12. LIMIT THEOREMS FOR EXPECTATION

Theorem 46 (Monotone convergence theorem (MCT))**.** *Suppose X_n, X are non-negative r.v.s and $X_n \uparrow X$ a.s. Then* $\mathbf{E}[X_n] \uparrow \mathbf{E}[X]$. *(valid even when $\mathbf{E}[X] = +\infty$).*

Theorem 47 (Fatou's lemma)**.** *Let X_n be non-negative r.v.s. Then* $\mathbf{E}[\liminf X_n] \leq \liminf \mathbf{E}[X_n]$.

Theorem 48 (Dominated convergence theorem (DCT))**.** *Let $|X_n| \leq Y$ where Y is a non-negative r.v. with* $\mathbf{E}[Y] < \infty$*. If $X_n \to X$ a.s., then,* $\mathbf{E}[|X - n - X|] \to 0$ *and hence we also get* $\mathbf{E}[X_n] \to \mathbf{E}[X]$.

Assuming MCT, the other two follow easily. For example, to prove Fatou's lemma, just define $Y_n = \inf_{n \geq k} X_n$ and observe that Y_ks increase to $\liminf X_n$ a.s and hence by MCT $\mathbf{E}[Y_k] \to \mathbf{E}[\liminf X_n]$. Since $X_n \geq Y_n$ for each n, we get $\liminf \mathbf{E}[X_n] \geq \liminf \mathbf{E}[Y_n] = \mathbf{E}[\liminf X_n]$.

To prove DCT, first note that $|X_n| \leq Y$ and $|X| \leq Y$ a.s. Consider the sequence of non-negative r.v.s $2Y - |X_n - X|$ that converges to $2Y$ a.s. Then, apply Fatou's lemma to get

$$\mathbf{E}[2Y] = \mathbf{E}[\liminf(2Y - |X_n - X|)] \leq \liminf\ \mathbf{E}[2Y - |X_n - X|] = \mathbf{E}[2Y] - \limsup \mathbf{E}[|X_n - X|].$$

Thus $\limsup \mathbf{E}[|X_n - X|] = 0$. Further, $|\mathbf{E}[X_n] - \mathbf{E}[X]| \leq \mathbf{E}[|X_n - X|] \to 0$.

13. LEBESGUE INTEGRAL VERSUS RIEMANN INTEGRAL

Consider the probability space $([0,1], \overline{\mathcal{B}}, \mathbf{m})$ (note that this is the Lebesgue σ-algebra, not Borel!) and a function $f : [0,1] \to \mathbb{R}$. Let

$$U_n := \frac{1}{2^n} \sum_{k=0}^{2^n - 1} \max_{\frac{k}{2^n} \leq x \leq \frac{k+1}{2^n}} f(x), \qquad U_n := \frac{1}{2^n} \sum_{k=0}^{2^n - 1} \min_{\frac{k}{2^n} \leq x \leq \frac{k+1}{2^n}} f(x)$$

be the upper and lower Riemann sums. Then, $L_n \leq U_n$ and U_n decrease with n while L_n increase. *If* $\lim U_n = \lim L_n$, we say that f is Riemann integrable and this common limit is defined to be the Riemann integral of f. The question of which functions are indeed Riemann integrable is answered precisely by

Lebesgue's theorem on Riemann integrals: *A bounded function f is Riemann integrable if and only if the set of discontinuity points has zero Lebesgue outer measure.*

Next consider the Lebesgue integral $\mathbf{E}[f]$. For this we need f to be Lebesgue measurable in the first place. Clearly any bounded and measurable function is integrable (why?). Plus, if f is continuous a.e., then f is measurable (why?). Thus, Riemann integrable functions are also Lebesgue integrable (but not conversely). What about the values of the two kinds of integrals? Define

$$g_n(x) := \sum_{k=0}^{2^n - 1} \left(\max_{\frac{k}{2^n} \leq x \leq \frac{k+1}{2^n}} f(x) \right) \mathbf{1}_{\frac{k}{2^n} \leq x \leq \frac{k+1}{2^n}}, \qquad h_n(x) := \sum_{k=0}^{2^n - 1} \left(\min_{\frac{k}{2^n} \leq x \leq \frac{k+1}{2^n}} f(x) \right) \mathbf{1}_{\frac{k}{2^n} \leq x \leq \frac{k+1}{2^n}}$$

so that $\mathbf{E}[g_n] = U_n$ and $\mathbf{E}[h_n] = L_n$. Further, $g_n(x) \downarrow f(x)$ and $h_n(x) \uparrow f(x)$ at all continuity points of f. By MCT, $\mathbf{E}[g_n]$ and $\mathbf{E}[h_n]$ converge to $\mathbf{E}[f]$, while by the assumed Riemann integrability L_n and U_n converge to $\int_0^1 f(x)dx$ (Riemann integral). Thus we must have $\mathbf{E}[f] = \int_0^1 f(x)dx$.

In short, when a function is Riemann integrable, it is also Lebesgue integrable, and the integrals agree. But there are functions that are measurable but not a.e. continuous. For example, consider the indicator function of a totally disconnected set of positive Lebesgue measure (like a Cantor set where an α middle portion is deleted at each stage, with α sufficiently small). Then at each point of the set, the indicator function is discontinuous. Thus, Lebesgue integral is more powerful than Riemann integral.

14. LEBESGUE SPACES:

Fix $(\Omega, \mathcal{F}, \mathbf{P})$. For $p \geq 1$, define $\|X\|_p := \mathbf{E}[|X|^p]^{\frac{1}{p}}$ for those r.v.s for which this number is finite. Then $\|tX\|_p = t\|X\|_p$ for $t > 0$, and Minkowski's inequality gives $\|X+Y\|_p \leq \|X\|_p + \|Y\|_p$ for any X and Y. However, $\|X\|_p = 0$ does not imply $X = 0$ but only that $X = 0$ a.s. Thus, $\|\cdot\|_p$ is a pseudo norm.

If we introduce the equivalence $X \sim Y$ if $\mathbf{P}(X = Y) = 1$, then for $p \geq 1$ only $\|\cdot\|_p$ becomes a genuine norm on the set of equivalence classes of r.v.s for which this quantity is finite (the L^p norm of an equivalence class is just $\|X\|_p$ for any X in the equivalence class). It is known as "L^p-space". With this norm (and the corresponding metric $\|X-Y\|_p$, the space L^p becomes a normed vector space. A non-trivial fact (proof left to measure theory class) is that L^p is a complete under this metric. A normed vector space which is complete under the induced metric is called a *Banach space* and L^p spaces are the prime examples.

The most important are the cases $p = 1, 2, \infty$. In these cases, (we just write X in place of $[X]$)

$$\|X-Y\|_1 := \mathbf{E}[|X-Y|] \qquad \|X-Y\|_2 := \sqrt{\mathbf{E}[|X-Y|^2]} \qquad \|X-Y\|_\infty := \inf\{t \,:\, \mathbf{P}(|X-Y| > t) = 0\}.$$

Exercise 49. For $p = 1, 2, \infty$, check that $\|X-Y\|_p$ is a metric on the space $L^p := \{[X] \,:\, \|X\|_p < \infty\}$ (here $[X]$ denotes the equivalence class of X under the above equivalence relation).

Especially special is the case $p = 2$, in which case the norm comes from an inner product $\langle [X], [Y] \rangle := \mathbf{E}[XY]$. L^2 is a complete inner product space, also known as a *Hilbert space*. For $p \neq 2$, the L^p norm does not come from an inner product as $\|\cdot\|_p$ does *not* satisfy the polarization identity $\|X+Y\|_p^2 + \|X-Y\|_p^2 = 2\|X\|_p^2 + 2\|Y\|_p^2$.

15. SOME INEQUALITIES FOR EXPECTATIONS

The following inequalities are very useful. We start with the very general, but intuitively easy to understand Jensen's inequality. For this we recall two basic facts about convex functions on $\mathbb{R}$.

Let $\phi : (a,b) \to \mathbb{R}$ be a convex function. Then, (i) ϕ is continuous. (ii) Given any $u \in \mathbb{R}$, there is a line in the plane passing through the point $(u, \phi(u))$ such that the line lies below the graph of ϕ. If ϕ is strictly convex, then the only place where the line and the graph of ϕ meet, is at the point $(u, \phi(u))$. Proofs for these facts may be found in many books, eg., Rudin's "Real and Complex Analysis" (ch. 3).

Lemma 50 (Jensen's inequality)**.** *Let $\phi : \mathbb{R} \to \mathbb{R}$ be a convex function. Let X be a r.v on some probability space. Assume that X and $\phi(X)$ both have expectations. Then, $\phi(\mathbf{E}X) \leq \mathbf{E}[\phi(X)]$. The same assertion holds if ϕ is a convex function on some interval (a,b) and X takes values in (a,b) a.s.*

Proof. Let $\mathbf{E}[X] = a$. Let $y = m(x-a) + \phi(a)$ be the 'supporting line' through $(a, \phi(a))$. Since the line lies below the graph of ϕ, we have $m(X-a) + \phi(a) \leq \phi(X)$, a.s. Take expectations to get $\phi(a) \leq E[\phi(X)]$. ■

Lemma 51. *(a) [Cauchy-Schwarz inequality] If X, Y are r.v.s on a probability space, then $\mathbf{E}[XY]^2 \leq \mathbf{E}[X^2]\mathbf{E}[Y^2]$.*

(b) [Hölder's inequality] If X, Y are r.v.s on a probability space, then for any $p, q \geq 1$ satisfying $p^{-1} + q^{-1} = 1$, we have $\|XY\|_1 \leq \|X\|_p \|Y\|_q$.

Proof. Cauchy-Schwarz is a special case of Hölder with $p = q = 2$.

The proof of Hölder inequality follows by applying the inequality $a^p/p + b^q/q \geq ab$ for $a, b \geq 0$ to $a = |X|/\|X\|_p$ and $b = Y/\|Y\|_q$ and taking expectations. The inequality $a^p/p + b^q/q \geq ab$ is evident by noticing that the rectangle $[0,a] \times [0,b]$ (with area ab) is contained in the union of the region$\{(x,y) \,:\, 0 \leq x \leq a,\ 0 \leq y \leq x^{p-1}\}$ (with area a^p/p) and the region $\{(x,y) \,:\, 0 \leq y \leq b,\ 0 \leq x \leq y^{q-1}\}$ (with area b^q/q) simply because the latter regions are the regions between the x and y axes (resp.) and curve $y = x^{p-1}$ which is also the curve $x = y^{q-1}$ since $(p-1)(q-1) = 1$. ■

Lemma 52 (Minkowski's inequality)**.** *For any $p \geq 1$, we have $\|X+Y\|_p \leq \|X\|_p + \|Y\|_p$.*

Proof. For the important cases of $p = 1, 2, \infty$, we know how to check this (for $p = 2$, use Cauchy-Schwarz). For general p, one can get it by applying Hölder to an appropriate pair of functions. We omit details (we might not use them, actually). ■