ABSTRACT INTEGRATION — II

Contents

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 the following reassuring result. Let $\Omega = \mathbb{R}$ and λ be the Lebesgue measure (considered on either Borel σ-algebra, or its completion the Lebesgue σ-algebra). Suppose that f is a Riemann integrable function on some interval $[a, b]$. Then, f is *Lebesgue*-measurable and its Lebesgue integral equals the Riemann integral:

$$\int_a^b f(x)\,dx = \int_{[a,b]} f\,d\lambda = \int_{\mathbb{R}} 1_{[a,b]} f\,d\lambda$$

In particular, every *Borel* function's Lebesgue integral coincides with its Riemann integral whenever the latter exists.

Proof (optional). Consider an arbitrary finite partition σ of $[a, b]$. Corresponding to each σ there is a piece-wise constant (hence simple) function $f_\sigma(x) \leq f(x)$ and $f'_\sigma(x) \geq f(x)$ such that

$$\int_{[a,b]} f_\sigma d\lambda = L(\sigma) \tag{1}$$

$$\int_{[a,b]} f'_\sigma d\lambda = U(\sigma) \tag{2}$$

where $L(\sigma)$ and $U(\sigma)$ are lower and upper Darboux sums (defined in Lecture 7). There exists a sequence of partitions σ_n, each refining the previous one, such that

$$L(\sigma_n) \nearrow \sup_\sigma L(\sigma) \tag{3}$$

$$U(\sigma_n) \searrow \inf_\sigma U(\sigma)\,. \tag{4}$$

On the other hand the corresponding sequences of functions f_{σ_n} and f'_{σ_n} are monotone, hence converging:

$$f_{\sigma_n}(x) \nearrow \underline{f}(x) \leq f(x) \qquad \forall x \in [a, b] \tag{5}$$

$$f'_{\sigma_n}(x) \searrow \overline{f}(x) \geq f(x) \qquad \forall x \in [a, b] \tag{6}$$

From (1)-(2) and Riemann integrability we conclude that

$$\int_{[a,b]} \underline{f} d\lambda = \int_{[a,b]} \overline{f} d\lambda = \int_a^b g(x)dx\,. \tag{7}$$

Consequently, $\int |\overline{f} - \underline{f}| d\lambda = 0$ and thus

$$\underline{f}(x) = \overline{f}x = f(x) \qquad \text{for a.e. } x$$

implying f is Lebesgue measurable (and coincides with some Borel measurable function except on a set of measure zero). The equality of Lebesgue and Riemann integrals of f in turn follows from (7). □

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

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

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*

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

and all three $\int$*'s exist or do not exist simultaneously.*

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

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

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 equalities in the theorem follow, 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 theorem.

□

2.3 The case of continuous random variables, described by PDFs

We can now revisit the development of continuous random variables (Lecture 4), 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

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

where λ is Lebesgue measure, and f is a nonnegative measurable function wint $\int_{\mathbb{R}} f d\lambda = 1$. Recall that by Theorem 4 of Lecture 4 to each CDF there corresponds a unique probability measure $\mathbb{P}_X$ on $(\mathbb{R}, \mathcal{B})$. In this case $\mathbb{P}_X$ has a particularly simple expression:

$$\mathbb{P}_X(A) = \int_A f \, d\lambda \tag{8}$$

for any Borel set A. (Obviously, the CDF of $\mathbb{P}_X$ is F_X. The fact that (8) defines a valid measure is **property 10** from Lecture 7.)

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

Theorem 2. *For any measurable function g we have*

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

where all $\mathbb{E}$ and $\int$'s exist or do not exist simultaneously.

Note: Since integrals of non-negative functions always exist, this also gives a convenient criterion: $\mathbb{E}[g(X)]$ is finite iff $\int |g| f d\lambda < \infty$.

Proof: The first equality was shown in Theorem 1. So, let us concentrate on the second. 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. (8). 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

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\,dx,$$

where the integral is now an ordinary (improper) Riemann integral.

3 FATOU'S LEMMA

Note that for any two random variables, we have $\min\{X, Y\} \leq X$ and $\min\{X, Y\} \leq Y$. Taking expectations, we obtain $\mathbb{E}[\min\{X, Y\}] \leq \min\{\mathbb{E}[X], \mathbb{E}[Y]\}$. 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 $f_n \geq 0$ be measurable, then*

$$\int \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int f_n d\mu$$

Proof: Fix some n. We have

$$\inf_{k\geq n} f_k \leq f_m, \qquad \forall\, m \geq n.$$

Integrating both sides, we obtain

$$\int \inf_{k\geq n} f_k\, d\mu \leq \int f_m\, d\mu, \qquad \forall\, m \geq n.$$

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

$$\int \inf_{k\geq n} f_k\, d\mu \leq \inf_{m\geq n} \int f_m\, d\mu \tag{9}$$

The statement of the Theorem follows from (9) after taking the limit $\lim_{n\to\infty}$. Indeed, the sequence $\inf_{k\geq n} f_k$ is nonnegative and nondecreasing with n, and converges to $\liminf_{n\to\infty} f_n$. Therefore, from MCT we obtain

$$\lim_{n\to\infty} \int \inf_{k\geq n} f_k\, d\mu = \int \lim_{n\to\infty} \inf_{k\geq n} f_k \triangleq \int \liminf_{n\to\infty} f_n\, d\mu$$

Similarly, the limit as $n \to \infty$ of the right-hand side of (9) converges to $\liminf \int f_n d\mu$. $\square$

Corollary 1. *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 $\mathbb{E}[\limsup_{n\to\infty} X_n] \ge \limsup_{n\to\infty} \mathbb{E}[X_n]$.

Proof: Apply Theorem 3 to $X_n - Y$ or $Y - X_n$. $\square$

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 a.e. Suppose that $|X_n| \le Y$, for all n, where Y is a nonnegative random variable that satisfies $\mathbb{E}[Y] < \infty$. Then, $\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$.*

Proof: Let $A \subset \Omega$ be the set of outcomes ω along which $X_n(\omega) \to X(\omega)$ as $n \to \infty$. Then $\mathbb{P}(A^c) = 0$. Let $\tilde{X}_n(\omega) = X_n(\omega)$ for $\omega \in A$ and $= 0$ otherwise. Similarly, let $\tilde{X}(\omega) = X(\omega), \omega \in A$ and $= 0$ otherwise. Then $\mathbb{E}[\tilde{X}_n] = \mathbb{E}[X_n], \mathbb{E}[\tilde{X}] = \mathbb{E}[X]$ and $\tilde{X}_n \to \tilde{X}$ for all ω. Thus we may assume, without the loss of generality that $X_n(\omega) \to X(\omega)$ for all ω.

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 $\mathbb{E}[X]$. $\square$

Remark: We note that the DCT remains valid for general measures, not just for probability measures (the proof is the same). However, the following statement (Bounded Convergence Theorem), is specific to probability measures: *If there exists a constant $c \in \mathbb{R}$ such that $|X_n| \le c$, a.s., for all n, then $\lim_n \mathbb{E}[X_n] = \mathbb{E}[\lim_n X_n]$.*

Corollary 2. *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. □

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

Theorems such as MCT and DCT impose assumptions additional to the assumption that $X_n \to X$ a.e. that insure that $\lim_n \mathbb{E}[X_n] = \mathbb{E}[X]$. It should not be surprising that, in general just having $X_n \to X$ a.e. is not enough. Here is a counter-example. Let $\Omega = [0, 1]$, let $\mathcal{F}$ be the Borel sigma-field $\mathcal{B}$ on $[0, 1]$, and let $\mathbb{P}$ be the uniform (Lebesgue) probability measure. Let $X(\omega) = 0$ for all $\omega \in [0, 1]$. Let

$$X_n(\omega) = \begin{cases} n, & \text{when } \omega \in (0, \frac{1}{n}); \\ 0, & \text{when } \omega = 0 \text{ or } \omega \in (\frac{1}{n}, 1]. \end{cases}$$

Verify that $X_n(\omega) \to 0$ for all ω, but $\mathbb{E}[X_n] = n(1/n) = 1$ and thus $\mathbb{E}[X_n] \to 0$ does not hold.

The example above shows that DCT does not hold unless we make an additional assumption, such as $|X_n| \leq Y$ for some random variable Y with $\mathbb{E}[Y] < \infty$. However, the sequence X_n is not increasing.

Exercise:

(a) Establish the following generalization of the MCT. Suppose X_n is a.e. increasing sequence of random variables, but suppose X_n are not necessarily non-negative. Let $\lim_n X_n = X$ a.e. Suppose $X_n \geq Y$ a.e. for some random variable Y. Finally, suppose the expectations of X_n, X and Y are all finite. Establish that $\lim_n \mathbb{E}[X_n] = \mathbb{E}[X]$.

(b) Construct a sequence of random variables X_n which is increasing a.e., but $\mathbb{E}[X_n]$ does not converge to $\mathbb{E}[X]$, where $X = \lim_n X_n$ a.e.