MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J Fall 2008
Lecture 12 10/20/2008

ABSTRACT INTEGRATION — II

Contents

1. Borel-Cantelli revisited
2. Connections between abstract integration and elementary definitions of integrals and expectations
3. Fatou's lemma
4. Dominated convergence theorem

In the previous lecture:

(a) We defined the notion of an integral of a measurable function with respect to a measure ($\int g\, d\mu$), which subsumes the special case of expectations ($\mathbb{E}[X] = \int X\, d\mathbb{P}$), where X is a random variable, and $\mathbb{P}$ is a probability measure.

(b) We saw that integrals are always well-defined, though possibly infinite, if the function being integrated is nonnegative.

(c) For a general function g, we decompose it as the sum $g = g_+ - g_-$ of a positive and a negative function, and integrate each piece separately. The integral is well defined unless both $\int g_+\, d\mu$ and $\int g_-\, d\mu$ happen to be infinite.

(d) We saw that integrals obey a long list natural properties, including linearity: $\int (g + h)\, d\mu = \int g\, d\mu + \int h\, d\mu$.

(e) We stated the Monotone Convergence Theorem (MCT), according to which, if $\{g_n\}$ is a nondecreasing sequence of nonnegative measurable functions that converge pointwise to a function g, then $\lim_{n\to\infty} \int g_n\, d\mu = \int g\, d\mu$.

(f) Finally, we saw that for every nonnegative measurable function g, we can find an nondecreasing sequence of nonnegative simple functions that converges (pointwise) to g.

1 BOREL-CANTELLI REVISITED

Recall that one of the Borel-Cantelli lemmas states that if $\sum_{i=1}^{\infty} \mathbb{P}(A_i) < \infty$, then $\mathbb{P}(A_i \text{ i.o.}) = 0$. In this section, we rederive this result using the new machinery that we have available.

Let X_i be the indicator function of the event A_i, so that $\mathbb{E}[X_i] = \mathbb{P}(A_i)$. Thus, by assumption $\sum_{i=1}^{\infty} \mathbb{E}[X_i] < \infty$. The random variables $\sum_{i=1}^{n} X_i$ are nonnegative and form an increasing sequence, as n increases. Furthermore,

$$\lim_{n\to\infty} \sum_{i=1}^{n} X_i = \sum_{i=1}^{\infty} X_i,$$

pointwise; that is, for every ω, we have $\lim_{n\to\infty} \sum_{i=1}^{n} X_i(\omega) = \sum_{i=1}^{\infty} X_i(\omega)$.

We can now apply the MCT, and then the linearity property of expectations (for finite sums), to obtain

$$\begin{aligned}
\mathbb{E}\Big[\sum_{i=1}^{\infty} X_i\Big] &= \lim_{n\to\infty} \mathbb{E}\Big[\sum_{i=1}^{n} X_i\Big] \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}[X_i] \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(A_i) \\
&= \sum_{i=1}^{\infty} \mathbb{P}(A_i) \\
&< \infty.
\end{aligned}$$

This implies that $\sum_{i=1}^{\infty} X_i < \infty$, a.s. (This is intuitively obvious, but a short formal proof is actually needed.) It follows that, with probability 1, only finitely many of the events A_i can occur. Equivalently, the probability that infinitely many of the events A_i occur is zero, i.e., $\mathbb{P}(A_i \text{ i.o.}) = 0$.

2 CONNECTIONS BETWEEN ABSTRACT INTEGRATION AND ELEMENTARY DEFINITIONS OF INTEGRALS AND EXPECTATIONS

Abstract integration would not be useful theory if it were inconsistent with the more elementary notions of integration. For discrete random variables taking values in a finite range, this consistency is automatic because of the definition of an integral of a simple function. We will now verify some additional aspects of this consistency.

2.1 Connection with Riemann integration.

We state here, without proof, the following reassuring result. Let $\Omega = \mathbb{R}$, endowed with the usual Borel σ-field. Let λ be the Lebesgue measure. Suppose that g is a measurable function which is Riemann integrable on some interval $[a, b]$. Then, the Riemann integral $\int_a^b g(x)\,dx$ is equal to $\int_{[a,b]} g\,d\lambda = \int 1_{[a,b]} g\,d\lambda$.

2.2 Evaluating expectations by integrating on different spaces

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $X : \Omega \to \mathbb{R}$, be a random variable. We then obtain a second probability space $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$, where $\mathcal{B}$ is the Borel σ-field, and $\mathbb{P}_X$ is the probability law of X, defined by

$$\mathbb{P}_X(A) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in A\}), \qquad A \in \mathcal{B}.$$

Consider now a measurable function $g : \mathbb{R} \to \mathbb{R}$, and use it to define a new random variable $Y = g(X)$, and a corresponding probability space $(\mathbb{R}, \mathcal{B}, \mathbb{P}_Y)$. The expectation of Y can be evaluated in three different ways, that is, by integrating over either of the three spaces we have introduced.

Theorem 1. *We have*

$$\int Y\,d\mathbb{P} = \int g\,d\mathbb{P}_X = \int y\,d\mathbb{P}_Y,$$

assuming that the integrals are well defined.

Proof: We follow the "standard program": first establish the result for simple functions, then take the limit to deal with nonnegative functions, and finally generalize.

Let g be a simple function, which takes values in a finite set $y_1, \ldots, y_k$. Using the definition of the integral of a simple function we have

$$\int Y\,d\mathbb{P} = \sum_{y_i} y_i \mathbb{P}(\{\omega \mid Y(\omega) = y_i\}) = \sum_{y_i} y_i \mathbb{P}(\{\omega \mid g(X(\omega)) = y_i\}).$$

Similarly,

$$\int g\,d\mathbb{P}_X = \sum_{y_i} y_i \mathbb{P}_X(\{x \mid g(x) = y_i\}).$$

However, from the definition of $\mathbb{P}_X$, we obtain

$$\begin{aligned}\mathbb{P}_X(\{x \mid g(x) = y_i\}) &= \mathbb{P}_X(g^{-1}(y_i)) \\ &= \mathbb{P}(\{\omega \mid X(\omega) \in g^{-1}(y_i)\}) \\ &= \mathbb{P}(\{\omega \mid g(X(\omega)) = y_i\}),\end{aligned}$$

and the first equality in the theorem follows, for simple functions.

Let now g be nonnegative function, and let $\{g_n\}$ be an increasing sequence of nonnegative simple functions that converges to g. Note that $g_n(X)$ converges monotonically to $g(X)$. We then have

$$\int Y\,d\mathbb{P} = \int g(X)\,d\mathbb{P} = \lim_{n\to\infty} \int g_n(X)\,d\mathbb{P} = \lim_{n\to\infty} \int g_n\,d\mathbb{P}_X = \int g\,d\mathbb{P}_X.$$

(The second equality is the MCT; the third is the result that we already proved for simple functions; the last equality is once more the MCT.)

The case of general (not just nonnegative) functions follows easily from the above – the details are omitted. This proves the first equality in the theorem.

For the second equality, note that by considering the special case $g(x) = x$, we obtain $Y = X$ and $\int X\,d\mathbb{P} = \int x\,d\mathbb{P}_X$. By a change of notation, we have also established that $\int Y\,d\mathbb{P} = \int y\,d\mathbb{P}_Y$. $\square$

2.3 The case of continuous random variables, described by PDFs

We can now revisit the development of continuous random variables (Lecture 8), in a more rigorous manner. We say that a random variable $X : \Omega \to \mathbb{R}$ is continuous if its CDF can be written in the form

$$F_X(x) = \mathbb{P}(X \le x) = \int 1_{(-\infty,x]} f\,d\lambda, \qquad \forall x \in \mathbb{R},$$

where λ is Lebesgue measure, and f is a nonnegative measurable function. It can then be shown that for any Borel subset A of the real line, we have

$$\mathbb{P}_X(A) = \int_A f\,d\lambda. \tag{1}$$

When f is Riemann integrable and the set A is an interval, we can also write $\mathbb{P}_X(A) = \int_A f(x)\,dx$, where the latter integral is an ordinary Riemann integral.

Theorem 2. *Let g be a measurable function which is either nonnegative, or satisfies $\int |g|\, d\mathbb{P}_X < \infty$. Then,*

$$\mathbb{E}[g(X)] = \int g\, d\mathbb{P}_X = \int (gf)\, d\lambda.$$

Proof: The first equality holds by definition. The second was shown in Theorem 1. So, let us concentrate on the third. Following the usual program, let us first consider the case where g is a simple function, of the form $g = \sum_{i=1}^{k} a_i 1_{A_i}$, for some measurable disjoint subsets A_i of the real line. We have

$$\begin{aligned}
\int g\, d\mathbb{P}_X &= \sum_{i=1}^{k} a_i \mathbb{P}_X(A_i) \\
&= \sum_{i=1}^{k} a_i \int_{A_i} f\, d\lambda \\
&= \sum_{i=1}^{k} \int a_i 1_{A_i} f\, d\lambda \\
&= \int \sum_{i=1}^{k} a_i 1_{A_i} f\, d\lambda \\
&= \int (gf)\, d\lambda.
\end{aligned}$$

The first equality is the definition of the integral for simple functions. The second uses Eq. (1). The fourth uses linearity of integrals. The fifth uses the definition of g.

Suppose now that g is a nonnegative function, and let $\{g_n\}$ be an increasing sequence of nonnegative functions that converges to g, pointwise. Since f is nonnegative, note that $g_n f$ also increases monotonically and converges to gf. Then,

$$\int g\, d\mathbb{P}_X = \lim_{n\to\infty} \int g_n\, d\mathbb{P}_X = \lim_{n\to\infty} \int (g_n f)\, d\lambda = \int (gf)\, d\lambda.$$

The first and the third equality above is the MCT. The middle equality is the result we already proved, for the case of a simple function g_n.

Finally, if g is not nonnegative, the result is proved by considering separately the positive and negative parts of g. □

When g and f are "nice" functions, e.g., piecewise continuous, Theorem 2 yields the familiar formula

$$\mathbb{E}[g(X)] = \int g(x)f(x)\,dx,$$

where the integral is now an ordinary Riemann integral.

3 FATOU'S LEMMA

Note that for any two random variables, we have $\min\{X, Y\} \le X$ and $\min\{X, Y\} \le Y$. Taking expectations, we obtain $\mathbb{E}[\min\{X, Y\}] \le \min\{\mathbb{E}[X], \mathbb{E}[Y]\}$. Fatou's lemma is in the same spirit, except that infinitely many random variables are involved, as well as a limiting operation, so some additional technical conditions are needed.

Theorem 3. *Let Y be a random variable that satisfies $\mathbb{E}[|Y|] < \infty$.*

(a) If $Y \le X_n$, for all n, then $\mathbb{E}[\liminf_{n\to\infty} X_n] \le \liminf_{n\to\infty} \mathbb{E}[X_n]$.

(b) If $X_n \le Y$, for all n, then $\mathbb{E}[\limsup_{n\to\infty} X_n] \ge \limsup_{n\to\infty} \mathbb{E}[X_n]$.

By considering the case where $Y = 0$, we see that parts (a) and (b) apply in particular to the case of nonnegative (respectively, nonpositive) random variables.

Proof: Let us only prove the first part; the second part frollows from a symmetrical argument, or by simply applying the first part to the random variables $-Y$ and $-X_n$.

Fix some n. We have

$$\inf_{k\ge n} X_k - Y \le X_m - Y, \qquad \forall\, m \ge n.$$

Taking expectations, we obtain

$$\mathbb{E}[\inf_{k\ge n} X_k - Y] \le \mathbb{E}[X_m - Y], \qquad \forall\, m \ge n.$$

Taking the infimum of both sides with respect to m, we obtain

$$\mathbb{E}[\inf_{k\ge n} X_k - Y] \le \inf_{m\ge n} \mathbb{E}[X_m - Y].$$

Note that the sequence $\inf_{k\ge n} X_k - Y$ is nonnegative and nondecreasing with n, and converges to $\liminf_{n\to\infty} X_n - Y$. Taking the limit of both sides of the preceding inequality, and using the MCT for the left-hand side term, we obtain

$$\mathbb{E}[\liminf_{n\to\infty} X_n - Y] \le \liminf_{n\to\infty} \mathbb{E}[X_n - Y].$$

Note that $\mathbb{E}[X_n - Y] = \mathbb{E}[X_n] - \mathbb{E}[Y]$. (This step makes use of the assumption that $\mathbb{E}[|Y|] < \infty$.) For similar reasons, the term $-\mathbb{E}[Y]$ can also be removed from the right-hand side. After canceling the terms $\mathbb{E}[Y]$ from the two sides, we obtain the desired result. □

We note that Fatou's lemma remains valid for the case of general measures and integrals.

4 DOMINATED CONVERGENCE THEOREM

The dominated convergence theorem complements the MCT by providing an alternative set of conditions under which a limit and an expectation can be interchanged.

Theorem 4. (DCT) *Consider a sequence of random variables $\{X_n\}$ that converges to X (pointwise). Suppose that $|X_n| \le Y$, for all n, where Y is a random variable that satisfies $\mathbb{E}[|Y|] < \infty$. Then, $\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$.*

Proof: Since $-Y \le X_n \le Y$, we can apply both parts of Fatou's lemma , to obtain

$$\mathbb{E}[X] = \mathbb{E}[\liminf_{n\to\infty} X_n] \le \liminf_{n\to\infty} \mathbb{E}[X_n] \le \limsup_{n\to\infty} \mathbb{E}[X_n] \le \mathbb{E}[\limsup_{n\to\infty} X_n] = \mathbb{E}[X].$$

This proves that

$$\mathbb{E}[X] = \liminf_{n\to\infty} \mathbb{E}[X_n] = \limsup_{n\to\infty} \mathbb{E}[X_n].$$

In particular, the limit $\lim_{n\to\infty} \mathbb{E}[X_n]$ exists and equals $\mathbb{E}[X]$. □

A special case of the DCT is the BCT (Bounded Convergence Theorem), which asserts that if there exists a constant $c \in \mathbb{R}$ such that $|X_n| \le c$, a.s., for all n, then $\mathbb{E}[X_n] \to X$.

Corollary 1. *Suppose that $\sum_{n=1}^{\infty} \mathbb{E}[|Z_n|] < \infty$. Then,*

$$\sum_{n=1}^{\infty} \mathbb{E}[Z_n] = \mathbb{E}\Big[\sum_{n=1}^{\infty} Z_n\Big].$$

Proof: By the monotone convergence theorem, applied to $Y_n = \sum_{k=1}^{n} |Z_k|$, we have

$$\mathbb{E}\Big[\sum_{n=1}^{\infty} |Z_n|\Big] = \sum_{n=1}^{\infty} \mathbb{E}[|Z_n|] < \infty.$$

Let $X_n = \sum_{i=1}^{n} Z_i$ and note that $\lim_{n\to\infty} X_n = \sum_{i=1}^{\infty} Z_n$. We observe that $|X_n| \leq \sum_{i=1}^{\infty} |Z_i|$, which has finite expectation, as shown earlier. The result follows from the dominated convergence theorem. $\square$

Exercise: Can you prove Corollary 1 directly from the monotonone convergence theorem, without appealing to the DCT or Fatou's lemma?

Remark: We note that the MCT, DCT, and Fatou's lemma are also valid for general measures, not just for probability measures (the proofs are the same); just replace expressions such as $\mathbb{E}[X]$ by integrals $\int g \, d\mu$.