Exercise Class 10

Convergence Results and Absolute Continuity[1]

The Convergence Theorems

Let μ be a Radon measure on $\mathbb{R}^n$ and Ω be μ-measurable. Let $f_1, f_2, \cdots : \Omega \to [0, +\infty]$ be a sequence of extended real valued non-negative measurable functions and suppose that $f_n(x)$ convergese pointwise μ-a.e. to a measurable limit f. A basic question in the subject of analysis is to determine the conditions under which such pointwise convergence would imply convergence of the integral:

$$\int_\Omega f_n d\mu \overset{?}{\to} \int_\Omega f d\mu.$$

That is, when can we ensure that one can interchange integrals and limits,

$$\lim_{n\to\infty} \int_\Omega f_n d\mu \overset{?}{=} \int_\Omega \lim_{n\to\infty} f_n d\mu?$$

We already know one case in which one can safely do this, namely uniform convergence on a finite measure space. Note that the finiteness assumption is indeed necessary as one of the following examples will demonstrate. There are further cases in which one cannot interchange limits and integrals. We consider three classic examples for that (see Exercise 9.5) — all of the "moving bump" type:

Example 1 (Escape to horizontal infinity)**.** Let $\Omega = \mathbb{R}$ equipped with the Lebesgue measure λ. Define $f_n := \mathbb{1}_{[n,n+1]}$. Then f_n converges pointwise to $f := 0$, but $\int_\mathbb{R} f_n d\lambda = 1$ does not converge to $\int_\mathbb{R} f d\lambda = 0$. Intuitively, all the mass in the f_n has escaped by moving off to infinity in a horizontal direction, leaving none behind for the pointwise limit.

Example 2 (Escape to width infinity)**.** Let $\Omega = \mathbb{R}$ be equipped with the Lebesgue measure λ. Define $f_n := \frac{1}{n}\mathbb{1}_{[0,n]}$. Then f_n converges *uniformly* to $f := 0$, but $\int_\mathbb{R} f_n d\lambda = 1$ still does not converge to $\int_\mathbb{R} f d\lambda = 0$. One could prevent this from happening if all the f_n were supported on a single set of finite measure. However, the increasingly wide nature of the support of the f_n violates that.

Example 3 (Escape to vertical infinity)**.** Let $\Omega = [0, 1]$ equipped with the Lebesgue measure λ (restricted from $\mathbb{R}$). Define $f_n := n\mathbb{1}_{[\frac{1}{n},\frac{2}{n}]}$. Now, we have finite measure, and f_n converges pointwise to f, but no uniform convergence. Again, we have that $\int_{[0,1]} f_n d\lambda = 1$ is not converging to $\int_{[0,1]} f d\lambda = 0$. This time, the mass has escaped vertically through the increasingly large values of f_n.

Once one shuts down these avenues of escape to infinity, it turns out that one can recover convergence of the integral. There are two major ways to accomplish this. One is to enforce *monotonicity*, which prevents each f_n from abandoning the location where the mass of the preceding $f_1, \ldots, f_{n-1}$ was concentrated and thus shuts down the above three escape scenarios. More precisely, we have the *monotone convergence theorem*:

[1]The section on convergence theorems is taken from [Tao11], while the section on absolute continuity is based on Tobias Castelberg's notes.

Theorem 4 (Monotone Convergence Theorem, Beppo–Levi)**.** *Let $f_n : \Omega \to [0, +\infty]$ be μ-measurable for all $n \geq 1$ and be such that $f_n \leq f_{n+1}$ for all $n \geq 1$. Then it holds*

$$\int_\Omega \lim_{n\to\infty} f_n d\mu = \lim_{n\to\infty} \int_\Omega f_n d\mu.$$

Remark. One can easily see that the result still holds if the monotonicity $f_n \leq f_{n+1}$ only holds μ-a.e. rather than everywhere.

Note that in the special case when each f_n is an indicator function $f_n = \mathbb{1}_{E_n}$ for measurable sets E_n, this theorem collapses to continuity from below. Conversely, one could use the continuity from below to prove the monotone convergence theorem (instead of using Fatou's lemma).

This theorem has a number of important corollaries. One can prove a version of Tonelli's theorem for sums and integrals, the Borel–Cantelli lemma, and Fatou's lemma (provided it was not used to prove Beppo–Levi in the first place). Fatou's lemma gives us an important inequality when one does not have monotonicity.

Theorem 5 (Fatou's lemma)**.** *Let $f_n : \Omega \to [0, +\infty]$ be μ-measurable for all $n \geq 1$. Then it holds that*

$$\int_\Omega \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int_\Omega f_n d\mu.$$

Proof. Define $F_N := \inf_{n\geq N} f_n$ for each N and apply the monotone convergence theorem. □

Remark. Informally, Fatou's lemma tells us that when taking the pointwise limit of measurable functions f_n, that mass $\int_\Omega f_n d\mu$ can be destroyed in the limit (as was the case in the moving bump examples), but it cannot be created in the limit.

Finally, we consider the other major way to shut down loss of mass via escape to infinity, which is to *dominate* all of the functions involved by an absolutely convergent one. This result is known as the *dominated convergence theorem*:

Theorem 6 (Dominated Convergence Theorem, Lebesgue)**.** *Let $g : \Omega \to [0, +\infty]$ be μ-summable and $f : \Omega \to [-\infty, \infty]$, $\{f_k\}_k : \Omega \to [-\infty, \infty]$ be μ-measurable. Suppose $|f_k| \leq g$ and $f_k \to f$ μ-a.e. as $k \to \infty$. Then*

$$\lim_{k\to\infty} \int_\Omega |f_k - f| \, d\mu = 0.$$

Moreover

$$\lim_{k\to\infty} \int_\Omega f_k d\mu = \int_\Omega f d\mu.$$

From the moving bump examples we see that this statement fails if there is no μ-summable dominating function g. Note also that when each of the f_n is an indicator function $f_n = \mathbb{1}_{E_n}$ for measurable sets E_n, the dominated convergence theorem collapes to (a version) of continuity from above.

Remark. In this lecture, we deduced the dominated convergence theorem and the monotone convergence theorem from Fatou's lemma. However, these theorems are so closely related that one can obtain these theorems in a different order, depending on one's taste. It is instructive to view a couple different derivations of these key results to get more of an intuitive understanding as to how they work.

Another corollary of the monotone convergence theorem is the following:

Corollary 7. *Let $f : \Omega \to [0, \infty)$ be a μ-measurable function and let*

$$\nu(E) := \int_E f d\mu, \quad E \text{ measurable.}$$

Then ν is a measure on Ω. Further, if $g : \Omega \to [0, \infty]$ is a μ-measurable function, then

$$\int_\Omega g d\nu = \int_\Omega g f d\mu.$$

Proof. Clearly $\nu(\varnothing) = 0$. Suppose that $\{E_n\}$ is a countable disjoint family of measurable sets. Then

$$\nu\left(\bigcup_i E_i\right) := \int_\Omega \mathbb{1}_{\cup_i E_i} f d\mu.$$

Now, observe that

$$\mathbb{1}_{\cup_i E_i} f = \lim_{n \to \infty} \mathbb{1}_{\cup_{i=1}^n E_i} f = \lim_{n\to\infty} \sum_{i=1}^{\infty} \mathbb{1}_{E_i} f,$$

where $(\sum_{i=1}^n \mathbb{1}_{E_i} f)$ is an increasing sequence of non-negative measurable functions. Hence by the monotone convergence theorem,

$$\int_\Omega \mathbb{1}_{\cup_i E_i} f d\mu = \lim_{n\to\infty} \int_\Omega \sum_{i=1}^{n} \mathbb{1}_{E_i} f = \sum_{i=1}^{\infty} \int_\Omega \mathbb{1}_{E_i} f.$$

Thus,

$$\nu\left(\bigcup_i E_i\right) = \sum_{i=1}^{\infty} \nu(E_i).$$

This completes the first part of the corollary.

Next, we observe that the second statements holds if g is a characteristic function of a measurable set. Indeed, if $g = \mathbb{1}_E$ for some measurable set, then

$$\int_\Omega g d\nu = \int_\Omega \mathbb{1}_E d\nu = \nu(E)$$

and

$$\int_\Omega g f d\mu = \int_\Omega \mathbb{1}_E f d\mu = \int_E f d\mu$$

so that the equality holds by definition. Now, using the linearity of the integral, the equality holds for all simple non-negative measurable functions as well. Since any measurable function $g : \Omega \to [0, \infty]$ is a pointwise limit of an increasing sequence of simple non-negative measurable functions, the proof can be completed by invoking the monotone convergence theorem (or Lebesgue's theorem). □

The relation in the previous corollary is usually written as

$$d\nu = f d\mu \quad \text{or} \quad \frac{d\nu}{d\mu} = f,$$

and f is called the *Radon–Nikodym derivative* of ν with respect to μ. The notation is very convenient as

$$\int_E d\nu = \int_E f d\mu = \int_E \frac{d\nu}{d\mu} d\mu.$$

Note that we have the following implication

$$\mu(E) = 0 \implies \nu(E) = 0$$

for all μ-measurable sets E.

This phenomenon is called *absolute continuity* and it will be the topic of the following section.

Absolute Continuity

The concept of absolute continuity is not just relevant in probability theory, but also in statistics. In statistics, you somehow do the opposite of what you do in probability theory: you start with samples and try to deduce which distribution generated the samples.

Example 8. Suppose you measure the number of customers in a supermarket during the span of a week from 9am to 12am. We get the following result:

$$(745, 692, 715, 1012, 557, 545).$$

We obtain 711 customers as an arithmetic mean and consequently, we consider the following models:

- $X \overset{P}{\sim} \text{Bin}(1000, 0.711)$;
- $X \overset{Q}{\sim} \text{Poisson}(711)$.

Both models satisfy $\mathbb{E}[X] = 711$ (without really defining what $\mathbb{E}$ means), however, the Binomial distribution does not make a lot of sense for this date. It is impossible that a binomially distributed random variable with parameters $n = 1000$ and $p \in [0, 1]$ takes the value 1012! Thus the sample could not have been generated by $\text{Bin}(1000, 0.711)$.

The concept behind this phenomenon is called absolute continuity. Let P and Q be two probability measures on the same sample space $(\Omega, \mathcal{A})$. Suppose you consider an arbitrary sample $\omega \in \Omega$ generated by Q. If it is possible to eliminate that ω was generated by P, then Q is not absolutely continuous with respect to P.

In Example 8, we have that $X(\omega) = 1012$ is a sample which was possibly generated by Q, since $Q[X = 1012] = e^{-711} \frac{711^{1012}}{1012!} > 0$. On the other hand, we have that $P[X = 1012] = 0$ and hence X could not have been generated by P. In other words, Q is not absolutely continuous with respect to P.

P, on the other hand, is absolutely with respect to Q, since every value $x = X(\omega) \in \{0, 1, \dots, 1000\}$ which can possibly be generated by P, can also by generated by Q. Thus, if P has generated the sample, we cannot say with certainty whether P or Q has generated the sample by just looking at it.

Definition 9. Suppose that μ and ν are measures on a measurable space (Ω, Σ). Then ν is *absolutely continuous with respect to* μ if for every $E \in \Sigma$

$$\mu(E) = 0 \implies \nu(E) = 0,$$

and we write $\nu \ll \mu$.

If we define ν as in Corollary 7, then we know that $\nu \ll \mu$. Now we can ask ourselves whether the converse is true as well. That is, if μ and ν are measures such that $\nu \ll \mu$, then does there exist a measurable function $f \geq 0$ such that

$$\nu(E) = \int_E f d\mu \quad \forall E \in \Sigma.$$

The answer is yes and this result is called the *Radon–Nikodym theorem.*

Theorem 10 (Radon–Nikodym)**.** *If μ and ν are σ-finite measures on a measurable space Ω such that $\nu \ll \mu$, then there exists a non-negative measurable function f, such that*

$$\nu(E) = \int_E f d\mu$$

for all measurable sets E.

The Radon–Nikodym theorem is important in probability theory due to the following consideration.

Recall from Exercise Class 6 that if $(\Omega, \mathcal{F}, P)$ is a probability space and $X : \Omega \to \mathbb{R}$ is a random variable on it, then the probability distribution of X is the probability measure ν defined on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ by the pushforward

$$\nu(B) = P(X^{-1}(B)), \quad B \in \mathcal{B}(\mathbb{R}).$$

Furthermore, we defined the distribution function of X as the function $F : \mathbb{R} \to \mathbb{R}$ with

$$F(x) := P(X \leq x).$$

Thus, if we can establish that ν is absolutely continuous with respect to the probability measure Radon measure μ on $\mathbb{R}$, then as a consequence of Radon–Nikodym, there exists a non-negative μ-measurable function $f : \mathbb{R} \to \mathbb{R}$ such that

$$F(t) = \int_{(-\infty, t]} f d\mu, \quad t \in \mathbb{R}.$$

Such a function f, given that it exists, is called the *probability density function* of the random variable X. In short, a probability density function is just a Radon–Nikodym derivative. This allows us to fully motivate the usual formula to compute expectations.

We usually consider two measures for the Radon measure μ. For discrete random variables, we take μ to be the counting measure. Recall that integration with respect to the counting measure is just summation. Combine this with the change of variables formula from Exercise Class 6 to obtain that for a discrete random variable with absolutely continuous distribution, we have

$$\mathbb{E}[X] = \int_\Omega X(\omega) dP(\omega) \overset{\text{C.o.V.}}{=} \int_\mathbb{N} x d\nu(x) \overset{\text{R.N.}}{=} \int_\mathbb{N} x f(x) d\mu(x) = \sum_{n=0}^{\infty} x P(X = x).$$

Similarly, if we have a continuous random variable such that its distribution is absolutely continuous with respect to the Lebesgue measure, we have

$$\mathbb{E}[X] = \int_\Omega X(\omega) dP(\omega) \overset{\text{C.o.V.}}{=} \int_\mathbb{R} x d\nu(x) \overset{\text{R.N.}}{=} \int_\mathbb{R} x f(x) d\mu(x).$$

Remark. Not every distribution is absolutely continuous with resepct to the Lebesgue or counting measure. We have already encountered the Cantor distribution as such an example in Exercise Class 4.

Sources

[Tao11] Terence Tao. *An introduction to measure theory*. Vol. 126. 2011. URL: https://terrytao.files.wordpress.com/2011/01/measure-book1.pdf.