Integration

References: Bass (Real Analysis for Graduate Students), Folland (Real Analysis), Athreya and Lahiri (Measure Theory and Probability Theory).

1 Measurable Functions

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

Definition 1 *A function $T : \Omega_1 \to \Omega_2$ is $(\mathcal{F}_1, \mathcal{F}_2)$-measurable if for every $E \in \mathcal{F}_2$, $T^{-1}(E) \in \mathcal{F}_1$.*

Terminology: If $(\Omega, \mathcal{F})$ is a measurable space and f is a real-valued function on Ω, it's called $\mathcal{F}$-measurable or simply measurable, if it is $(\mathcal{F}, \mathcal{B}(\Re))$-measurable. A function $f : \Re \to \Re$ is called Borel measurable if the σ-algebra used on the domain and codomain is $\mathcal{B}(\Re)$. If the σ-algebra on the domain is Lebesgue, f is called Lebesgue measurable.

Example 1 *Measurability of a function is related to the σ-algebras that are chosen in the domain and codomain. Let $\Omega = \{0, 1\}$. If the σ-algebra is $\mathcal{P}(\Omega)$, every real valued function is measurable. Indeed, let $f : \Omega \to \Re$, and $E \in \mathcal{B}(\Re)$. It is clear that $f^{-1}(E) \in \mathcal{P}(\Omega)$ (this includes the case where $f^{-1}(E) = \emptyset$).*

However, if $\mathcal{F} = \{\emptyset, \Omega\}$ is the σ-algebra, only the constant functions are measurable. Indeed, if $f(x) = a, \forall x \in \Omega$, then for any Borel set E containing a, $f^{-1}(E) = \Omega \in \mathcal{F}$. But if f is a function s.t. $f(0) \neq f(1)$, then, any Borel set E containing $f(0)$ but not $f(1)$ will satisfy $f^{-1}(E) = \{0\} \notin \mathcal{F}$.

It is hard to check for measurability of a function using the definition, because it requires checking the preimages of *all* sets in $\mathcal{F}_2$. The following proposition relaxes this.

Proposition 1 *Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be measurable spaces, where $\mathcal{F}_2$ is the σ-algebra generated by a collection $\mathcal{E}$ of subsets of Ω_2. Then, f is $(\mathcal{F}_1, \mathcal{F}_2)$-measurable if and only if $f^{-1}(E) \in \mathcal{F}_1$ for every $E \in \mathcal{E}$.*

Proof. Only if is obvious. If part: Suppose $f^{-1}(E) \in \mathcal{F}_1$, for all $E \in \mathcal{E}$. Note that $\mathcal{V} = \{E \subset \Omega_2 : f^{-1}(E) \in \mathcal{F}_1\}$ is a σ-algebra. (Indeed, if $E \in \mathcal{V}$, $f^{-1}(E) \in \mathcal{F}_1$, so its complement $(f^{-1}(E))^c \in \mathcal{F}_1$, and note that $(f^{-1}(E))^c = f^{-1}(E^c)$. So, $E^c \in \mathcal{V}$. Similarly, if $(E_j)_1^\infty$ are all sets in $\mathcal{V}$, their union is also in $\mathcal{V}$, because for all j, $f^{-1}(E_j) \in \mathcal{F}_1$, and $f^{-1}(\cup_1^\infty E_j) = \cup_1^\infty f^{-1}(E_j)$.)

Since $\mathcal{V}$ is a σ-algebra containing $\mathcal{E}$, it also contains $\mathcal{F}_2$, the σ-algebra generated by $\mathcal{E}$. So by definition of $\mathcal{V}$, it must be that whenever $E \in \mathcal{F}_2$, $f^{-1}(E) \in \mathcal{F}_1$. ■

This proposition has the following consequence.

Corollary 1 *If S and T are topological spaces and $f : S \to T$ is continuous, then it is $(\mathcal{B}(S), \mathcal{B}(T))$-measurable.*

Proof. Let E be an open set in T. Since f is continuous, $f^{-1}(E)$ is an open set in S, and is therefore in $\mathcal{B}(S)$. $\mathcal{B}(T)$ is the σ-algebra generated by the collection of open sets in T, so now apply the above proposition. ■

The next proposition is another implication of Proposition 1. It has to do with equivalent definitions of measurability when a function f maps to $\Re$ (and we have the Borel σ-algebra on $\Re$). Before stating the proposition, note generally that for a topological space (S, τ), the Borel σ-algebra $\mathcal{B}(S)$ is a rich collection. It has all the open sets of S, their complements the closed sets, all countable intersections of open sets (these are called G_δ sets, G for open and δ for Durchschnitt or intersection); their complements the countable unions of closed sets (called F_σ sets, F for closed, σ for union or sum); countable unions of G_δ sets (called $G_{\sigma\delta}$ sets), and so on. It turns out that this richness makes $\mathcal{B}(S)$ the σ-algebra generated by a lot of other families of sets, as in the proposition below.

Proposition 2 *Let $(S, \mathcal{S})$ be a measurable space, and $f : S \to \Re$. Let $\mathcal{E}_5 - \mathcal{E}_8$ be respectively, the families of all open and closed rays of types $(a, \infty), (-\infty, a), [a, \infty), (-\infty, a]$. Then, "$f$ is measurable" is equivalent to saying that $f^{-1}(E) \in \mathcal{S}$, for every set E in any of these families.*

Proof. Take, say, $\mathcal{E}_8$. We claim $\mathcal{B}(\Re)$ is the σ-algebra generated by $\mathcal{E}_8$. Indeed, for any $a \in \Re$, $(-\infty, a] = \cap_{n=1}^{\infty}(-\infty, a + \frac{1}{n})$, that is $(-\infty, a]$ is a G_δ set. So $\mathcal{E}_8 \subset \mathcal{B}(\Re)$, so $\sigma(\mathcal{E}_8) \subset \mathcal{B}(\Re)$. On the other hand, let E be an open set in $\Re$. Then for some countable collection of open intervals (a_n, b_n), we have $E = \cup_{n=1}^{\infty}(a_n, b_n)$. Now, every $(a_n, b_n) = \cup_{j=k}^{\infty}\left((-\infty, b_n - \frac{1}{j}] - (-\infty, a_n]\right)$, ($k$ is chosen so that $a_n < b_n - (1/k)$); and therefore belongs to $\sigma(\mathcal{E}_8)$; therefore so does their union. So every open set $E \subset \sigma(\mathcal{E}_8)$. So $\mathcal{B}(\Re) \subset \sigma(\mathcal{E}_8)$.

So by Proposition 1 above, f is $\mathcal{S}$-measurable. ∎

Note that from $\mathcal{E}_j, j = 7, 8$, f is measurable if we can show that for every a, the lower (or upper) contour set of f w.r.t. is a set in $\mathcal{S}$. This motivates the following definition.

Definition 2 *Let $(\Omega, \mathcal{F}, \mu)$ be a probability space. A* **random variable** *is a function $X : \Omega \to \Re$ s.t. $X^{-1}(-\infty, a] \equiv \{\omega \in \Omega | X(\omega) \in (-\infty, a]\} \in \mathcal{F}$, for all $a \in \Re$.*

It is straightforward to check that a composition of measurable functions is a measurable function. Now some other standard results on measurability of functions, in the next 2 propositions.

Proposition 3 *Let* $f_1, ..., f_k$ *be* $(\mathcal{F}, \mathcal{B}(\Re))$*-measurable functions from* Ω *to* $\Re$*. Then*

(i) $f = (f_1, ..., f_k)$ *is* $(\mathcal{F}, \mathcal{B}(\Re^k))$*-measurable.*

(ii) $g = f_1 + ... + f_k$ *is* $(\mathcal{F}, \mathcal{B}(\Re))$*-measurable.*

(iii) $h = \Pi_{i=1}^k f_i$ *is* $(\mathcal{F}, \mathcal{B}(\Re))$*-measurable.*

(iv) $\xi = \psi \circ f$ *is* $(\mathcal{F}, \mathcal{B}(\Re^p))$*-measurable, where* $f = (f_1, ..., f_k)$ *and* $\psi : \Re^k \to \Re^p$ *is a continuous function.*

Proof. (i) We know that the σ-algebra generated by the class of open rectangles is $\mathcal{B}(\Re^k)$. So we show that for every open rectangle $R = (a_1, b_1) \times ... \times (a_n, b_n)$, its preimage $f^{-1}(R) \in \mathcal{F}$. Indeed,

$$f^{-1}(R) = \{\omega \in \Omega | a_1 < f_1(\omega) < b_1, ..., a_k < f_k(\omega) < b_k\}$$

$$= \cap_{i=1}^k \{\omega \in \Omega | a_i < f_i(\omega) < b_i\}$$

$$= [\cap_{i=1}^k f_i^{-1}((a_i, b_i))] \in \mathcal{F}$$

since each f_i is measurable.

(ii) $g(x) = f_1(x) + ... + f_k(x) = g_1(f(x))$, where $g_1(y) = y_1 + ... + y_k$. Since f is measurable and g_1 is linear, hence continuous, hence measurable, and g is a composition of measurable functions, it is measurable.

(iii) and (iv). Similar proofs as for (ii). ∎

There are 2 other standard results about measurable functions. First, the sup, inf, limit etc. of a sequence of measurable functions is a measurable function. Second, that the (pointwise) limit function of a sequence of measurable functions, if it exists, is measurable. To state these, note that $\bar{\Re} = \Re \cup \{\infty, -\infty\}$, and $\mathcal{B}(\bar{\Re}) = \sigma\langle \mathcal{B}(\Re) \cup \{\infty\} \cup \{-\infty\}\rangle$.

Proposition 4 *Let $(S, \mathcal{S})$ be a measurable space and (f_j) a sequence of $\bar{\Re}$-valued, $(\mathcal{S}, \mathcal{B}(\bar{\Re}))$-measurable functions on S. Then g_1, g_2, g_3, g_4 defined below are measurable.*

$$g_1(x) = \sup_j f_j(x), g_2(x) = \inf_j f_j(x), g_3(x) = \lim \sup_{j\to\infty} f_j(x), g_4(x) = \lim \inf f_j(x)$$

Moreover, if $f(x) = \lim_{j\to\infty} f_j(x)$ exists for all $x \in S$, then f is measurable.

Proof. Let $a \in \Re$. $g_1^{-1}((a,\infty])$ is the set of all $x \in S$ s.t. $g_1(x) > a$. For any such x, $f_j(x) > a$ for some j. Conversely, if $f_j(x) > a$ for at least one x, then the sup $g_1(x) > a$. So $g_1^{-1}((a,\infty]) = \cup_1^\infty f_j^{-1}((a,\infty])$. By Proposition 2, all the sets on the right are in $\mathcal{S}$, hence so is their countable union, and so again by Proposition 2, g_1 is measurable. Similarly, $g_2^{-1}([-\infty, a)) = \cup_1^\infty f_j^{-1}([-\infty, a))$, and so g_2 is measurable. Similarly, h_k, defined by $h_k(x) = \sup_{j>k} f_j(x)$ is measurable for each k. Since $g_3(x) = \inf_{k\geq 1} h_k(x)$, g_3 is measurable. Similarly for g_4. And if the pointwise limit function f exists, $f = g_3 = g_4$ and so it is measurable. ■

Note therefore that the max and min of functions is measurable.

This section closes with introducing the notion of the σ-algebra generated by a function or family of functions.

Definition 3 *Let $\{f_\lambda : \lambda \in \Lambda\}$ be a family of functions from Ω_1 to Ω_2, and let $\mathcal{F}_2$ be a σ-algebra on Ω_2. Then*

$$\sigma\langle\{f_\lambda^{-1}(A) | A \in \mathcal{F}_2, \lambda \in \Lambda\}\rangle$$

is called the **σ-algebra generated by** $\{f_\lambda | \lambda \in \Lambda\}$ **w.r.t.** $\mathcal{F}_2$.

It is denoted $\sigma\langle\{f_\lambda | \lambda \in \Lambda\}\rangle$. Note that this is the **smallest** σ-algebra that makes all the f_λ's measurable (wrt $\mathcal{F}_2$).

2 Induced Measures and Probability Distributions

This will be fleshed out properly only later. For a short introduction, you could refer to Chapter 21 of Bass.

Let $(S, \mathcal{S}, \mu)$ be a probability space and let $X : S \to \Re$ be a r.v. Then the collection of sets

$$X^{-1}(\mathcal{B}) = \{A \subset \mathcal{S} : A = X^{-1}(B) \text{ for some B} \in \mathcal{B}\}$$

is a σ-algebra. We say this is the σ-algebra generated by the r.v. X, and denote it by $\mathcal{S}_X$.

Now, on the measurable space $(\Re, \mathcal{B})$, we define the measure P_X, the probability distribution of the r.v. X, by setting, for every Borel set B,

$$P_X(B) = \mu(X^{-1}(B))$$

(It can be shown that P_X is a (probability) measure, so that $(\Re, \mathcal{B}, \mathcal{P_X})$ is a probability space).

Definition 4 *Two r.v.s X, Y are independent if the σ-algebras generated by them are independent. That is, for $A, B \in \mathcal{B}$,*

$$\mu(X^{-1}(A) \cap Y^{-1}(B)) = \mu(X^{-1}(A))\mu(Y^{-1}(B))$$

3 Integration

Riemann Integral: We partition the domain of the function, and take the sum of areas of lower rectangles. We do this for increasingly fine partitions, getting an increasing sequence of numbers (each being a lower sum): we take the supremum over lower sums over **all** partitions, or what's the same thing, the limit of any increasing sequence from the above iterative process.

Alternatively, we could take sums of areas of upper rectangles: finer partitions would give a decreasing sequence of numbers, and we take its infimum. If the Riemann integral exists, these lower and upper limits are equal.

More specifically, consider $f : [a, b] \to [0, \infty)$. To compute the lower Riemann integral, we use sums like $\sum_1^n y_i(a_i - a_{i-1})$, over a partition or grid of points $a = a_0 < a_1 < ... < a_n = b$, with $y_i \leq f(x), x \in [a_{i-1}, a_i], i = 1, ..., n$. The supremum over all grids is the lower Riemann integral.

Now consider the following example.

Example 2 *Let* $f : [0, 1] \to [0, 1]$ *be defined by*

$$f(x) = \begin{cases} 1 & \textit{if } x \textit{ is rational} \\ 0 & \textit{if } x \textit{ is irrational} \end{cases}$$

Then f *is not Riemann integrable. Indeed, take any partition of* $[0, 1]$*. Every member of the partition will have both irrational and rational numbers. So, in computing the lower Riemann integral, the* y_i*'s must all be* ≤ 0*. And the sup is obtained by noticing that we can choose* $y_i = 0, \forall i$*. So the lower Riemann integral equals* 0*. Similarly, the upper Riemann integral equals* 1*.*

On the other hand, the Lebesgue integral of a function f works with sums like $\sum_1^n y_i \lambda(A_i)$, where the grid is on the **vertical axis**, given by $0 = y_1 < ... < y_n$, $A_i = \{x : f(x) \in [y_i, y_{i+1})\}$, and $\lambda(A_i)$ is the Lebesgue measure of A_i. (Note that $A_i = f^{-1}([y_i, y_{i+1})))$. A limiting operation over increasingly fine vertical axis grids is then performed. A diagram makes it visually clear that when the Riemann integral exists, it equals the Lebesgue integral. In fact for a monotone function, the two approximation procedures can visually produce the same sequence of rectangles.

In the case of the Dirichlet function (the indicator function of the rationals defined in the example above), the Lebesgue integral exists. (Note that this is a measurable function: because the rationals, being a countable union of closed sets, constitute a Borel set, and the irrationals, being their complement, constitute a Borel set as well). Since the Lebesgue measure on each rational number is 0 and the rationals are countable, by countable additivity, the Lebesgue measure on the set of rationals is 0; so that on its complement (the set of irrationals in $[0, 1]$), it is 1. Take any grid on the range of the function, say $0 = y_1 < y_2 < y_3 < y_4 = 1$. Now $A_1 = f^{-1}([y_1, y_2))$ is the set of irrationals; so $\lambda(A_1) = 1$; this will be multiplied by $y_1 = 0$. $A_2 = f^{-1}([y_2, y_3)) = \emptyset$. Its Lebesgue measure is again 0. Finally, $f^{-1}([y_3, y_4]) \equiv A_3$ is the set of rationals, whose Lebesgue measure is 0, by finite additivity. So, the sum $\sum_1^n y_i\lambda(A_i) = 0$. As the grid becomes finer, this remains $= 0$.

Just as the real numbers "complete" the set of rational numbers, and constitute a much larger set of numbers, the functions that are Lebesgue integrable complete the set of Riemann integrable functions. We will do integration with respect to arbitrary measures, of which Lebesgue measure will be a special case.

The sums used to approximate the Lebesgue integral above took values in a finite set. Such functions are called *simple functions*; an increasing sequence of these is used to approximate a Lebesgue (or more general) integral.

Simple Functions.

Let $I_A : S \to \Re$ be the indicator function taking a value of 1 if $x \in A$ and 0 otherwise.

Definition 5 *A simple function is a finite linear combination of indicator functions of sets in $\mathcal{S}$.*

If the sets are in $\mathcal{S}$, the indicator functions are measurable. And since finite linear combinations of measurable functions are measurable (see Exercise 7.4 in Stokey-Lucas-Prescott), simple functions are measurable.

Equivalently, ϕ is a simple function if it is measurable and and its range is a finite subset of $\Re$. Indeed, if $Range(\phi) = \{a_1, ..., a_n\}$ (n distinct elements), then let $A_j = \phi^{-1}(\{a_j\})$. So, $\phi = \sum_1^n a_j I_{A_j}$. This is called the **standard representation** of ϕ. It shows ϕ as a linear combination, with *distinct* coefficients, of indicator functions of *disjoint* measurable sets A_j whose union is S. Note that sums and products of simple functions are simple.

We first start by integrating nonnegative functions, and then generalize, just like for Riemann integrals. The basic idea is to approximate a nonegative function f with a sequence (ϕ_n) of nonnegative *simple* functions increasing to f. And to approximate the integral of f with the limit of integrals of these simple functions.

We first show that any measurable nonnegative function can be approximated with an increasing sequence of simple functions. To define ϕ_n, we divide the range $[0, \infty]$ of f into 2^{2n} intervals of size $1/2^n$, starting with $[0, 1/2^n)$, and then to take a final interval $[2^n, \infty]$. ϕ_n takes 2^{2n} values: for each interval $[k/2^n, (k+1)/2^n)$, this corresponds to the lowest value of f (i.e. $k/2^n$) in this interval of the range. The set of points in the domain on which ϕ_n takes this value is the set on which f takes values in $[k/2^n, (k+1)/2^n)$.

Theorem 1 *Let $(S, \mathcal{S})$ be a measurable space and $f : S \to [0, \infty]$ a measurable function. Then there is a sequence $(\phi_n)_1^\infty$ of simple functions such that (pointwise) $0 \le \phi_1 \le \phi_2 \le ... \le f$, ϕ_n converges to f pointwise, and ϕ_n converges to f uniformly (i.e. in the sup norm) on any set on which f is bounded.*

Proof. For each $n = 1, 2, ...,$ define ϕ_n as follows. For integers $k \in \{0, 1, ..., 2^{2n} - 1\}$, let $A_n^k = f^{-1}([k/2^n, (k+1)/2^n))$. Let $B_n = f^{-1}([2^n, \infty])$. Let

$$\phi_n = \sum_{k=0}^{2^{2n}-1} (k/2^n) I_{A_n^k} + 2^n I_{B_n}$$

That is, whenever $f(x)$ takes values in $[k/2^n, (k+1)/2^n)$, $\phi_n(x)$ takes the lowest, or rather infimum of these values. Similarly for the interval $[2^n, \infty]$. Thus $\phi_n \leq f$. It's also easy to see that $\phi_n \leq \phi_{n+1}$. Finally, on the set of x s.t. $f(x) \leq 2^n$, $0 \leq f - \phi_n \leq 1/2^n$. (The reason is that when f takes values in $[0, 2^n]$, for every subinterval of type $[k/2^n, (k+1)/2^n)$ in it, ϕ_n is constructed to take the value $k/2^n$, which is at most $1/2^n$ different from f in this subinterval). The convergence results follow. ■

Flesh out the diagram drawn in class for the above proof. $k \in \{0, 1, ..., 2^{2n} - 1\}$ because in the codomain, we partition $[0, 2^n]$ in intervals of length $1/2^n$, so there are 2^{2n} such intervals.

In what follows, $(S, \mathcal{S}, \mu)$ is the measure space, $M(S, \mathcal{S})$ the space of measurable, extended real-valued functions on S, and $M^+(S, \mathcal{S})$ the subset of nonnegative functions.

Definition 6 *Let $\phi \in M^+(S, \mathcal{S})$ be a measurable simple function, with the standard representation $\phi(x) = \sum_1^n a_i I_{A_i}(x)$. Then the integral of ϕ w.r.t. μ, written $\int_S \phi(x)\mu(dx)$, or $\int \phi d\mu$, is given by $\sum_1^n a_i \mu(A_i)$.*

Notice that if $\mu = \lambda$, i.e. Lebesgue measure, then the integral of a simple function is just a sum of rectangle areas: indeed, a simple function often looks like steps going up and coming down.

Definition 7 *For any nonnegative function $f \in M^+(S, \mathcal{S})$, the integral of f w.r.t. μ is*

$$\int f d\mu = \sup\{\int \phi d\mu : \phi \in M^+(S, \mathcal{S}), \text{ and } 0 \le \phi \le f\}$$

If $A \in \mathcal{S}$, the integral of f over A w.r.t. μ is

$$\int_A f d\mu = \int_S f(x) I_A(x) \mu(dx)$$

The monotone convergence theorem will show that the supremum over all $0 \le \phi \le f$ can be interpreted as using a monotone sequence $0 \le \phi_1 \le \phi_2 \le ... \le f$. This illustrates the meaning of integrating f by taking increasingly finer grids of a_i's on the vertical axis.

Claim 2 *Linearity of the Integral for simple functions: If ϕ, ψ are nonnegative simple functions and $c \ge 0$, then*

$$\int (\phi + \psi) d\mu = \int \phi d\mu + \int \psi d\mu, \text{ and } \int c\phi d\mu = c \int \phi d\mu$$

Proof. (Sketch). Let $\phi(x) = \sum_1^n a_i I_{A_i}(x)$, $\psi(x) = \sum_1^m b_j I_{B_j}(x)$ be standard representations of ϕ and ψ. Notice that $A_i = \cup_{j=1}^m (A_i \cap B_j)$, a disjoint union. Similarly, $B_j = \cup_{i=1}^n (A_i \cap B_j)$. So, $\int \phi d\mu = \sum_{i=1}^n a_i \sum_{j=1}^m \mu(A_i \cap B_j)$, by finite additivity of μ. Similarly, $\int \psi d\mu = \sum_{j=1}^m b_j \sum_{i=1}^n \mu(A_i \cap B_j)$. So

$$\int \phi d\mu + \int \psi d\mu = \sum_{i,j} (a_i + b_j) \mu(A_i \cap B_j)$$

Similarly, the right hand equals $\int (\phi + \psi) d\mu$. ■

The next lemma shows that the map $A \mapsto \int_A \phi d\mu$ is a measure. Notice that the function $\phi I_A = \sum_i a_i I_{A \cap A_i}$, if the standard representation of the function $\phi = \sum_i a_i I_{A_i}$. This is because $\phi(x) I_A(x)$ takes a value of a_i if and only if $x \in A_i$ and $x \in A$. Therefore, $\int_A \phi d\mu \equiv \int \phi I_A d\mu = \sum_i a_i \mu(A \cap A_i)$.

Lemma 1 *Let $\phi \in M^+(S, \mathcal{S})$ be a simple function and let $\lambda : \mathcal{S} \to \Re$ be defined by $\lambda(A) = \int_A \phi d\mu$, for all $A \in \mathcal{S}$. Then λ is a measure on $\mathcal{S}$.*

Proof. Note first that if $\phi(x) = \sum_i a_i I_{A_i}(x)$ in standard representation, then $\phi(x) I_A(x) = (\sum_i a_i I_{A_i}(x)) I_A(x) = \sum_i a_i I_{(A \cap A_i)}(x)$. So, $\int_A \phi d\mu = \int \phi I_A d\mu = \sum_i a_i \mu(A \cap A_i)$.

Now, let $(B_j)_1^\infty$ be a disjoint sequence of sets whose union is A. To show countable additivity of λ, note that

$$\int_A \phi d\mu = \sum_i a_i \mu(A \cap A_i) = \sum_i a_i \sum_j \mu(B_j \cap A_i) = \sum_j \int_{B_j} \phi d\mu$$

where the second last equality follows from countable additivity of μ. ■

Note. If f, g are nonnegative measurable functions and $f \leq g$, then $\int f d\mu \leq \int g d\mu$. The reason is that the set of simple functions between 0 and f is a subset of the set of simple functions between 0 and g; so the supremum is lower in the case of f. A corollary: If f is nonnegative and measurable, and $A \subset B$ are measurable sets, then $\int_A f d\mu \leq \int_B f d\mu$. Indeed, $f I_A \leq f I_B$.

The Monotone convergence theorem below shows that to integrate f w.r.t. μ, it suffices to consider the integrals of a monotone sequence of simple functions converging to f.

Theorem 3 *(Monotone Convergence Theorem). If (f_n) is a monotone increasing sequence of functions in $M^+(S, \mathcal{S})$ converging pointwise to f, then*

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

Proof. Since $\int f_n d\mu$ is an increasing sequence of reals, its limit (possibly ∞ exists). Also, $\int f_n d\mu \leq \int f d\mu, \forall n$. So, $\lim \int f_n d\mu \leq \int f d\mu$. For the converse, fix $\alpha \in (0,1)$ and $0 \leq \phi \leq f$. Let $A_n = \{x \in S : f_n(x) \geq \alpha\phi(x)\}$. Since $f_{n+1}(x) \geq f_n(x)$, $A_n \subset A_{n+1}$. Also, for every x, there must be some n for which $f_n(x) \geq \alpha\phi(x)$, since the pointwise limit $f(x) > \alpha\phi(x)$. So, $(A_n)_1^\infty$ is an increasing sequence of sets $A_n \subset A_{n+1}$ whose union is S. So $\int f_n d\mu \geq \int_{A_n} f_n d\mu \geq \alpha \int_{A_n} \phi d\mu$. Now we take limits on both sides as $n \to \infty$

(and so $A_n \uparrow S$). Since $A_n \mapsto \int_{A_n} \phi d\mu$ defines a measure (say $\nu(A_n)$), by mcfb we have $\lim \nu(A_n) = \nu(S)$, i.e. $\lim \int_{A_n} \phi d\mu = \int \phi d\mu$.

So, $\lim \int f_n d\mu \geq \alpha \int \phi d\mu$. Since this holds for arbitrary $\alpha \in (0,1)$, it holds for $\alpha = 1$. And so, the inequality also holds for the supremum on the RHS: taking the supremum over all $0 \leq \phi \leq f$, $\lim \int f_n d\mu \geq \int f d\mu$. ■

Example 3 *Let $S = (0,1]$, $\mathcal{S}$ the Borel σ-algebra on it, let $f : S \to \Re$ be defined by $f(x) = \frac{1}{\sqrt{x}}$. Let λ be lebesgue measure. (Verify that f is measurable). We want to find $\int_S f d\lambda$ (or $\int f$ for short).*

If we follow usual procedure for Riemann integration, this will integrate to 2. But notice that the function does not have a Riemann integral on S. Indeed, let $a_0 = 0 < a_1 < ... < a_n = 1$ be the endpoints of any partition of S. $\max\{f(x)|a_0 \leq x < a_1\}$ does not exist, so we can't assess an area to this rectangle; so the upper Riemann sum does not exist - neither does the Riemann integral.

We use the limiting procedure in the MCT. For $n = 1, 2, ...$ define $f_n = fI_{[1/n,1]}$. So, $0 \leq f_1 \leq f_2 \leq ... \leq f_n \leq ... \leq f$. (To show this, show that 0 (the zero function) is less than or equal to f_n (point by point), $f_n \leq f$ and $f_n \leq f_{n+1}$, all point by point for all x).

Since (f_n) is a sequence of measurable functions increasing to the limit function f, by the MCT, the integral of this limiting function, $\int f$, equals the limit of the sequence of integrals $\int f_n$. Notice that for every n, the Riemann integral $\int_0^1 f_n$ exists and must therefore equal the Lebesgue integral.

$$\int_0^1 f_n = \left[\frac{x^{(-1/2)}+1}{-(1/2)+1}\right]_{(1/n)}^{1} = 2\left[1 - \frac{1}{\sqrt{n}}\right]$$

The limit as $n \to \infty$ equals 2. ■

An implication of the monotone convergence theorem (MCT) is that the integral is *additive for all measurable nonnegative functions*, not just the simple functions.

Theorem 4 *Let (f_n) be a countable sequence of functions in $M^+(S, \mathcal{S})$, and let $f = \sum_n f_n$. Then $\int f d\mu = \sum_n \int f_n d\mu$.*

Proof. First take the case of 2 functions, say f_1 and f_2. By Theorem 1, there are increasing sequences (ϕ_j) and (ψ_j) of nonnegative simple functions that converge to f_1 and f_2 respectively. So, the increasing sequence $(\phi_j + \psi_j)$ converges to $(f_1 + f_2)$, so by the monotone convergence theorem,

$$\int (f_1 + f_2) d\mu = \lim \int (\phi_j + \psi_j) d\mu = \lim \int \phi_j d\mu + \lim \int \psi_j d\mu$$

The last inequality is due to linearity for simple functions; and the last expression equals $\int f_1 d\mu + \int f_2 d\mu$.

Next, by induction, additivity holds for a finite sequence $(f_n)_1^N$. So $\int (\sum_1^N f_n)\, d\mu = \sum_1^N (\int f_n d\mu)$. Now the partial sums $(\sum_1^N f_n)_N$ form a sequence of increasing functions with limit $\sum_1^\infty f_n$, so applying the monotone convergence theorem to them, the integral of this limit (the limit $\lim_{N\to\infty} \sum_1^N f_n \equiv \sum_1^\infty f_n$) equals the limit of the integrals. The limit of the integrals is

$$\lim_{N\to\infty} \int \sum_1^N f_n = \lim_{N\to\infty} \sum_1^N \int f_n \equiv \sum_1^\infty \int f_n$$

.

So,

$$\int \sum_1^\infty f_n = \sum_1^\infty \int f_n$$

■

We can refine the hypothesis of a sequence of functions increasing to f in the monotone convergence theorem, to them increasing to f almost everywhere (a.e.), without changing the conclusion of the integral of f being the limit of the integrals of the f_n's. To see this, we first have the following lemma. To shorten notation, we write $\int f$ for $\int f d\mu$, as the measure μ is understood. The lemma applies for instance to the Dirichlet function discussed earlier (the indicator function on the rationals), but is much more general.

Lemma 2 *If f is a nonnegative measurable function, then $\int f = 0$ if and only if $f = 0$ a.e.*

Proof. Suppose f is simple, and so let $f = \sum_1^n a_i I_{A_i}$ be the standard representation. $\int f = \sum_1^n a_i \mu(A_i) = 0$ if and only if, for every i, either $a_i = 0$ or $\mu(A_i) = 0$. For the set on which $f(x) > 0$, the relevant a_i's are greater than 0, so the $\mu(A_i)$'s are equal to 0: the set comprising the union of these has measure zero. Now take the general case. If $f = 0$ a.e., and ϕ is a simple function with $0 \leq \phi \leq f$, then $\phi = 0$ a.e. So $\int f = \sup_{0 \leq \phi \leq f} \int \phi = 0$. Conversely, suppose $f = 0$ a.e. is not true. Then, letting $E_n = \{x : f(x) > 1/n\}$, $\mu(E_n) > 0$ for some n. (Because $\{x : f(x) > 0\} = \cup_1^\infty E_n$, and this set has positive measure). But then, $f > I_{E_n}/n$; integrating both sides, $\int f \geq \mu(E_n)/n > 0$. ∎

The monotone convergence theorem under weaker hypothesis now follows.

Corollary 2 *If (f_n) is a sequence of measurable nonnegative functions that increase to $f \in M^+(S, \mathcal{S})$ almost everywhere, then $\int f = \lim \int f_n$.*

Proof. Suppose $(f_n(x))$ increases to $f(x)$ for all $x \in E$, and $\mu(E^c) = 0$. Then, $f = f I_E$ a.e., and for all n, $f_n = f_n I_E$ a.e. Moreover, since $f = f I_E + f I_{E^c}$ and $f I_{E^c} = 0$ a.e., (and a similar thing holds for the f_n)'s, by linearity of the integral and the above lemma we get the first and the last equality below.

$$\int f = \int f I_E = \lim \int f_n I_E = \lim \int f_n$$

The middle equality follows from the monotone convergence theorem, since $(f_n I_E)$ is a sequence of functions increasing to fI_E everywhere. ∎

Note. The monotone convergence theorem (MCT) says that the integral of the limit function f is the limit of the integrals $\int f_n d\mu$. The assumption that the (f_n)'s form an increasing sequence is required. Indeed, consider $S = \Re$, $\mu =$ Lebesgue measure, and the sequence of functions $(I_{(n,n+1)})$. As $n \to \infty$, $I_{(n,n+1)} \to 0$, pointwise. So the zero function is the limit function, and its integral equals zero. However, the integral $\int I_{(n,n+1)} d\mu = 1$, for all n, (since μ is Lebesgue measure). So, $\lim \int I_{(n,n+1)} d\mu = 1$. Thus although the functions converge to the zero function, the integral of the limit function (the zero function) is not equal to the limit of the intergrals.

A weaker version than the MCT, one that holds for *all* sequences of measurable nonnegative functions, is roughly that the integral of the limit function is less than or equal to the limit of the integrals. This is Fatou's Lemma. We suppress the measure μ w.r.t. which integrals are computed, for cleaner notation.

Note that the notion of limit or convergence used in these basic theorems is that of pointwise limit. In the MCT, the sequence (f_n) of functions increases to f for every (or almost every) point x. So the sequence $(f_n(x))$ converges to $f(x)$. Fatou's lemma on the other hand has a sequence (f_n) that does not necessarily increase or decrease. Thus for a given x, the sequence $(f_n(x))$ need not converge anywhere; so we can't talk of a pointwise limit function. But the $(f_n(x))$ must have a limit point (if it's bounded, then this follows from the Bolzano-Weierstrass' Theorem; otherwise, the limit point is at $\pm\infty$). So a liminf exists.

Lemma 3 *(Fatou's Lemma). If (f_n) is a sequence of measurable nonnegative* *functions, then*

$$\int (\liminf f_n) \leq \liminf \int f_n$$

Proof. For each k, $\inf_{n \geq k} f_n \leq f_j$, for all $j \geq k$. So, $\int \inf_{n\geq k} f_n \leq \int f_j$, $j \geq k$. So, $\int \inf_{n\geq k} f_n \leq \inf_{j\geq k} \int f_j$. Now we take limits as $k \to \infty$. But the $(\inf_{n\geq k} f_n)$'s are an increasing sequence of functions (indexed by $k = 1, 2, ...$), so applying the monotone convergence theorem,

$$\int (\liminf f_n) = \lim_{k\to\infty} \int (\inf_{n\geq k} f_n) \leq \liminf \int f_n$$

∎

As a corollary, we can now consider a sequence of functions converging pointwise to a limit function a.e.

Corollary 3 *Let (f_n) be a sequence of measurable nonnegative functions converging pointwise to $f \in M^+(S, \mathcal{S})$ a.e. Then $\int f \leq \liminf \int f_n$.*

Proof. If $f_n \to f$ everywhere, then $\liminf f_n = \lim f_n = f$, so the conclusion follows from Fatou's lemma. For $f_n \to f$ a.e., we modify f_n and f on a set of measure zero s.t. f_n converges to f everywhere. By Lemma 2, the integrals are unchanged and we may apply Fatou's lemma. ∎

Integration of Functions in $\mathbf{M}(\mathbf{S}, \mathcal{S})$

Finally, we extend integration of non-negative measurable functions to functions taking values in the entire $\Re$. By the same convention followed in Riemann integration, in regions where f takes negative values, the integral is negative. Specifically, we exploit the fact that if f^+ and f^- are the positive and negative parts of f, then $f = f^+ - f^-$, and since both f^+ and f^- are non-negative, we can integrate them with the theory above. (Recall that $f^+(x) = \max\{f(x), 0\}$ and $f^-(x) = -\min\{f(x), 0\}$. So $f = f^+ - f^-$, and $|f| = f^+ + f^-$, where all these equations are pointwise equalities). We have:

Definition 8

$$\int f = \int f^+ - \int f^-$$

If $\int f^+$ and $\int f^-$ are both finite, we say f is *integrable*. Note that $f^+ \leq |f|$ and $f^- \leq |f|$ (pointwise). So, if $\int |f| < \infty$, f is integrable. Conversely, suppose f is integrable, so both $\int f^+$ and $\int f^-$ are finite. So their sum is also finite. Now $|f| = f^+ + f^-$ pointwise, so integration yields $\int |f| = \int f^+ + \int f^- < \infty$. We have

Proposition 5 *f is integrable if and only if $\int |f| d\mu < \infty$.*

We also have

Proposition 6 *Let $f \in M(S, \mathcal{S})$ be integrable. Then $|\int f| \leq \int |f|$.*

Proof.

$$|\int f| = |\int f^+ - \int f^-| \leq \int f^+ + \int f^- = \int |f|$$

■

We note the following fact.

Proposition 7 *The set of all integrable, real-valued functions is a real vector space, and the integral is a linear functional.*

Proof. Linear combinations of integrable functions are integrable, since $|af + bg| \leq |a||f| + |b||g|$.

To show that the integral is linear, first note that $\int cf = c \int f, c \in \Re$ (this is straightforward). Finally, suppose $h = f + g$, where f and g are integrable. So, $h^+ - h^- = f^+ - f^- + g^+ - g^-$. Rearranging to get addition, $h^+ + f^- + g^- = h^- + f^+ + g^+$. By linearity of the integral for nonnegative functions, it follows that $\int h^+ + \int f^- + \int g^- = \int h^- + \int f^+ + \int g^+$. Rearranging back gives $\int h = \int f + \int g$. ■

Now for the other major convergence theorem.

Theorem 5 *(Dominated Convergence Theorem). Let (f_n) be a sequence of integrable functions in $M(S, \mathcal{S})$ s.t. $f_n \to f$ a.e. Suppose there exists an integrable nonnegative function g that dominates the f_n's, i.e. $|f_n| \leq g, \forall n$. Then f is integrable and $\int f = \lim_{n\to\infty} \int f_n$.*

Proof. From Proposition 4, we know convergence of measurable f_n to f implies f is measurable (maybe after redefining it on a set of measure 0, due to the a.e. in the hypothesis, to have convergence everywhere: this redefinition won't change the value of the integral). Also, $|f| \leq g$ (a.e.), so f is integrable. Now, by the dominance of g, we have both $g + f_n \geq 0$, and $g - f_n \geq 0$, a.e., i.e. these are nonnegative functions (to which we can apply Fatou's lemma). Now, $g \pm f_n \to g \pm f$. Applying the corollary to Fatou's lemma, (the first inequalities below)

$$\int g + \int f \leq \liminf \int (g + f_n) = \int g + \liminf \int f_n$$

$$\int g - \int f \leq \liminf \int (g - f_n) = \int g - \limsup \int f_n$$

Cancelling $\int g$ and multiplying the 2nd line above by (-1), we have $\int f = \lim \int f_n$. ■