The Lebesgue Differentiation Theorem

In this short note, we discuss the *Lebesgue Differentiation Theorem*. Recall from elementary real analysis that, if $f \in C(\mathbb{R})$, then[1]

$$\lim_{h \to 0^+} \frac{1}{2h} \int_{x-h}^{x+h} f(y)\, dy = f(x).$$

In other words, the average value of f on $[x-h, x+h]$ converges to $f(x)$ as the length of the interval tends to zero. The Lebesgue Differentiation Theorem generalizes this result to locally integrable functions. Since these functions are defined a.e., we of course cannot have that the above limit holds for *every* $x \in \mathbb{R}$. However, we can be sure that it holds for almost every x.

There are several ways to state the Lebesgue Differentiation Theorem. We begin with the most familiar version:

Theorem 1 (Lebesgue Differentiaion I)**.** *If $f \in L^1_{loc}(\mathbb{R}^n)$ and $B_r(x)$ is the n-ball of radius r centered at x, then*

$$\lim_{r \to 0} \frac{1}{m(B_r(x))} \int_{B_r(x)} f(y)\, dy = f(x) \text{ for a.e. } x \in \mathbb{R}^n.$$

The proof of the Lebesgue Differentiation is a bit more delicate than the case when f is continuous, and relies on the *Maximal Theorem*, which quantifies the measure for which the *Hardy-Littlewood maximal function* given by

$$Hf(x) := \sup_{r>0} \frac{1}{m(B_r(x))} \int_{B_r(x)} |f(y)|\, dy \text{ for } f \in L^1_{\text{loc}}(\mathbb{R}^n)$$

is greater than some fixed number $\alpha > 0$. Precisely, the theorem says:

Theorem 2 (The Maximal Theorem)**.** *There is a constant $C > 0$ such that for all $f \in L^1$ and all $\alpha > 0$ we have*

$$m(\{x : Hf(x) > \alpha\}) \leq \frac{C}{\alpha} \int_{\mathbb{R}^n} |f(x)|\, dx.$$

With Theorem 1 in mind, it makes sense to consider the *Lebesgue set L_f of f*:

$$L_f := \Big\{ x : \lim_{r \to 0} \frac{1}{m(B_r(x))} \int_{B_r(x)} |f(y) - f(x)|\, dy = 0 \Big\}.$$

With just a little bit more work, one can prove a stronger version of the Lebesgue Differentiation Theorem which quantifies the measure of the Lebesgue set of f.

Theorem 3 (Lebesgue Differentiation II)**.** *If $f \in L^1_{loc}$, then $m((L_f)^c) = 0$.*

One can also consider families of sets more general than balls. We say a family $\{E_r\}_{r>0}$ of Borel subsets of $\mathbb{R}^n$ *shrinks nicely* to $x \in \mathbb{R}^n$ if

(i) $E_r \subset B_r(x)$ for each r and

(ii) There is a constant $\alpha > 0$, independent of r, such that $m(E_r) > \alpha m(B_r(x))$.

[1] If you haven't seen the proof, it is a good exercise to prove it. The proof is rather straightforward and relies on the continuity of f.

In plain language, a family $\{E_r\}_{r>0}$ of Borel sets shrinks nicely to $x \in \mathbb{R}^n$ if each set E_r in the family is contained in the ball $B_r(x)$ and the measure of each set shrinks proportionally to that of the ball $B_r(x)$ as $r \to 0$. We note that the sets E_r do not even need to contain x. For example, if U is any Borel subset of $B_1(0)$ such that $m(U) > 0$ and $E_r := \{x + ry : y \in U\}$, then $\{E_r\}$ shrinks nicely to x. This leads to the most general version of the Lebesgue Differentiation Theorem:

Theorem 4 (Lebesgue Differentiation III)**.** *Suppose $f \in L^1_{loc}(\mathbb{R}^n)$. For every $x \in L_f$, we have*

$$\lim_{r\to 0} \frac{1}{m(E_r)} \int_{E_r} |f(y) - f(x)|\, dy = 0 \text{ and } \lim_{r\to 0} \frac{1}{m(E_r)} \int_{E_r} f(y)\, dy = f(x)$$

for every family $\{E_r\}_{r>0}$ that shrinks nicely to x.

At this point, one may wonder why the theorem is called the Lebesgue Differentiation Theorem since, at first glance, it doesn't appear to have anything to do with differentiation. This can be made clear by considering the function

$$F(x) := \int_0^x f(y)\, dy \text{ for } f \in L^1_{\text{loc}}(\mathbb{R}).$$

We have:

$$\lim_{h\to 0} \frac{F(x+h) - F(x)}{h} = \frac{1}{h} \int_x^{x+h} f(y)\, dy = f(x) \text{ for a.e. } x \in \mathbb{R},$$

where the last equality follows from the Lebesgue Differentiation Theorem since the sets $[x, x+h]$ shrink nicely to x. In other words, if $f \in L^1_{\text{loc}}(\mathbb{R})$, then $F'(x) = f(x)$ for a.e. $x \in \mathbb{R}$. In particular, we have recovered the Fundamental Theorem of Calculus and, in the process, have justified the name of the Lebesgue Differentiation Theorem.

Another interesting and useful application of the Lebesgue Differentiation Theorem is to study the *density* of measurable subsets of $\mathbb{R}^n$. Let $E \in \mathcal{L}^n$. Applying the Lebesgue Differentiation Theorem to χ_E, we obtain

$$\lim_{r\to 0} \frac{m(E \cap B_r(x))}{m(B_r(x))} = \lim_{r\to 0} \frac{1}{m(B_r(x))} \int_{B_r(x)} \chi_E(y)\, dy = \chi_E(x) \text{ for a.e. } x \in \mathbb{R}^n.$$

In particular, if $m(E) > 0$

$$\lim_{r\to 0} \frac{m(E \cap B_r(x))}{m(B_r(x))} = 1 \text{ for a.e. } x \in E$$

and

$$\lim_{r\to 0} \frac{m(E \cap B_r(x))}{m(B_r(x))} = 0 \text{ for a.e. } x \in \mathbb{R}^n \setminus E.$$

That is, if $x \in E$ the set $E \cap B_r(x)$ has nearly full measure in $B_r(x)$ as r gets small. This means that for a.e. $x \in E$ the set E is "dense" in the ball $B_r(x)$.[2] Likewise, for a.e. $x \in E^c$ points of E do not cluster too much around x.

Before closing out the note, we quickly mention the *Vitali Covering Lemma* which is used in several of the proofs of the above theorems. The Vitali Covering Lemma is a technical lemma that is often used to extract a subcollection of sets from a larger collection. This is often useful in approximation arguments. Though it doesn't often show up on the qual, it is very useful if you are interested in continuing coursework in analysis.

Lemma 1 (Vitali Covering Lemma)**.** *Let $\mathscr{C}$ be a collection of open balls in $\mathbb{R}^n$ and let $U = \bigcup_{B\in\mathscr{C}} B$. If $c < m(U)$, then there exist disjoint $B_1, \ldots, B_k \in \mathcal{C}$ such that $\sum_1^k m(B_j) > 3^{-n}c$.*

Reference: *Real Analysis: Modern Techniques and Their Applications*, 2nd ed., Gerald B. Folland.

[2]Here, we are using dense colloquially. It does not mean the set E is dense in the topological sense.