2 Integration and Differentiation

2.1 Bonjour, Monsieur Lesbesgue!

What is the Lesbesgue measure?

- Conceptually, the Lesbesgue measure is a function defined on a particular subset of the power set of $\mathbb{R}$ that inscribes the "size" of the given set.
- When defining the Lebesgue measure, one first defines the outer measure. Specifically, the Lesbesgue measure on $\mathbb{R}$ is defined as the restriction of the outer measures on $\mathcal{M}$.

Given this definition, we can also consider Lesbesgue-measurability:

- A set $E \subseteq \mathbb{R}$ is measurable if $\forall A \subseteq \mathbb{R}$:

$$\lambda^*(A) = \lambda^*(A \cap E) + \lambda^*(A \cap E^c)$$

- Note this implies that for measurable sets A, B and $A \cap B = \emptyset$, then:

$$\lambda^*(A \cup B) = \lambda^*(A) + \lambda^*(B)$$

Why do we need this concept of measurability?

- We need it to define a σ-algebra.
- Since the intersections of σ-algebras are still σ-algebra, there exists the *smallest* σ-algebra that contains all the open sets. We call this the *Borel* algebra, $\mathcal{B}$.
- **This is a way to connect topology to measure (probability).**

2.2 Measurable Functions

Remember continuous functions? See the parallel:

- In fact, measurable functions are kinda similar to continuous functions in spirit.
- A function is said to be *measurable* (continuous) if for any measurable (open) set on the line, the pre-image of this set is measurable (open).
- An easy way to check measurability: for any real number c, take the collection of xs such that $f(x) > c$ and check if this collection is measurable. This is a result of Dynkin's $\pi - \lambda$ theorem.

Measurable functions come with nice properties:

- If f is measurable and g continuous, then the composition $f \circ g$ is measurable.
- Measurable functions can be "approximated" by a family of more tractable functions.
- Measurability is preserved under pointwise convergence: if f_n is a sequence of real-valued measurable functions on X and f_n converges to f pointwisely almost everywhere, then f is measurable.

2.3 Lesbesgue Integral

Why can't we just live with a Riemann integral?

- Consider $f(x) = 1\{x \in \mathbb{Q}\}, x \in [0, 1]$. Then f is not Riemann integrable.

In fact, you need some assumptions for a function to be Riemann integrable:

- $f : [a, b] \to \mathbb{R}$ is Riemann integrable iff f is ocntinuous almost everywhere on $[a, b]$.

So the Lesbesgue measure is our savor. Here's how you construct it:

- First, define a simple function $\varphi : X \to \mathbb{R}$:

$$\varphi := \sum_{k=1}^{n} a_k 1\{x \in A_k\}$$

  and define

$$\int_X \varphi d\lambda = \sum_{k=1}^{n} a_k \lambda(A_k)$$

- Second, let $f : X \to \mathbb{R}, f \geq 0$ be a measurable function.
  - Simple Approximation Theorem tells us that $\exists \varphi_n$ such that $0 \leq \varphi_n \leq \varphi_{n+1} \leq f$ for all $n \in \mathbb{N}$.
- Third, we define the integral for this non-negative function:

$$\int_X f d\lambda = \lim_{n \to \infty} \int_X \varphi_n d\lambda$$

- Finally, consider an arbirary measurable function $f : X \to \mathbb{R}$ and let $f^+ = \max\{f, 0\}, f^- = -\min\{f, 0\}$.
  - Notice that both are non-negative, measurable, and $f = f^+ - f^-$

- If
$$\int_X f^+ d\lambda < \infty, \int_X f^- d\lambda < \infty$$
then
$$\int_X f d\lambda = \int_X f^+ d\lambda - \int_X f^- d\lambda$$
is well-defined!

- We say that f is Lesbesgue-integrable if $\int_X |f| d\lambda < \infty$.

The Lesbesgue integral comes with nice properties:

- Monotonicity, linearity, and countable additivity
- Almost everywhere equivalence (both ways)
- Characterization of integrability: f is integrable if and only if for $\epsilon > 0$, there exists $\delta > 0$ such that
$$\int_A f dA < \epsilon$$
for any $A \in \mathcal{M}$ with $\lambda(A) < \delta$.

Here's the link between the two integrals: let f be a Riemann integrable function. Then f is Lesbesgue integrable and:
$$\int_a^b f(x)dx = \int_{[a,b]} f d\lambda$$
Before moving onto the next topic, let us visit the notion of L^p space. Fix $X \subseteq \mathbb{R}, X \in \mathcal{M}, \lambda(X) < \infty$. Then:
$$L^p(X) = \left\{ f : X \to \mathbb{R} \,\middle|\, \int_X |f|^p d\lambda < \infty \right\}$$
and further define:
$$||f||_p := \left(\int_X |f|^p d\lambda \right)^{1/p}$$
which leads us to the famous Holder inequality:

- For any $f \in L^p(X), g \in L^q(x), p, q \in (1, \infty)$ and $1/p + 1/q = 1$, then:
$$\int_X |fg| d\lambda \leq ||f||_p \cdot ||g||_q$$

2.4 Convergence Theorems

Here's some motivation for studying convergence:

- Economics is (mostly) micro-founded, which inevitably requires aggregation.
  - If there's a sequence at the micro-level, is it true that at the aggregate level the whole thing would behave as expected?
  - This means we need to think about the convergence of such sequences.
- For example, denote firms by $j \in [0, 1]$ and let $f_t(j)$ denotes it production level at time t. Suppose $f_t(j) \to f(j)$. Then would it be true that:

$$\lim_{t\to\infty} \int_0^1 f_t(j)dj = \int_0^1 f(j)dj$$

We want to know when the limit and the integral can be exchanged. The conditions rely on the following three results:

- Fatou's Lemma: "Here's the best you can do if you don't make any extra assumptions about the functions."
  - *Layman's Words:* For a sequence of non-negative measurable functions that converges pointwisely to f, the integral of f cannot be larger than the infimum of the sequence of integrals in the limit.
- Monotone Convergence Theorem (MCT) and Dominated Convergence Theorem (DCT): "If you place restrictions on both f_n and f, then you can go ahead and exchange them."
  - *MCT in Layman's words:* If f_n is non-decreasing, you're good!
  - *DCT in Layman's words:* As long as f_ns are dominated, you're good!
- Still confused? See **here**.

What if you can't find a dominating function to be used in the DCT? We have the *Vitali Convergence Theorem*. A few comments:

- It requires two concepts: uniform integrability and tightness.
- Instances that necessitate the use of Vitali are quite rare.

In economics, set of increasing functions $\mathcal{F}$ is quite a common construction. Helley's selection theorem gives us a nice property:

- It says that for any sequence $\{f_n\} \subseteq \mathcal{F}$, there exists a subsequence of the sequence that converges pointwisely to some increasing functions.

- This is a *compactness*-type theorem.
- Corollary: $\mathcal{F}$ is compact in $L^p([0,1])$ because for any f_n, there is a subsequence that converges pointwisely, and since the subsequence is between 0 and 1, we can use the DCT:

$$\forall p \in (1, \infty), \lim_{n\to\infty} \int_0^1 |f_{n_k} - f|^p d\lambda = 0$$

2.5 Stochastic Dominance

The famous integration by parts:

$$\int_a^b f(x)g'(x)dx = f(b)g(b) - f(a)g(a) - \int_a^b f'(x)g(x)dx$$

We also cover concepts of stochastic dominance:

- *First-order Stochastic Dominance*: if $\forall x \in \mathbb{R}, F(x) \leq G(x)$, we say that $F \succ_{FSD} G$.
  - X first-order stochastic dominates Y means that for each possible value of x, the probability that X has a realization greater than x is larger than that of Y.
  - If you have an increasing function $u : [a,b] \to \mathbb{R}$, and $X \succ_{FSD} Y$, then:

$$\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$$

  - A distribution F first-order stochastic dominates G if and only if for any decision maker who prefers more X, she is better-off under lottery F.
- *Second-order Stochastic Dominance*: if $\forall x \in \mathbb{R}, \int_a^x F(t)dt \leq \int_a^x G(t)dt$
  - If you have an increasing and concave function $u : [a,b] \to \mathbb{R}$, and $X \succ_{SSD} Y$, then:

$$\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$$

  - A distribution F second-order stochastic dominates G if and only if for any risk averse decision maker who prefers more X, she is better-off under lottery G.
- Intuitively, while first-order stochastic dominance describes the notion of ranking in levels, second order stochastic dominance describes the notion of dispersion.

2.6 Fubini's Theorem

How do we integrate in multi-dimensional space?

- *Fubini's theorem* allows us to decompose a multidimensional integration into many one-dimensional integrations, which we are more familiar with.

2.7 Differentiation

Motivation - differentiation is everywhere in economics! What's more interesting is connecting integration and differentiation:

- If $f : [a, b] \to \mathbb{R}$ is an increasing function, then f is differentiable almost everywhere.
- If $f : [a, b] \to \mathbb{R}$ is a convex function, then f is differentiable except for countably many points.

Unfortunately, there are cases when f is differentiable but f' is not integrable.

- *Cantor Set:* The Cantor Set, as is well-known, has measure zero. A special function defined on this set is continuous and is differentiable, but it is not integrable.

So what do we need to uphold the Fundamental Theorem of Calculus? Introduce *absolute continuity*!

- Note that when people say that the CDF of a random variable is continuous, they are referring to absolute continuity.
- In many cases, you can decompose an increasing function into an absolutely continuous component and a singular component such as the Cantor function.

Partial derivatives:

- The notion here is to fix other variables and examine one-directional derivative of a function on a projected space
- *Directional derivative* is when we examine $f(\mathbf{x})$ changes when $\mathbf{x}$ changes to $\mathbf{x} + r\mathbf{v}$ for some $r \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^n$.
- Note that the existence of a gradient vector does not imply differentiability. Nor does the existence of a Hessian imply twice differentiability.

Differentiability = existence of a linear approximation:

- For a scalar valued function of two variables, this means the existence of a tangent plane.
- If f is differentiable at $\mathbf{x}^0$, then the gradient vector exists at $\mathbf{x}^0$ and the linear functional is precisely the inner product of the gradient vector and the vector in question.
- See **here** for more clarification.

2.8 Application: Mechanism Design

A little background:

- In all the theories - consumer theory, game theory - we're studying a situation in which for a given fixed environment, how the individuals behave and how their actions aggregate. Mechanism design is the other way around - a form of engineering.

- Are there ways to efficiently allocate items? Is it possible for a social planner to design something and propose it to the society and let things play out, yet the produced outcomes have an equilibrium that is desirable?

The setup:

- Monopolist has a good, indivisible and zero cost of production

- Consumer has value $v \in [0, 1]$ where $v \sim F$ and F has density $f > 0$ on $[0, 1]$.

- Given the probability of trade p and the payment is t, the consumer gets $p\delta - t$ and the monopolist gets t.

- The monopolist proposes the mechanism $(\mathcal{M}, p, t)$, which consists of :

  - $\mathcal{M}$, a set of strategies for the consumer to use
  - $p : \mathcal{M} \to [0, 1]$, a function that specifies the probability of trade
  - $t : \mathcal{M} \to \mathbb{R}$, amount of payment based on the strategy of choice

- The consumer, given the mechanism, seeks to choose $\sigma^*(v) = m \in \mathcal{M}$ that maximizes

$$[p(m)v - t(m), 0]^+$$

Introducing the *revelation principle*:

- All mechanisms can be reduced to incentive-compatible mechanisms - it is optimal for the consumer to report the truth (the true value).

Now let's solve the monopolist's problem:

- Choose (p, t) to maximize $\mathbb{E}[t(v)]$ such that

  1. $p(v)v - t(v) \geq p(v')v' - t(v'), v, v' \in [0, 1]$ (Incentive Compatibility)
  2. $p(v)v - t(v) \geq 0, \forall v \in [0, 1]$

- Fortunately, the incentive compatibility constraint reduces the problem to just with a choice variable p, upto a constant.

- Given that $t(v) = t(0) + p(v)v - \int_0^v p(x)dx$, the payoff is now

$$\mathbb{E}[t(v)] = t(0) + \int_0^1 p(v)vf(v)dv - \int_0^1 \left(\int_0^v p(x)dx \right) f(v)dv$$

to which we apply Fubini's Theorem:

$$= t(0) + \int_0^1 p(v)vf(v)dv - \int_0^1 \left(\int_v^1 f(v)dv \right) p(x)dx$$

So the problem now is to maximize:

$$\int_0^1 p(v) \left(v - \frac{1 - F(v)}{f(v)} \right) f(v)dv$$

such that $p : [0, 1] \to [0, 1]$ is increasing.

- Why is this new characterization so valuable? Because the objective function is linear (and hence continuous). In addition, the solution *must* exist because the collection of uniformly bounded increasing functions on $[0, 1]$ is compact under the L^1 norm, and a continuous function on a compact set must have a solution.

- Note that

$$v - \frac{1 - F(v)}{f(v)}$$

is called *virtual value*. Consider the second term to be "information rent." If you maximize ignoring the second term, you're maximizing total surplus.

- Notice that we put sup for a problem where the existence of a solution is not verified. Put max if the existence is verified.