11. EXPECTATION

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. We define *Expectation* or *Lebesgue integration on measure space* in three steps.

(1) If X can be written as $X = \sum_{i=1}^{n} c_i \mathbf{1}_{A_i}$ for some $A_i \in \mathcal{F}$, we say that X is a *simple r.v.*. We define its *expectation* to be $\mathbf{E}[X] := \sum_{i=1}^{n} c_i \mathbf{P}(A_i)$.
(2) If $X \geq 0$ is a r.v., we define $\mathbf{E}[X] := \sup\{\mathbf{E}[S] \,:\, S \leq X \text{ is a simple, nonegative r.v.}\}$. Then, $0 \leq \mathbf{E}X \leq \infty$.
(3) If X is any r.v. (real-valued!), let $X_+ := X\mathbf{1}_{X \geq 0}$ and $X_- := -X\mathbf{1}_{X<0}$ so that $X = X_+ - X_-$ (also observe that $X_+ + X_- = |X|$). If both $\mathbf{E}[X_+]$ and $\mathbf{E}[X_-]$ are finite, we say that X is integrable (or that $\mathbf{E}[X]$ exists) and define $\mathbf{E}[X] := \mathbf{E}[X_+] - \mathbf{E}[X_-]$.

Naturally, there are some arguments needed to complete these steps.

(1) In the first step, one should check that $\mathbf{E}[X]$ is well-defined, as a simple r.v. can be represented as $\sum_{i=1}^{n} c_i \mathbf{1}_{A_i}$ in many ways. It helps to note that there is a unique way to write it in this form with A_k p.w disjoint. Finite additivity of $\mathbf{P}$ is used here.
(2) In addition, check that the expectation defined in step 1 has the properties of *positivity* ($X \geq 0$ implies $\mathbf{E}[X] \geq 0$) and *linearity* ($\mathbf{E}[\alpha X + \beta Y] = \alpha\mathbf{E}[X] + \beta\mathbf{E}[Y]$).
(3) In step 2, again we would like to check positivity and linearity. It is clear that $\mathbf{E}[\alpha X] = \alpha\mathbf{E}[X]$ if $X \geq 0$ is a r.v and α is a non-negative real number (why?). One can also easily see that $\mathbf{E}[X+Y] \geq \mathbf{E}[X] + \mathbf{E}[Y]$ using the definition. To show that $\mathbf{E}[X+Y] \geq \mathbf{E}[X] + \mathbf{E}[Y]$, one proves using countable additivity of $\mathbf{P}$ -

   Monotone convergence theorem (provisional version). If S_n are non-negative simple r.v.s that increase to X, then $\mathbf{E}[S_n]$ increases to $\mathbf{E}[X]$.

   From this, linearity follows, since $S_n \uparrow X$ and $T_n \uparrow Y$ implies that $S_n + T_n \uparrow X + Y$. One point to check is that there do exist simple r.v S_n, T_n that increase to X, Y. For example, we can take $S_n(\omega) = \sum_{k=0}^{2^{2n}} \frac{k}{2^n} \mathbf{1}_{X(\omega) \in [k2^{-n}, (k+1)2^{-n})}$.

   An additional remark: It is convenient to allow a r.v. to take the value $+\infty$ but adopt the convention that $0 \cdot \infty = 0$ (infinite value on a set of zero probability does not matter).
(4) In step 3, one assumes that both $\mathbf{E}[X_+]$ and $\mathbf{E}[X_-]$ are finite, which is equivalent to assuming that $\mathbf{E}[|X|] < \infty$. In other words, we deal with "absolutely integrable r.v.s" (no "conditionally convergent" stuff for us).

Let us say "$X = Y$ a.s" or "$X < Y$ a.s" etc., to mean that $\mathbf{P}(X = Y) = 1$, $P(X < Y) = 1$ etc. We may also use a.e. (almost everywhere) or w.p.1 (with probability one) in place of a.s (almost surely). To summarize, we end up with an expectation operator that has the following properties.

(1) *Linearity:* X, Y integrable imples $\alpha X + \beta Y$ is also integrable and $\mathbf{E}[\alpha X + \beta Y] = \alpha\mathbf{E}[X] + \beta\mathbf{E}[Y]$.
(2) *Positivity:* $X \geq 0$ implies $\mathbf{E}[X] \geq 0$. Further, if $X \geq 0$ and $\mathbf{P}(X = 0) < 1$, then $\mathbf{E}[X] > 0$. As a consequence, whenever $X \leq Y$ and $\mathbf{E}[X], \mathbf{E}[Y]$ exist, then $\mathbf{E}[X] \leq \mathbf{E}[Y]$ with equality if and only if $X = Y$ a.s.
(3) If X has expectation, then $|\mathbf{E}\, X| \leq \mathbf{E}|X|$.
(4) $\mathbf{E}[\mathbf{1}_A] = \mathbf{P}(A)$, in particular, $\mathbf{E}[1] = 1$.

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 $\mathbf{E}[\liminf X_n] \le \liminf \mathbf{E}[X_n]$.*

Theorem 48 (Dominated convergence theorem (DCT))**.** *Let $|X_n| \le Y$ 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 $Y_n = \inf_{n \ge k} X_n$ 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 $X_n \ge Y_n$ for each n, we get $\liminf \mathbf{E}[X_n] \ge \liminf \mathbf{E}[Y_n] = \mathbf{E}[\liminf X_n]$.

To prove DCT, first note that $|X_n| \le Y$ and $|X| \le Y$ 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

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

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

13. LEBESGUE INTEGRAL VERSUS RIEMANN INTEGRAL

Consider the probability space $([0,1], \mathcal{B}, \mathbf{m})$ and a function $f : [0,1] \to \mathbb{R}$. Let

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

be the upper and lower Riemann sums. Then, $L_n \le U_n$ 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 number 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 measure.

Next consider the Lebesgue integral $\mathbf{E}[f]$. For this we need f to be Borel measurable in the first place. Clearly and 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

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

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.