^n A_i) \;=\; & \textstyle\sum_{1 \le i \le n} \nu(A_i) - \\
& \textstyle\sum_{1 \le i < j \le n} \nu(A_i \cap A_j) + \\
& \textstyle\sum_{1 \le i < j < k \le n} \nu(A_i \cap A_j \cap A_k) - \\
& \cdots + \\
& (-1)^{n+1} \nu(A_1 \cap \cdots \cap A_n).
\end{aligned}$$

If $\mathcal{F}$ is the Borel σ-field, the most common measure is the *Lebesgue measure*, which is defined by the relation

$$\nu((a, b)) = b - a.$$

If $\mathcal{F}$ is a countable σ-field, the most common measure is the *counting measure*, which is defined by the relation

$$\nu(A) = \#(A).$$

If $\mathcal{F}$ is the collection of all subsets of Ω, a useful measure at a fixed point $\omega \in \Omega$ is the *Dirac measure* concentrated at ω, usually denoted by δ_ω, and defined by

$$\delta_\omega(A) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise.} \end{cases}$$

Measurable Sets

If ν is a measure with domain $\mathcal{F}$ then every set in $\mathcal{F}$ is said to be ν-*measurable*, or just *measurable*.

Note that unlike other terms above that involve "measurable", this term is defined in terms of a given measure.

Measure Space: The Structure $(\Omega, \mathcal{F}, \nu)$

If Ω is a sample space, $\mathcal{F}$ is a σ-field over Ω, and ν is a measure with domain $\mathcal{F}$, the triple $(\Omega, \mathcal{F}, \nu)$ is called a *measure space* (compare *measurable space*, above).

The elements in the measure space can be any kind of objects. They do not need to be numbers.

σ-Finite Measure

A measure ν is σ-*finite* on $(\Omega, \mathcal{F})$ iff there exists a sequence $A_1, A_2, \ldots$ in $\mathcal{F}$ such that $\cup_i A_i = \Omega$ and $\nu(A_i) < \infty$ for all i.

The Lebesgue measure is σ-finite, as can be seen from the sequence of open intervals $(-i, i)$.

The counting measures is σ-finite iff Ω is countable.

Almost Everywhere (a.e.)

Given a measure space, a property that holds for any element of the σ-field with positive measure is said to hold *almost everywhere*, or a.e.

Probability Measure

A measure whose domain is a σ-field defined on the sample space Ω with the property that $\nu(\Omega) = 1$ is called a *probability measure*. We often use P to denote such a measure.

A property that holds a.e. with respect to a probability measure is said to hold *almost surely*, or a.s.

Support of a Pprobability Measure

If the probability measure P is defined with respect to the σ-field $\mathcal{F}$, $S \in \mathcal{F}$, and $P(S) = 1$, S is called a *support* of the probability measure (or of the probability space, or of the "distribution"). If S_1 and S_2 are supports of P, then $S_1 = S_2$ a.s.

Probability Space

If P in the measure space $(\Omega, \mathcal{F}, P)$ is a probability measure, the triple $(\Omega, \mathcal{F}, P)$ is called a *probability space*.

The elements in the probability space can be any kind of objects. They do not need to be numbers.

Cumulative Distribution Function (CDF)

If $(\mathbb{R}, \mathcal{B}, P)$ is a probability space, and F is defined by

$$F(x) = P((-\infty, x]) \quad \forall\, x \in \mathbb{R},$$

then F is called a cumulative distribution function.

The probability space completely determines F, and likewise, F completely determines P a.s.

Induced Measure

If $(\Omega, \mathcal{F}, \nu)$ is a measure space and $(\Lambda, \mathcal{G})$ is a measurable space, and f is a function from Ω to Λ that is measurable with respect to $\mathcal{F}$, then $\nu \circ f^{-1}$ is an *induced measure* on $\mathcal{G}$.

Note that the domain and range of the function $\nu \circ f^{-1}$ is $\mathcal{G}$. Note, further, that it is a measure.

Simple Functions

If $A_1, \ldots, A_k$ are measurable subsets of Ω and $a_1, \ldots, a_k$ are constant real numbers, a function φ is a *simple function* if for $\omega \in \Omega$,

$$\varphi(\omega) = \sum_{i=1}^{k} a_i \mathrm{I}_{A_i}(\omega),$$

where $\mathrm{I}_S(x)$ is the *indicator function*, i.e., $\mathrm{I}_S(x) = 1$ if $x \in S$ and $\mathrm{I}_S(x) = 0$ otherwise.

Real-Valued Functions over Real Domains

Random variables are real-valued functions

Measurable real-valued functions are called Borel functions.

Now we will define integrals and derivatives of real-valued functions.

For real-valued functions over real domains we identify some additional properties.

We have already defined similar properties for functions that are measures (real-valued set functions), and we could define them for other functions from one measure space to another. The definitions for real-valued functions over real domains are simpler and more useful.

Continuous Functions

Let f be a real-valued function whose domain is a set $D \in \mathbb{R}$. We say that f is *continuous at the point* $x \in D$ if, given $\epsilon > 0, \exists \delta \ni \forall y \in D \ni |x - y| < \delta, |f(x) - f(y)| < \epsilon$.

If f is continuous at each point in a subset of its domain, we say it is continuous on that subset.

If f is continuous at each point in its domain, we say it is *continuous*.

From this definition we have an immediate useful fact about continuous functions: the inverse image of an open set is open.

Absolutely Continuous Functions

Let f be a real-valued function defined on $[a, b]$ (its domain may be larger). We say that f is *absolutely continuous* on $[a, b]$ if, if, given $\epsilon > 0$, there exists a δ such that for every finite collection of nonoverlapping intervals $(x_i, y_i) \subset [a, b]$ with

$$\sum_{i=1}^{n} |x_i - y_i| < \delta,$$

$$\sum_{i=1}^{n} |f(x_i) - f(y_i)| < \epsilon.$$

If f is absolutely continuous on $[a, b]$, it is continuous on $[a, b]$.

An absolutely continuous function is of bounded variation; it has a derivative almost everywhere; and if the derivative is 0 a.e., the function is constant.

Simple Functions

Let f be real and measurable. If $f(\omega) \geq 0$, there exists a sequence $\{f_n\}$ of simple functions such that

$$0 \leq f_n(\omega) \nearrow f(\omega) \quad \text{a.s.},$$

and if $f(\omega) \leq 0$, there exists a sequence $\{f_n\}$ of simple functions such that

$$0 \geq f_n(\omega) \searrow f(\omega) \quad \text{a.s.}$$

The sequence is

$$f_n(\omega) = \begin{cases} -n & \text{if} \quad f(\omega) \leq -n, \\ -(k-1)2^{-n} & \text{if} \quad -k2^{-n} < f(\omega) \leq -(k-1)2^{-n}, \text{ for } 1 \leq k \leq n \\ (k-1)2^{-n} & \text{if} \quad (k-1)2^{-n} < f(\omega) < k2^{-n}, \text{ for } 1 \leq k \leq n2^{-n} \\ n & \text{if} \quad n \leq f(\omega). \end{cases}$$

In particular, if X is a nonnegative random variable, there exists a sequence of simple (degenerate) random variables $\{X_n\}$ such that

$$0 \leq X_n \nearrow X \quad \text{a.e.}$$

Integration

Integrals are some of the most important functionals of real-valued functions. Integrals and the action of integration are defined using measures.

Integrals of nonnegative functions are themselves measures.

There are various types of integrals, Lebesgue, Riemann, Riemann-Stieltjes, Ito, and so on. The most important is the Lebesgue, and when we use the term "integral" without qualification that will be the integral meant.

The (Lebesgue) Integral of a Function with Respect to a Given Measure

An *integral of a function f with respect to a given measure* ν, if it exists, is a functional whose value is an average of the function weighted by the measure. It is denoted by $\int f \,\mathrm{d}\nu$.

We build the definition of an integral of a function in three steps.

1. Simple function.
   If f is a simple function defined as $f(\omega) = \sum_{i=1}^{k} a_i \mathrm{I}_{A_i}(\omega)$, where the A_is are measurable with respect to ν, then

   $$\int f \,\mathrm{d}\nu = \sum_{i=1}^{k} a_i \nu(A_i).$$

   (Note that a simple function over measurable A_is is measurable.)

2. Nonnegative Borel function.
We define the integral of a nonnegative Borel function in terms of the supremum of a collection of simple functions. Let f be a nonnegative Borel function with respect to ν on Ω, and let S_f be the collection of all nonnegative simple functions such that

$$\varphi \in S_f \Rightarrow \varphi(\omega) \leq f(\omega) \forall\, \omega \in \Omega$$

We define the integral of f with respect to ν as

$$\int f \,\mathrm{d}\nu = \sup \left\{ \int \varphi \,\mathrm{d}\nu | \varphi \in S_f \right\}.$$

3. General Borel function.

For a general Borel function f, we form two nonnegative Borel functions f_+ and f_- such that $f = f_+ - f_-$:

$$f_+(\omega) = \max\{f(\omega), 0\}$$

$$f_-(\omega) = \max\{-f(\omega), 0\}$$

We define the integral of f with respect to ν as the difference of the integrals of the two nonnegative functions:

$$\int f \, \mathrm{d}\nu = \int f_+ \, \mathrm{d}\nu - \int f_- \, \mathrm{d}\nu,$$

so long as either $\int f_+ \, \mathrm{d}\nu$ or $\int f_- \, \mathrm{d}\nu$ is finite. ($\infty - \infty$ is not defined.)

The (Lebesgue) Integral of a Function with Respect to a Given Measure

This definition of an integral immediately yields some important properties of integrals:

- linearity: for real a and Borel f and g,
  $\int af + g \,\mathrm{d}\nu = a \int f \,\mathrm{d}\nu + \int g \,\mathrm{d}\nu$.

- $\int |f| \,\mathrm{d}\nu$ is a norm: $\int |f| \,\mathrm{d}\nu = 0 \Rightarrow f = 0$ a.e.
  This fact together with the linearity means that $\int |f| \,\mathrm{d}\nu$ is a *norm* for functions. A more general form norm is $(\int |f|^p \,\mathrm{d}\nu)^{1/p}$ for $1 \leq p$.

- finite monotonicity: for integrable f and g,
  $f \leq g$ a.e. $\Rightarrow \int f \,\mathrm{d}\nu \leq \int g \,\mathrm{d}\nu$.

Integration

There are some conditions for interchange of an integration operation and a limit operation that are not so obvious.

- monotone convergence: if $0 \le f_1 \le f_2$ and $\lim_{n\to\infty} f_n = f$ a.e., then $\int \lim_{n\to\infty} f_n \, \mathrm{d}\nu = \lim_{n\to\infty} \int f_n \, \mathrm{d}\nu$.

- Fatou's lemma (follows from finite monotonicity and monotone convergence): if $0 \le f_n$ then $\int \lim_n \inf f_n \, \mathrm{d}\nu \le \lim_n \inf \int f_n \, \mathrm{d}\nu$.

- dominated convergence: if $\lim_{n\to\infty} f_n = f$ a.e. and there exists an integrable function g such that $|f_n| \le g$ a.e., then $\int \lim_{n\to\infty} f_n \, \mathrm{d}\nu = \lim_{n\to\infty} \int f_n \, \mathrm{d}\nu$.

Shao gives proofs of these (but he should prove them in the order given above).

Integration

If the measure is a probability measure P with associated CDF F, all of the following notations are equivalent:
$\int f \, \mathrm{d}P$, $\int f(x) \, \mathrm{d}P$, $\int f \, \mathrm{d}F$, $\int f(x) \, \mathrm{d}F(x)$

The Riemann integral is one of the simplest integrals. We can define the Riemann integral of a real function f over the interval $(a, b]$ in terms of the Lebesgue measure λ as the real number r such that for any $\epsilon > 0$, there exists a δ such that

$$\left| r - \sum_i f(x_i)\lambda(I_i) \right| < \epsilon$$

where $\{I_i\}$ is any finite partition of $(a, b]$ such that for each i, $\lambda(I_i) < \delta$ and $x_i \in I_i$. If the Riemann integral exists, it is the same as the Lebesgue integral.

A classic example for which the Lebesgue integral exists, but the Riemann integral does not, is the function g defined over $(0, 1]$ as $g(x) = 1$ is x is rational, and $g(x) = 0$ otherwise.

The Lebesgue integral $\int_0^1 g(x)\,\mathrm{d}x$ exists and equals 0, because $g(x) = 0$ a.e.

The Riemann integral, on the other had does not exist because for an arbitrary partition $\{I_i\}$, the integral is 1 if $x_i \in I_i$ is taken as a rational, and the integral is 0 if $x_i \in I_i$ is taken as a irrational.

Integrable Function

If $f = f_+ - f_-$, where f_+ and f_- are nonnegative Borel functions, and both $\int f_+ \, d\nu$ and $\int f_- \, d\nu$ are finite, we say f is *integrable*.

Note that being Borel does not imply that a function is integrable.

A random variable is not necessarily integrable.

Change of Variables

Consider two measurable spaces $(\Omega, \mathcal{F})$ and $(\Lambda, \mathcal{G})$, let f be a measurable function from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$, and let ν be a measure on $\mathcal{F}$. As we have seen, $\nu \circ f^{-1}$ is an induced measure on $\mathcal{G}$.

Now let g be a Borel function on $(\Lambda, \mathcal{G})$.

Then the integral of $g \circ f$ over Ω with respect to ν is the same as the integral of g over Λ with respect to $\nu \circ f^{-1}$:

$$\int_\Omega g \circ f \, \mathrm{d}\nu = \int_\Lambda g \, \mathrm{d}(\nu \circ f^{-1})$$

Integration in a Product Space (Fubini's Theorem)

Given two measure spaces $(\Omega_1, \mathcal{F}_1, \nu_1)$ and $(\Omega_2, \mathcal{F}_2, \nu_2)$ and a Borel function f on $\Omega_1 \times \Omega_2$, the integral over Ω_1, if it exists, is a function of $\omega_2 \in \Omega_2$ a.e., and likewise, the integral over Ω_2, if it exists, is a function of $\omega_2 \in \Omega_1$ a.e.

Fubini's theorem shows that if one of these marginal integrals, say

$$g(\omega_2) = \int_{\Omega_1} f(\omega_1, \omega_2)\,\mathrm{d}\nu_1,$$

exists a.e., then the natural extension of an integral to a product space, resulting in the *double integral*,

$$\int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2)\,\mathrm{d}\nu_1 \times \mathrm{d}\nu_2$$

is the same as the *iterated integral*,

$$\int_{\Omega_2}\left(\int_{\Omega_1} f(\omega_1, \omega_2)\, \mathrm{d}\nu_1\right) \mathrm{d}\nu_2.$$

The Integral over a Domain

For $A \subset \Omega$, we define

$$\int_A f \, \mathrm{d}\nu = \int \mathrm{I}_A f \, \mathrm{d}\nu$$

If f is a nonnegative Borel function, then $\int_A f \, \mathrm{d}\nu$ for $A \subset \Omega$ is a measure over Ω.

[Probability of an Event

Given a probability space $(\Omega, \mathcal{F}, P)$, a set $A \in \mathcal{F}$ is called an "event", and $\int_A \mathrm{d}P$ is the *probability of* A, written $P(A)$ or $\Pr(A)$.

Expected Value

Given a probability space $(\Omega, \mathcal{F}, P)$ and a random variable with respect to $\mathcal{F}$, X, we define the *expected value* of X with respect to P as

$$\int X \, \mathrm{d}P,$$

and denote it as $\mathrm{E}(X)$ or for clarity, $\mathrm{E}_P(X)$.
Sometimes we limit this definition to integrable random variables X.

Conditional Expectation over a Sub-σ-Field

Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\mathcal{A}$ be a sub-σ-field of $\mathcal{F}$, and let X be an integrable random variable over Ω.

The *conditional expectation* of X given $\mathcal{A}$, denoted by $\mathrm{E}(X|\mathcal{A})$ is a random variable such that $\mathrm{E}(X|\mathcal{A})$ is a measurable function from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B})$ and $\int_A \mathrm{E}(X|\mathcal{A})\,\mathrm{d}P = \int_A X\,\mathrm{d}P$ for any $A \in \mathcal{A}$. (The existence and uniqueness of this random variable follows from the Radon-Nikodym theorem (Theorem 1.4)).

Conditional Probability

Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\mathcal{A}$ be a sub-σ-field of $\mathcal{F}$, and let $B \in \mathcal{F}$. The *conditional probability* of B given $\mathcal{A}$, denoted by $\mathrm{Pr}(B|\mathcal{A})$ is defined as $\mathrm{E}(\mathrm{I}_B|\mathcal{A})$.

Conditional Expectation with Respect to Another Measurable Function

Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\mathcal{A}$ be a sub-σ-field of $\mathcal{F}$, let X be an integrable random variable over Ω, and let Y be a measurable function from $(\Omega, \mathcal{F}, P)$ to any measurable space $(\Lambda, \mathcal{G})$. Then the *conditional expectation* of X given Y, denoted by $\mathrm{E}(X|Y)$ is defined as the conditional expectation of X given the sub-σ-field generated by Y, that is, $\mathrm{E}(X|\sigma(Y))$.

Continuity

Given two measures ν and λ on the same measurable space, $(\Omega, \mathcal{F})$, if $\forall A \in \mathcal{F}$

$$\nu(A) = 0 \quad \Rightarrow \quad \lambda(A) = 0,$$

then λ is said to be *absolutely continuous* with respect to ν. We denote that λ is said to be *absolutely continuous* with respect to ν by

$$\lambda \ll \nu.$$

As an example, let $(\Omega, \mathcal{F}, \nu)$ be a measure space and let f be a nonnegative Borel function on Ω. Define the measure λ by

$$\lambda(A) = \int_A f \, \mathrm{d}\nu$$

for any $A \in \mathcal{F}$. Then $\nu(A) = 0 \Rightarrow \lambda(A) = 0$, and so λ is absolutely continuous with respect to ν.

Radon-Nikodym Theorem

Given two measures ν and λ on the same measurable space, $(\Omega, \mathcal{F})$, such that $\lambda \ll \nu$ and ν is σ-finite. Then there exists a unique a.e. nonnegative Borel function f on Ω such that $\lambda(A) = \int_A f \mathrm{d}\nu \, \forall \, A \in \mathcal{F}$.
Uniqueness a.e. means that if also, for some g, $\lambda(A) = \int_A g \mathrm{d}\nu \, \forall \, A \in \mathcal{F}$ then $f = g$ a.e.

The proof of the Radon-Nikodym theorem is complicated.

Radon-Nikodym Derivative

If $\lambda(A) = \int_A f \mathrm{d}\nu \; \forall A \in \mathcal{F}$, then f is called the Radon-Nikodym derivative of λ with respect to ν, and we write $f = \mathrm{d}\lambda/\mathrm{d}\nu$.