Appendix to Notes 8 (a)

13 Comparison of the Riemann and Lebesgue integrals.

Recall Let $f : [a,b] \to \mathbb{R}$ be bounded. Let D be a partition of $[a,b]$ such that

$$D = \{a = x_0 < x_1 < ... < x_n = b\}.$$

Let

$$\begin{aligned} m_i &= \inf\{f(x) : x_{i-1} \le x \le x_i\} \\ M_i &= \sup\{f(x) : x_{i-1} \le x \le x_i\}. \end{aligned}$$

Define the step functions (therefore, simple functions, since we have assumed that f is bounded and so $M_i < \infty$ for all i).

$$\alpha_D(x) = m_i \text{ on } [x_{i-1}, x_i) \text{ for all } 1 \le i \le n,$$

and

$$\beta_D(x) = M_i \text{ on } [x_{i-1}, x_i) \text{ for all } 1 \le i \le n.$$

So

$$\alpha_D(x) \le f(x) \le \beta_D(x) \quad \text{for all } x \in [a,b].$$

Note that if $D \supseteq D'$ then $\alpha_{D'}(x) \le \alpha_D(x)$ and $\beta_D(x) \le \beta_{D'}(x)$. That is, with a finer partition we get better approximations to f. With the notation of integrals of simple functions we have, with Lebesgue measure on $\mathbb{R}$,

$$I(\alpha_D) = \sum_{i=1}^n m_i(x_i - x_{i-1}) \qquad \text{and} \qquad I(\beta_D) = \sum_{i=1}^n M_i(x_i - x_{i-1}),$$

which are normally known as $L(D,f)$ and $U(D,f)$ in the theory of Riemann integration. Then we obviously have $I(\alpha_D) \le I(\beta_D)$ for all D, and if $D \supseteq D'$ then $I(\alpha_{D'}) \le I(\alpha_D)$ and $I(\beta_D) \le I(\beta_{D'})$. Let

$$\underline{\int_a^b} f(x)dx = \sup_D I(\alpha_D) \qquad \text{and} \qquad \overline{\int_a^b} f(x)dx = \inf_D I(\beta_D).$$

Then f is Riemann integrable if, and only if,

$$\underline{\int_a^b} f(x)dx = \overline{\int_a^b} f(x)dx.$$

The common value is denoted by

$$R\text{-}\int_a^b f(x)dx.$$

Theorem 1

If f is Riemann integrable on a finite interval $[a,b]$ then it is Lebesgue integrable with the same value.

Proof For each $n \geq 1$ we can find, by the definition of supremum, a partition D_n^α such that

$$0 \leq \underline{\int_a^b} f(x)dx - I(\alpha_{D_n^\alpha}) < \frac{1}{n},$$

when, in particular,

$$I(\alpha_{D_n^\alpha}) \to \underline{\int_a^b} f(x)dx \quad \text{as } n \to \infty.$$

Similarly choose a sequence of partitions D_n^β such that

$$I(\beta_{D_n^\beta}) \to \overline{\int_a^b} f(x)dx \quad \text{as } n \to \infty.$$

Set $D_n = D_n^\alpha \cup D_n^\beta$ then

$$I(\alpha_{D_n^\alpha}) \leq I(\alpha_{D_n}) \leq \underline{\int_a^b} f(x)dx$$

and

$$I(\beta_{D_n^\beta}) \geq I(\beta_{D_n}) \geq \overline{\int_a^b} f(x)dx.$$

Thus

$$I(\alpha_{D_n}) \to \underline{\int_a^b} f(x)dx \quad \text{and} \quad I(\beta_{D_n}) \to \overline{\int_a^b} f(x)dx \quad \text{as } n \to \infty. \tag{1}$$

Replacing the sequence $D_1, D_2, D_3, ...$ by $D_1, D_1 \cup D_2,\ D_1 \cup D_2 \cup D_3, ...$and relabeling we can assume that $D_n \subseteq D_{n+1}$ for all $n \geq 1$ while (1) still holds. Yet $D_n \subseteq D_{n+1}$ means that

$$\alpha_{D_n}(x) \leq \alpha_{D_{n+1}}(x) \quad \text{and} \quad \beta_{D_n}(x) \geq \beta_{D_{n+1}}(x) \text{ for all } n \text{ and } x.$$

In particular $\{\alpha_{D_n}\}_{n\geq 1}$ in an increasing sequence bounded above by f. So $\lim_{n\to\infty} \alpha_{D_n} = g$ exists, and satisfies $g \leq f$. Similarly $\{\beta_{D_n}\}_{n\geq 1}$ in an decreasing sequence bounded below by f. So $\lim_{n\to\infty} \beta_{D_n} = h$ exists, and satisfies $h \geq f$.

Now $\{\alpha_{D_n} - \alpha_{D_1}\}_{n\geq 1}$ is an increasing sequence of non-negative simple $\mathcal{F}$-measurable functions tending to $g - \alpha_{D_1}$. So by Lebesgue's Monotone Convergence Theorem we have

$$\begin{aligned} L\text{-}\int_a^b (g - \alpha_{D_1})d\mu &= \lim I(\alpha_{D_n} - \alpha_{D_1}) \\ &= \lim I(\alpha_{D_n}) - I(\alpha_{D_1}) \\ &= \underline{\int_a^b} f(x)dx - I(\alpha_{D_1}). \end{aligned}$$

Since α_{D_1} is a simple function we have $L\text{-}\int_a^b \alpha_{D_1} d\mu = I(\alpha_{D_1})$ and so

$$L\text{-}\int_a^b g d\mu = \underline{\int_a^b} f(x)dx. \tag{2}$$

Similarly, by examining $\beta_{D_1} - \beta_{D_n}$ we find that

$$L\text{-}\int_a^b h d\mu = \overline{\int_a^b} f(x)dx.$$

So, if f is Riemann integrable, that is, $\underline{\int_a^b} f(x)dx = \overline{\int_a^b} f(x)dx$, then $L \int_a^b (g - h)d\mu = 0$. Yet $h - g \geq 0$, so $h = g$ a.e.(μ) on $[a, b]$. But $g \leq f \leq h$ and so $f = g$ a.e.(μ) on $[a, b]$. Hence

$$\begin{aligned} L\text{-}\int_a^b f d\mu &= L\text{-}\int_a^b g d\mu && \text{since } f = g \text{ a.e.}(\mu) \text{ on } [a, b] \\ &= \underline{\int_a^b} f(x)dx && \text{by (2)} \\ &= R\text{-}\int_a^b f(x)dx && \text{since } f \text{ is Riemann integrable} \end{aligned}$$

■

Let $\Delta(D) = \max_{1 \le i \le n}(x_i - x_{i-1})$. In Theorem 1 it is possible, by adding extra points to each of the partitions D_n, to assume that $\Delta(D_n) \to 0$ as $n \to \infty$. With the notation and assumptions of Theorem 1 we can prove

Lemma 1

Assume that $\Delta(D_n) \to 0$ as $n \to \infty$. Then for any $x \notin \bigcup_{k=1}^{\infty} D_k$ we have that f is continuous at x if, and only if, $g(x) = f(x) = h(x)$.

Proof Recall that f is continuous at x if, and only if,

$$\forall \varepsilon > 0 \ \exists \delta > 0 : \forall y \text{ if } |y - x| < \delta \text{ then } |f(y) - f(x)| < \varepsilon. \tag{3}$$

For each k let I_k be the subinterval of D_k containing x. This is unique since $x \notin \bigcup_{k=1}^{\infty} D_k$. Write $I_k = [x_{i-1}, x_i]$.

($\Rightarrow$) Let $\varepsilon > 0$ be given. From (D6) we find a $\delta > 0$. Since $\Delta(D_n) \to 0$ as $n \to \infty$ there exists N such $\Delta(D_n) < \delta$ for all $n \ge N$. Then $\ell(I_k) < \delta$ so if $y \in I_k$ we have that $|y - x| < \delta$. In which case, from (D6) we get that $|f(y) - f(x)| < \varepsilon$. In turn this means that we have both

$$\left| \inf_{I_k} f(y) - f(x) \right| < \varepsilon \quad \text{and} \quad \left| \sup_{I_k} f(y) - f(x) \right| < \varepsilon.$$

Yet $\inf_{I_k} f(y)$ and $\sup_{I_k} f(y)$ are the values of α_k and β_k at x. Hence, combining the inequalities, $|\beta_k(x) - \alpha_k(x)| < 2\varepsilon$. Let $k \to \infty$ to deduce $|h(x) - g(x)| < 2\varepsilon$. True for all $\varepsilon > 0$ gives $h(x) = g(x)$.

($\Leftarrow$) Assume f is not continuous at x. So

$$\exists \varepsilon > 0 \ \forall \delta > 0 : \exists y \text{ with } |y - x| < \delta \text{ and } |f(y) - f(x)| \ge \varepsilon. \tag{4}$$

For each $k \ge 1$ choose $\delta_k = \min(x - x_{i-1}, x_i - x)$ so $(x - \delta_k, x + \delta_k) \subseteq I_k$. But then by (4) we can find $y_k \in (x - \delta_k, x + \delta_k)$ such that $|f(y_k) - f(x)| \ge \varepsilon$. In particular,

$$\sup_{I_k} f - \inf_{I_k} f \ge \varepsilon,$$

in which case $\beta_k(x) - \alpha_k(x) \ge \varepsilon$ and $h(x) - g(x) \ge \varepsilon$. Hence $h(x) \ne g(x)$.■

This leads to

Theorem 2

Assume $f : [a, b] \to \mathbb{R}$ is bounded. Then f is Riemann integrable if, and only if, f is continuous a.e.(μ) *on $[a, b]$.*

Proof Choose a sequence of partitions, D_k, as in Lemma 1. Then.

$$
\begin{aligned}
& f \text{ is continuous a.e.}(\mu) \text{ on } [a,b] \\
\text{iff} \quad & f \text{ is continuous a.e.}(\mu) \text{ outside } \bigcup_k D_k \text{ on } [a,b] \\
\text{iff} \quad & g = h \text{ a.e.}(\mu) \text{ on } [a,b] \qquad \text{by Lemma 1,} \\
\text{iff} \quad & \int g d\mu = \int h d\mu \\
\text{iff} \quad & \underline{\int_a^b} f(x)dx = \overline{\int_a^b} f(x)dx \\
\text{iff} \quad & f \text{ is Riemann integrable.}
\end{aligned}
$$

■

Measure preserving Transformations

These are a special case of measurable functions.

Definition $T : (X, \mathcal{F}, \mu) \to (X, \mathcal{F}, \mu)$ is a *measure preserving transformation* if

(i) $T^{-1}A \in \mathcal{F}$ for all $A \in \mathcal{F}$,
(ii) $\mu(T^{-1}A) = \mu(A)$ for all $A \in \mathcal{F}$.

Definition Let $A \in \mathcal{F}$. A point $x \in A$ is said to be *recurrent with respect to* A if there exists a $k \geq 1$ such that $T^k x \in A$.

Theorem 3 Poincare's Recurrence Theorem

Assume that $\mu(X) < \infty$. *Let* F *be the set of points of* A *which are not recurrent with respect to* A. *Then* $\mu(F) = 0$.

(So for every $A \in \mathcal{F}$, almost all points of A are recurrent.)

Proof

Let $x \in F$. If there exists $n \geq 1$ such that $T^n x \in F$ then we have both $x \in F \subseteq A$ and $T^n x \in F \subseteq A$, i.e. x is a recurrent point with respect to A, which contradicts the definition of F. So $T^n x \notin F$ for all $n \geq 1$, that is, $T^n F \cap F = \emptyset$ for all $n \geq 1$. Now, the preimage of an empty set is empty, so given any $k, n \geq 1$ we have

$$\emptyset = T^{-k-n}(T^n F \cap F) = T^{-k} F \cap T^{-(n+k)} F.$$

Hence the sets $F, T^{-1}F, T^{-2}F, \ldots$ are pairwise disjoint. So

$$\begin{aligned}
\infty &> \mu(X) \geq \mu\left(\bigcup_{k\geq 0} T^{-k}F\right) \\
&= \sum_{k=0}^{\infty} \mu(T^{-k}F) \\
&= \sum_{k=0}^{\infty} \mu(F) \qquad \text{by part (ii) of definition.}
\end{aligned}$$

Hence $\mu(F) = 0$. ■

We can ask how long it takes a point $x \in A$ to wander back into A. To this end define

$$n_A(x) = \min\{n \geq 1 : T^n x \in A\}.$$

Assume throughout the rest of this section that $\mu(X) = 1$.

Definition A measure preserving map $T : X \to X$ is *Ergodic* if either of the following hold.

(i) Whenever $A \in \mathcal{F}$ is such that $\mu(T^{-1}A \bigtriangleup A) = 0$ then either $\mu(A) = 0$ or 1.

(ii) Whenever an integrable function f satisfies $f(Tx) = f(x)$ for a.e. (μ) x in X then f is constant a.e.(μ) on X.

The first definition here means that if $T^{-1}A$ is almost exactly A then either $\mu(A) = 0$ or 1. So if $0 < \mu(A) < 1$ then $T^{-1}A$ must differ "quite a lot" from A. We say that T is *mixing up* the space.

We do not prove here that (i) and (ii) are equivalent.

It can be shown that if T is ergodic then

$$\int_A n_A d\mu = 1.$$

Since $\int_A d\mu = \mu(A)$ we have, that in some sense, n_A is of size $1/\mu(A)$. This is connected with the question of how often a point $x \in X$ will wander into the set $A \in \mathcal{F}$. It will be shown below that for $B \in \mathcal{F}$ and T ergodic,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_B(T^k x) = \mu(B) \text{ a.e. } (\mu), \tag{5}$$

where χ_B is the characteristic function of the set B, i.e. $\chi_B(x) = 1$ if $x \in B$, 0 otherwise. So if $\mu(B) > 0$ then almost every point of X wanders into B infinitely often.

Let $S_n(x) = \#\{1 \leq i \leq n : T^i x \in B\}$ and $A_n(x) = S_n(x)/n$. It is not obvious that the limit $lim_{n\to\infty} A_n(x)$ will exist. We will show that it does by looking at the limsup and liminf of the sequence $\{A_n(x)\}$. So let $\overline{A}(x) = \limsup A_n(x)$ which is trivially ≤ 1.

Lemma 2 *Let* $\{a_n\}$ *be a sequence for which* $\limsup a_n < \infty$. *Let* $\{b_n\}$ *be a sequence for which* $\lim b_n = 0$. *Then* $\limsup(a_n + b_n) = \limsup a_n$.

Proof Write $A = \limsup a_n$.

Let $\varepsilon > 0$ be given. There exists N_1 such that $-\varepsilon < b_n < \varepsilon$ for all $n \geq N_1$ and there exists N_2 such that

$$A - \varepsilon < \sup_{r \geq n} a_r < A + \varepsilon$$

for all $n \geq N_2$. Choose $N = \max(N_1, N_2)$, so that for all $n \geq N$ we have

$$A - 2\varepsilon < \sup_{r \geq n}(a_r + b_r) < A + 2\varepsilon$$

which gives the result. ■

Lemma 3

$$\overline{A}(Tx) = \overline{A}(x).$$

Proof

$$\begin{aligned} A_n(Tx) &= \frac{1}{n} \sum_{1 \leq i \leq n} \chi_B(T^i(Tx)) = \frac{1}{n} \sum_{2 \leq i \leq n+1} \chi_B(T^i x) \\ &= A_n(x) + \frac{\chi_B(x) - \chi_B(T^{n+1}x)}{n} \end{aligned}$$

and an application of Lemma 2 gives the result. ■

Theorem 4

$$\textit{The limit } \lim_{n\to\infty} A_n(x) \textit{ exists.}$$

Proof

For a given $x \in X$ we follow the *orbit of* x, namely $x, Tx, T^2x, T^3x, \ldots$. We call the exponent n in $T^n x$, the *time*.

Let $\varepsilon > 0$ be given.

It might be that for all sufficiently large n we have $A_n(x) > \overline{A}(x) - \varepsilon$ which obviously shows that the limit exists. Otherwise the sequence $\{m_j\}$ defined by

$$m_j = \min\{m > m_{j-1} : A_m(x) > \overline{A}(x) - \varepsilon\}$$

has infinitely many gaps. Note that this sequence depends on x. The question must be how large can these gaps be?

Define

$$\tau(x) = \min\{n : A_n(x) > \overline{A}(x) - \varepsilon\}.$$

We first assume that there exists M such that $\tau(x) < M$ a.e. (μ). Let $\mathcal{S}$ be the exceptional set here.

Assume there is a gap after m_j so $A_{m_j}(x) > \overline{A}(x) - \varepsilon$ but $A_{m_j+1}(x) \leq \overline{A}(x) - \varepsilon$. Then if $T^{m_j}x \notin \mathcal{S}$ we know there exists $n < M$ such that $A_n(T^{m_j}x) > \overline{A}(T^{m_j}x) - \varepsilon = \overline{A}(x) - \varepsilon$ by the lemma above. So we have both

$$\sum_{1\leq i\leq m_j} \chi_B(T^i x) > m_j(\overline{A}(x) - \varepsilon)$$

and

$$\sum_{m_j+1\leq i\leq m_j+n} \chi_B(T^i x) = \sum_{1\leq i\leq n} \chi_B(T^i(T^{m_j}x)) > n(\overline{A}(x) - \varepsilon).$$

Adding these two inequalities gives

$$\sum_{1\leq i\leq m_j+n} \chi_B(T^i x) > (m_j + n)(\overline{A}(x) - \varepsilon),$$

that is

$$A_{m_j+n}(x) > \overline{A}(x) - \varepsilon.$$

Thus $m_{j+1} \leq m_j + n < m_j + M$. So if $T^{m_j}x \notin \mathcal{S}$, the gap $m_{j+1} - m_j$ is less than M. So if

$$x \notin \bigcup_{k=1}^{\infty} T^{-k}\mathcal{S}$$

a set of measure zero, all the gaps are less than M. Thus given $x \notin \bigcup_{k=1}^{\infty} T^{-k}\mathcal{S}$ and given N choose j (which will depend on x as well N since the sequence of m_j depends on x) such that $m_j \leq N < m_{j+1}$, then for almost all x we have

$$\begin{aligned} S_N(x) &\geq S_{m_j}(x) > m_j(\overline{A}(x) - \varepsilon) \\ &> (N - M)(\overline{A}(x) - \varepsilon). \end{aligned}$$

Since this inequality is true for almost all x we can integrate to get

$$\begin{aligned} \int_X (\overline{A}(x) - \varepsilon) d\mu &\leq \frac{1}{N-M} \sum_{n \leq N} \int_X \chi_{T^{-n}B} d\mu \\ &= \frac{1}{N-M} \sum_{n \leq N} \mu(T^{-n}B) \\ &= \frac{N\mu(B)}{N-M}, \end{aligned}$$

since T is measure preserving. Let $N \to \infty$ and then $\varepsilon \to 0$ to deduce

$$\int_X \overline{A}(x) d\mu \leq \mu(B). \tag{6}$$

The assumption above concerning M may not hold. Define $\mathcal{S}_M = \{x : \tau(x) > M\}$, the collection of which is a nested sequence of sets, $\mathcal{S}_1 \supseteq \mathcal{S}_2 \supseteq \mathcal{S}_3 \supseteq ....$ From the definition of $\limsup A_n(x)$ we know that given $\varepsilon > 0$ and $x \in X$ there exists (infinitely many) N such that $A_N(x) > \overline{A}(x) - \varepsilon$ in which case $\tau(x) \leq N$ and so $x \notin \mathcal{S}_N$. Let $\mathcal{S}^* = \bigcap_{M \geq 1} \mathcal{S}_M$, then we have seen that $x \notin \mathcal{S}^*$ for all $x \in X$, that is, $\mathcal{S}^* = \emptyset$. Since the $\mathcal{S}_N$ are nested we can thus find an M such that $\mu(\mathcal{S}_M) < \varepsilon$.

Let $B' = B \cup \mathcal{S}_M$. We apply the arguments above with B replaced by B', so $S'_n(x) = \#\{1 \leq i \leq n : T^i x \in B'\}$, $A'_n(x) = S'_n(x)/n$ and $A'(x) = \limsup A'_n(x)$. We follow the method above looking at the gaps in the sequence of $\{m_j\}$. If after some m_j we have a gap then $A'_{m_j}(x) > \overline{A}'(x) - \varepsilon$ but $A'_{m_j+1}(x) \leq \overline{A}'(x) - \varepsilon$, in particular $A'_{m_j+1}(x) < A'_{m_j}(x)$. If it were the case that $T^{m_j+1}x \in B'$ then

$$\begin{aligned} A'_{m_j+1}(x) &= \frac{S'_{m_j}(x) + \chi_{B'}(T^{m_j+1}x)}{m_j + 1} = \frac{S'_{m_j}(x) + 1}{m_j + 1} \\ &> \frac{S'_{m_j}(x)}{m_j} = A'_{m_j}(x), \end{aligned}$$

having used the observation that $D > C > 0$ implies $\frac{C+1}{D+1} > \frac{C}{D}$. Hence we must have $T^{m_j+1}x \notin B'$. In particular $T^{m_j+1}x \notin \mathcal{S}_M$, in which case

$\tau(T^{m_j+1}x) \leq M$ and so, as in the argument above, the gap after m_j is bounded by M. Thus following the argument that led to (6) will lead to

$$\int_X \overline{A}'(x)d\mu \leq \mu(B').$$

But $A'_n(x) \geq A_n(x)$ for all n in which case $\overline{A}'(x) \geq \overline{A}(x)$, while $\mu(B') < \mu(B) + \varepsilon$. Let $\varepsilon \to 0$ to deduce

$$\int_X \overline{A}(x)d\mu \leq \mu(B).$$

Hence this inequality holds whatever we assume about M.

A similar argument gives

$$\mu(B) \leq \int_X \underline{A}(x)d\mu$$

where $\underline{A}(x) = \liminf A_n(x)$. Hence

$$\int_X \underline{A}(x)d\mu = \int_X \overline{A}(x)d\mu = \mu(B) \tag{7}$$

and so $\underline{A}(x) = \overline{A}(x)$ for a.e. (μ) $x \in X$. Hence $\lim_{n\to\infty} A_n(x)$ exists for almost every (μ) $x \in X$ (and is integrable). ■

The remaining question must be waht is the value of this limit?

From the definition of ergodic above the appropriate one for the present situation states that whenever an integrable function f satisfies $f(Tx) = f(x)$ for a.e. (μ) x in X then f is constant a.e.(μ) on X.

From the lemma above we have that $\overline{A}(Tx) = \overline{A}(x)$ for all x and so for the points at which the limit exists we have $\lim_{n\to\infty} A_n(Tx) = \lim_{n\to\infty} A_n(x)$, i.e. this holds a.e.$(\mu)$ on X. Hence if T is ergodic we have that $\lim_{n\to\infty} A_n(x) = c$, a constant, a.e.$(\mu)$. For the value of c use (7) that shows that

$$\mu(B) = \int_X \lim_{n\to\infty} A_n(x)d\mu = \int_X cd\mu = c\mu(X) = c,$$

since $\mu(X) = 1$. Hence

$$\lim_{n\to\infty} \frac{\#\{1 \leq i \leq n : T^i x \in B\}}{n} = \mu(B)$$

a.e.(μ).