4 Expectation & the Lebesgue Theorems

Let $X$ and $\{X_n : n \in \mathbb{N}\}$ be random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If $X_n(\omega) \to X(\omega)$ for each $\omega \in \Omega$, does it follow that $\mathbb{E}[X_n] \to \mathbb{E}[X]$? That is, may we exchange expectation and limits in the equation

$$\lim_{n \to \infty} \mathbb{E}[X_n] = \mathbb{E} \left[ \lim_{n \to \infty} X_n \right] ? \quad (1)$$

In general the answer is no; for a simple example take $\Omega = (0, 1]$, the unit interval, with Borel sets $\mathcal{F} = \mathcal{B}(\Omega)$ and Lebesgue measure $\mathbb{P} = \lambda$, and for $n \in \mathbb{N}$ set

$$X_n(\omega) = 2^n 1_{[0,2^{-n}]}(\omega). \quad (2)$$

For each $\omega \in \Omega$, $X_n(\omega) = 0$ for all $n > \log_2(1/\omega)$, so $X_n(\omega) \to 0$ as $n \to \infty$ for every $\omega$, but $\mathbb{E}[X_n] = 1$ for all $n$.

We will want to find conditions that allow us to compute expectations by taking limits, i.e., to force an equality in Equation (1). The two most famous of these conditions are both attributed to Henri Lebesgue: the Monotone Convergence Theorem (MCT) and the Dominated Convergence Theorem (DCT). We will see stronger results later in the course— but let’s look at these now.

4.1 Expectation

Let $\mathcal{E}$ be the linear space of finite-valued $\mathcal{F}$-measurable random variables, and let $\mathcal{E}_+$ be the positive members of $\mathcal{E}$. Each $X \in \mathcal{E}$ may be represented in the form

$$X(\omega) = \sum_{j=1}^{k} a_j 1_{A_j}(\omega)$$

for some $k \in \mathbb{N}$, $\{a_j\} \subset \mathbb{R}$ and $\{A_j\} \subset \mathcal{F}$. The representation is unique if we insist that the $\{a_j\}$ be distinct and that the $\{A_j\}$ be disjoint (why?), in which case $X \in \mathcal{E}_+$ if and only if each $a_j \geq 0$. In general we will not need uniqueness of the representation, so don’t demand that the $\{a_j\}$ be distinct nor that the $\{A_j\}$ be disjoint.
We define the expectation for simple functions in the obvious way:

$$EX = \sum_{j=1}^{k} a_j P(A_j).$$

For this to be a “definition” we must verify that it is well-defined in the sense that it doesn’t depend on the (non-unique) representation; that’s easy. Now we extend the definition of expectation to all non-negative $\mathcal{F}$-measurable random variables as follows:

**Definition 1** The expectation of any nonnegative random variable $Y \geq 0$ on $(\Omega, \mathcal{F}, P)$ is

$$EX = \sup \{EX : X \in \mathcal{E}_+, X \leq Y \}.$$

The expectation can be evaluated using:

**Proposition 1**

$$EY = \lim EX_n$$

for any simple sequence $X_n \in \mathcal{E}_+$ such that $X_n(\omega) \not\succ Y(\omega)$ for each $\omega \in \Omega$.

Proof. First let’s check that such a sequence of simple random variables exists and that the limit makes sense. In a homework exercise you’re asked to prove that

$$X_n = \min \left(2^n, 2^{-n} \lfloor 2^n X \rfloor \right)$$

is simple and nonnegative, and increases monotonically to $Y$. Thus at least one such sequence exists.

By monotonicity the expectations $E[X_n]$ are increasing, so $\lim E[X_n] = \sup E[X_n] \leq \infty$ is just their least upper bound and always exists in $\mathbb{R}_+$. Now let’s show that $EX_n$ for any such sequence converges to $EY$. Fix $\epsilon > 0$ and, by the definition of $EY$, find $X_s \in \mathcal{E}_+$ with $X_s \leq Y$ and $EX_s \geq EY - \epsilon$.

Since $X_s \in \mathcal{E}$ takes only finitely many values, it must be bounded for all $\omega$ by $0 \leq X_s \leq B$ for some $0 < B < \infty$. The events

$$A_n = \{ \omega : X_n(\omega) < X_s(\omega) \}$$

are decreasing (i.e., $A_{n+1} \subset A_n$) and $\cap A_n = \emptyset$ since $\lim X_n(\omega) = Y(\omega) \geq X_s(\omega)$, so $P[A_n] \to 0$. Fix $N_\epsilon$ large enough that $P[A_n] \leq \epsilon / B$. Then

$$EX_n = EX_s + E(X_n - X_s)$$

$$= EX_s + E(X_n - X_s)1_{A_n} + E(X_n - X_s)1_{A_n^c}$$

$$\geq EX_s + E(X_n - X_s)1_{A_n}$$
since \((X_n - X_s) \geq 0\) on \(A_n^c\)

\[ \geq EX_s - BP[A_n] \]

since \((X_n - X_s) \geq -B\) on \(A_n\)

\[ \geq EX_s - \varepsilon \geq EY - 2\varepsilon \]

Thus, since \(X_n \leq Y\) and \(X_n \in \mathcal{E}_+\), for \(n \geq N_{\varepsilon}\)

\[ EX_n \leq EY \leq EX_n + 2\varepsilon \]

and so \(EX_n \rightarrow EY\) as claimed. \(\blacksquare\)

Now that we have \(EX\) well-defined for random variables \(X \geq 0\) we may extend the definition to not-necessarily-non-negative RVs by

\[ EX = EX_+ - EX_- \]

to all random variables for which either of the nonnegative random variables \(X_+ = (X \lor 0), X_- = (-X \lor 0)\) has finite expectation. If both \(EX_+\) and \(EX_-\) are infinite, we must leave \(EX\) undefined. If both are finite, we call \(X\) integrable and note that

\[ |EX| \leq EX_+ + EX_- = E|X| \]

\[ 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots = \sum_{k \in \mathbb{N}} \frac{(-1)^{k+1}}{k} \quad (3) \]

converge? Let’s look closely. First, for any \(p \in \mathbb{R}\) define

\[ S(n) = \sum_{k=1}^{n} k^{-p}\quad I(n) = \int_{1}^{n} x^{-p} \, dx = \begin{cases} \frac{n^{1-p} - 1}{1-p} & p \neq 1 \\ \log n & p = 1. \end{cases} \]

For \(p < 0\) the function \(x^{-p}\) is increasing on \(\mathbb{R}_+\), so \(I(n) + 1 \leq S(n) < I(n+1)\) and so

\[ p < 0 \Rightarrow \frac{n^{1-p} + p}{1-p} \leq \sum_{k=1}^{n} k^{-p} < \frac{(n+1)^{1-p} - 1}{1-p}, \]

\[ \text{Page 3} \quad \text{September 27, 2011} \]
and $S(n) \propto n^{1-p} \to \infty$ as $n \to \infty$.

For $p > 0$ the function $x^{-p}$ is decreasing on $\mathbb{R}_+$, so $I(n+1) < S(n) \leq I(n)+1$ and so

$$p > 0 \Rightarrow \frac{(n+1)^{1-p} - 1}{1 - p} < \sum_{k=1}^{n} k^{-p} \leq \frac{n^{1-p} - p}{1 - p}$$

for $p \neq 1$. For $0 < p < 1$ we again have $S(n) \propto n^{1-p} \to \infty$ as $n \to \infty$, but for $p > 1$ the series converges to some limit $S(\infty) \in (1, p)/\{p - 1\}$. For example, with $p = 2$ we have $S(\infty) = \pi^2/6 \approx 1.644934 \in (1, 2)$. For any $p > 1$ the limit is called the Riemann-zeta function $S(\infty) = \zeta(p)$.

For $p = 1$ we again have divergence, with bounds

$$\log(n+1) < S(n) \leq \log(n) + 1,$$

so the harmonic series $S(n) = \sum_{k=1}^{n} \propto \log n$. In fact $[S(n) - \log n] \to \gamma_e$ converges as $n \to \infty$, to Euler's constant $\gamma_e = 0.577215665$.

Thus in the Lebesgue sense, the alternating series of Equation (3) does not converge, since its negative and positive parts

$$S_-(n) = \sum_{j=1}^{n/2} \frac{1}{2j} \quad \text{and} \quad S_+(n) = \sum_{j=1}^{n/2} \frac{1}{2j - 1}$$

each approach $\infty$ as $n \to \infty$. Notice that the even partial sums are

$$\sum_{k=1}^{2^n} \frac{(-1)^{k+1}}{k} = \left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \left(\frac{1}{5} - \frac{1}{6}\right) + \cdots = \sum_{j=1}^{n} \frac{1}{(2j-1)(2j)},$$
bounded above by \( \pi^2/8 \) for all \( n \) (why?), making the example interesting. More precisely, the difference
\[
\sum_{k=1}^{n} \frac{(-1)^{k+1}}{k} = S_+(n) - S_-(n) = \frac{1}{2} \left[ \log(2n) - \log(n/2) \right] + o(1)
\]
converges to \( \log 2 \) as \( n \to \infty \). What do you think happens with \( \sum_{k=1}^{n} \xi_k/n \), for independent random variables \( \xi_k = \pm 1 \) with probability 1/2 each?

**Theorem 1 (MCT)** Let \( X \) and \( X_n \geq 0 \) be random variables (not necessarily simple) for which \( X_n \not\sim X \). Then
\[
\lim_{n \to \infty} E[X_n] = E[X] = E \left[ \lim_{n \to \infty} X_n \right],
\]
i.e., Equation (1) is satisfied.

For the proof we must find for each \( n \) an approximating sequence \( Y_n^{(m)} \subset \mathcal{E}_+ \) such that \( Y_n^{(m)} \not\sim X_n \) as \( m \to \infty \) and, from it, construct a single sequence
\[
Z_m = \max_{1 \leq n \leq m} Y_n^{(m)} \in \mathcal{E}_+
\]
that satisfies \( Z_m \leq X_m \) for each \( m \) (this is true because, for each \( n \leq m \), \( Y_n^{(m)} \leq X_n \leq X_m \)) and \( Z_m \not\sim X \) as \( m \to \infty \) (to see this, take \( \omega \in \Omega \) and \( \epsilon > 0 \); first find \( n \) such that \( X_n(\omega) \geq X(\omega) - \epsilon \), then find \( m \geq n \) such that \( Y_n^{(m)}(\omega) \geq X_n(\omega) - \epsilon \), and verify that \( Z_m(\omega) \geq X(\omega) - 2\epsilon \), and verify that
\[
\lim_{n \to \infty} E[X_n] \geq \lim_{m \to \infty} E[Z_m] = E[X] \geq E \left[ \lim_{n \to \infty} E[X_n] \right].
\]

**Theorem 2 (Fatou)** Let \( X_n \geq 0 \) be random variables. Then
\[
E \left[ \liminf_{n \to \infty} X_n \right] \leq \liminf_{n \to \infty} E[X_n].
\]
To prove this, just set \( Y_n \equiv \inf_{m \geq n} X_n \) and apply MCT to \( Y_n \leq X_n \):
\[
E \left[ \liminf_{n \to \infty} X_n \right] = E \left[ \lim_{n \to \infty} Y_n \right] = \lim_{n \to \infty} E[Y_n] \leq \liminf_{n \to \infty} E[X_n]
\]
Notice that equality may fail, as in the example of Equation (2). The condition \( X_n \geq 0 \) isn’t entirely superfluous, but it can be weakened to \( X_n \geq Z \) for any integrable random variable \( Z \) (i.e., one with \( E[Z] < \infty \)).

For indicator random variables \( X_n \equiv 1_{A_n} \) of events \( \{A_n\} \), since \( E X_n = P(A_n) \), Fatou’s lemma asserts that
\[
P \left( \liminf_{n \to \infty} A_n \right) \leq \liminf_{n \to \infty} P(A_n) \leq \limsup_{n \to \infty} P(A_n) \leq P \left( \limsup_{n \to \infty} A_n \right)
\]
Finally we have the most important result of this section:

**Theorem 3 (DCT)** Let $X$ and $X_n$ be random variables (not necessarily simple or positive) for which $X_n \to X$, and suppose that $|X_n| \leq Y$ for some integrable random variable $Y$ with $\mathbb{E}Y < \infty$. Then

$$\lim_{n \to \infty} \mathbb{E}[X_n] = \mathbb{E}X = \mathbb{E}\left[\lim_{n \to \infty} X_n\right],$$

i.e., Equation (1) is satisfied if $\{X_n\}$ is “dominated” by $Y \in L_1$. Moreover, we have $\mathbb{E}|X_n - X| \to 0$.

To show this just apply Fatou’s lemma to both $(X_n - X)$ and to $(X - X_n)$; each is bounded below by $-2Y$. For the “moreover” part, apply DCT separately to the positive and negative parts $(X_n - X)_+ \equiv 0 \lor (X_n - X)$ and $(X_n - X)_- \equiv 0 \lor (X - X_n)$; each is dominated by $2Y$ and converges to zero. Then use

$$\mathbb{E}|X_n - X| = \mathbb{E}(X_n - X)_+ + \mathbb{E}(X_n - X)_- \to 0.$$  

We will see later that the condition “$\forall \omega \in \Omega \ X_n(\omega) \to X(\omega)$” can be weakened to “$(\forall \epsilon > 0) \ P[|X_n - X| > \epsilon] \to 0$”.

### 4.2 $L_p$ Spaces and some Expectation Inequalities

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and, for any number $p > 0$, let “$L_p$” (or “$L_p(\Omega, \mathcal{F}, \mathbb{P})$”, pronounced “ell pee”) denote the vector space of real-valued (or sometimes complex-valued) random variables $X$ for which $\mathbb{E}|X|^p < \infty$. Note that this is a vector space, since

- For any $X \in L_p$ and $a \in \mathbb{R}$,
  $$\mathbb{E}|aX|^p = |a|^p \mathbb{E}|X|^p < \infty.$$

- For any $X,Y \in L_p$,
  $$\mathbb{E}|X + Y|^p \leq \mathbb{E}\{\max(|X|,|Y|)^p\} \leq \mathbb{E}\{2\max(|X|,|Y|)^p\} = 2^p \mathbb{E}\{\max(|X|^p,|Y|^p)\} \leq 2^p \mathbb{E}\{|X|^p + |Y|^p\} = 2^p \mathbb{E}|X|^p + \mathbb{E}|Y|^p < \infty,$$

and hence $aX \in L_p$ and $X + Y \in L_p$. By **Minkowski’s Inequality** (see item (7) below), the function

$$\|X\|_p = \left(\mathbb{E}|X|^p\right)^{1/p}$$
is a norm on the space $L_p$ for $p \geq 1$, inducing a metric $d(X,Y) = \|X - Y\|_p$ that obeys the triangle inequality$^1$. We can show that $L_p$ is a complete separable metric space in this metric (what does “complete” mean? Why “separable”? What do we need to show to prove each of these?) For $0 < p < 1$ the space $L_p$ is still a complete separable metric space, but (because $\varphi(x) = |x|^p$ isn’t convex for $p < 1$) $\|X - Y\|_p$ doesn’t satisfy the triangle inequality and so isn’t a metric— but $\|X - Y\|_p = E|X - Y|^p$ is a metric for $0 < p < 1$. By Jensen’s Inequality (see item (5) below) for the convex function $\varphi(x) = |x|^{q/p}$,

$$0 < p < q < \infty \implies \|X\|_p = \{E|X|^p\}^{1/p} \leq \{E|X|^q\}^{1/q} = \|X\|_q$$

and hence $L_p \supset L_q$ for all $0 < p < q < \infty$.

It is common to treat any two random variables $X, Y$ for which $P[X = Y]$ as “equivalent,” and regard $L_p$ not as a space of functions, but rather as a space of equivalence classes of functions where $X \equiv Y$ if and only if $P[X = Y] = 1$. Distances and norms in $L_p$ depend only on the equivalence class. The distinction is only important when we assert the uniqueness of random variables with some specific property.

For example, by Hölder’s Inequality (item (6) below), for each $Y \in L_q$ the linear functional $\ell_Y$ defined on $L_p$ by

$$\ell_Y[X] = E[XY]$$

is continuous if $1 < p < \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$. It happens that these are the only continuous linear functionals on $L_p$, so $L_p$ and $L_q$ are mutually dual Banach spaces and, in particular, $L_2$ is a (self-dual) Hilbert space with inner product $\langle X, Y \rangle = E[XY]$.

Since $\|X\|_p$ is non-decreasing in $p$ for each random variable $X$, we can define

$$\|X\|_\infty = \sup_{p < \infty} \|X\|_p \quad L_\infty = \{X : \|X\|_\infty < \infty\}$$

One can show (it’s a good exercise, you should do it) that $\|X\|_\infty$ finite if and only if $X$ is “essentially bounded,” i.e., if $P[|X| \leq B] = 1$ for some constant $B < \infty$, and that $\|X\|_\infty$ is the l.u.b. of the constants $B$ with this property. The space $L_\infty$ is also a complete metric space, but except in some trivial cases it isn’t separable. Can you prove $L_\infty(\Omega, \mathcal{F}, P)$ isn’t separable for $\Omega = (0, 1]$, $\mathcal{F} = 2\mathcal{B}$, and $P = \lambda$? What if instead $P$ has finite or countable support $\{\omega_j\}$, with $P[\{\omega_j\}] = p_j > 0$, $\sum p_j = 1$?

$^1$Strictly speaking, $d$ is only a metric if we identify any two random variables $X, Y$ with $d(X,Y) = 0$, i.e., if we regard $L_p$ as a space of equivalence classes $[X] = \{Y : \Omega \to \mathbb{R} : P[X \neq Y] = 0\}$ of $p$-integrable functions; see paragraph below.
1. For any $p > 0$, $E|X|^p = \int_0^\infty p x^{p-1} P(|X| > x) dx$ and $E|X|^p < \infty \iff \sum_{n=1}^\infty n^{p-1} P(|X| \geq n) < \infty$. The case $p = 1$ is easiest and most important: if $S \equiv \sum_{n=1}^\infty P(|X| \geq n) < \infty$, then $S \leq E|X| < S+1$. If $X$ takes on only nonnegative integer values then $EX = S$.

2. If $\mu_X$ is the distribution of $X$, and if $f$ is a measurable real-valued function on $\mathbb{R}$, then $E f(X) = \int_\Omega f(X(\omega)) d\mathcal{P} = \int_\mathbb{R} f(x) \mu_X(dx)$ if either side exists. In particular, $\mu = EX = \int x \mu_X(dx)$ and $\sigma^2 = E(X^2) - (EX)^2 = \int (x-\mu)^2 \mu_X(dx)$.

3. Markov’s & Chebychev’s Inequalities: If $\varphi$ is positive and increasing, then $P[|X| \geq u] \leq E[\varphi(|X|)]/\varphi(u)$. In particular $P[|X-\mu| > u] \leq \sigma^2/u^2$, $P[|X| > u] \leq \sigma^2/(u-\mu)^2$, and for any $t > 0$, $P[X > u] \leq M(t) e^{-tu}$.

4. One-Sided Version: $P[X > u] \leq \sigma^2/(2u^2)$.

5. Jensen’s Inequality: Let $\varphi(x)$ be a convex function on $\mathbb{R}$, $X$ an integrable RV. Then $\varphi(EX) \leq E[\varphi(X)]$. Examples: $\varphi(x) = |x|^p$, $p \geq 1$; $\varphi(x) = e^{ax}$; $\varphi(x) = |0 \vee x|$. (Introduce $L_p \supset L_q$). The equality is strict if $\varphi(\cdot)$ is strictly convex and $X$ has a non-degenerate distribution.

6. Hölder’s Inequality: Let $p > 1$ and $q = \frac{p}{p-1}$ (e.g., $p = q = 2$ or $p = 1.01$, $q = 101$). Then $E|XY| \leq [E|X|^p]^{1/p} [E|Y|^q]^{1/q}$.

7. Minkowski’s Inequality: Let $1 \leq p \leq \infty$ and let $X, Y \in L_p(\Omega, \mathcal{F}, P)$. Then the norm $\|X\|_p \equiv (E|X|^p)^{1/p}$ obeys the triangle ineq on $L_p(\Omega, \mathcal{F}, P)$:

$$\|X + Y\|_p \leq (E|X|^p)^{1/p} + (E|Y|^p)^{1/p}$$

(what if $p < 1$?). Pf: $E|X+Y|^p \leq E(|X|+|Y|)|X+Y|^p \leq E|X|^p + E|Y|^p$, then Hölder.

8. Höfling’s Inequality: If $\{X_j\}$ are independent and $(\exists \{a_j, b_j\})$ s.t. $P[a_j \leq X_j \leq b_j] = 1$, then $(\forall c > 0)$, $S_n := \sum_{j=1}^n X_j$ satisfies $P[S_n - ES_n \geq c] \leq \exp \left( -2c^2 / \sum_{j=1}^n [b_j-a_j]^2 \right)$. If $X_j$ are iid and bounded
by $|X_j| \leq 1$, e.g., then $P[|\bar{X}_n - \mu| \geq \epsilon] \leq e^{-n\epsilon^2/2}$. Wassily Hoeffding proved this improvement on Chebychev’s inequality in 1963 at UNC. Follows from Hoeffding’s Lemma: $E[e^{\lambda(X_j - \mu_j)}] \leq \exp(\lambda^2(b_j - a_j)^2/8)$, proved in turn from Jensen’s ineq and Taylor’s thm. The importance is that the bound decreases exponentially, while Chebychev only decreases like a power; the price is that $\{X_j\}$ must be bounded in $L_\infty$, not merely in $L_2$. See also related Azuma’s inequality (1967), Bernstein’s inequality (1937), and Chernoff bounds (1952).

Here’s a proof for the important special case of $X_j = \pm 1$ with probability 1/2 each (and hence $\mu = 0$):

$$P[\bar{X}_n \geq \epsilon] = P[S_n \geq n\epsilon]$$
$$= P[e^{\lambda S_n} \geq e^{n\lambda \epsilon}] \text{ for any } \lambda > 0$$
$$\leq E[e^{\lambda S_n} e^{-n\lambda \epsilon}]$$
$$= \left\{ \frac{1}{2} e^{\lambda} + \frac{1}{2} e^{-\lambda} \right\}^n e^{-n\lambda \epsilon}$$
$$\leq \left\{ e^{\lambda^2/2} \right\}^n e^{-n\lambda \epsilon}$$
$$= \exp(n\lambda^2/2 - n\lambda \epsilon)$$

The exponent is minimized at $\lambda = \epsilon$, so

$$P[\bar{X}_n \geq \epsilon] \leq \exp(n\epsilon^2/2 - n\epsilon \epsilon) = e^{-n\epsilon^2/2}.$$

The general case isn’t much harder, but proving $E[e^{\lambda X}] \leq e^{\lambda^2/2}$ is a bit more delicate. Note the Chebychev bound would only have guaranteed $P[\bar{X}_n \geq \epsilon] \leq 1/n\epsilon^2$, algebraic instead of exponential.