**STA 711: Probability & Measure Theory**
Robert L. Wolpert

# 4 Expectation & the Lebesgue Theorems

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

$$\lim_{n\to\infty} \mathsf{E}\,[X_n] \stackrel{?}{=} \mathsf{E}\left[\lim_{n\to\infty} X_n\right]? \tag{1}$$

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 $\mathsf{P} = \lambda$, and for $n \in \mathbb{N}$ set

$$X_n(\omega) := 2^n\, \mathbf{1}_{(0,2^{-n}]}(\omega). \tag{2}$$

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 $\mathsf{E}[X_n] = 1$ for all $n$.

We will want to find conditions that allow us to compute expectations by taking limits, *i.e.*, to force equality in Eqn (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 two now. First, we have to define "expectation."

## 4.1 Expectation

Let $\mathcal{E}$ be the linear space of real-valued $\mathcal{F}$-measurable random variables taking only finitely-many values (these are called *simple*), and let $\mathcal{E}_+$ be the positive members of $\mathcal{E}$. Each $X \in \mathcal{E}$ may be represented in the form

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

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 $\{A_j\}$ form a partition— *i.e.*, be disjoint with $\Omega = \cup_j A_j$ (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:

$$\mathsf{E}X := \sum_{j=1}^{k} a_j \mathsf{P}[A_j].$$

For this to be a "definition" we must verify that the right-hand side doesn't depend on the (non-unique) representation. That's easy— you should do it.

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 $Y \ge 0$ on $(\Omega, \mathcal{F}, \mathsf{P})$ is*

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

The expectation can be evaluated using:

**Proposition 1**

$$\mathsf{E}Y = \lim_{n\to\infty} \mathsf{E}X_n$$

*for any simple sequence $X_n \in \mathcal{E}_+$ such that $X_n(\omega) \nearrow Y(\omega)$ 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

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

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

By monotonicity the expectations $\mathsf{E}[X_n]$ are increasing, so $\lim \mathsf{E}[X_n] = \sup \mathsf{E}[X_n] \le \infty$ is just their least upper bound and always exists in the extended positive reals $\bar{\mathbb{R}}_+ = [0, \infty]$. Now let's show that $\mathsf{E}X_n$ for *any* such sequence converges to $\mathsf{E}Y$ (note $\mathsf{E}Y$ may be infinite).

Fix any $\lambda < \mathsf{E}Y$ and any $\epsilon > 0$. By the definition of $\mathsf{E}Y$, find $X_* \in \mathcal{E}_+$ with $X_* \le Y$ and $\mathsf{E}X_* \ge \lambda$. Since $X_* \in \mathcal{E}$ takes only finitely many values, it must be bounded for all $\omega$ by $0 \le X_* \le B$ for some $0 < B < \infty$. Because $X_n \le X_{n+1}$ and $X_n(\omega) \to Y(\omega) \ge X_*(\omega)$ as $n \to \infty$ for each $\omega \in \Omega$, the events

$$A_n := \{\omega : \ X_n(\omega) < X_*(\omega) - \epsilon\}$$

are decreasing (*i.e.*, $A_n \supset A_{n+1}$) with empty intersection $\cap A_n = \emptyset$, so $\mathsf{P}[A_n] \to 0$. Fix $N_\epsilon$ large enough that $\mathsf{P}[A_n] \le \epsilon/B$ for all $n \ge N_\epsilon$. Then for $n \ge N_\epsilon$,

$$\begin{aligned}
\mathsf{E}X_n &= \mathsf{E}X_* - \epsilon + \mathsf{E}(X_n - X_* + \epsilon) \\
&= \mathsf{E}X_* - \epsilon + \mathsf{E}(X_n - X_* + \epsilon)\mathbf{1}_{A_n} + \mathsf{E}(X_n - X_* + \epsilon)\mathbf{1}_{A_n^c} \\
&\ge \mathsf{E}X_* - \epsilon + \mathsf{E}(X_n - X_* + \epsilon)\mathbf{1}_{A_n}
\end{aligned}$$

since $(X_n - X_* + \epsilon) \ge 0$ on $A_n^c$ and, since $(X_n + \epsilon)\mathbf{1}_{A_n} \ge 0$,

$$\ge \mathsf{E}X_* - \epsilon - \mathsf{E}X_*\mathbf{1}_{A_n}.$$

Since $X_* \le B$, $\mathsf{E}X_*\mathbf{1}_{A_n} \le B\mathsf{P}[A_n] \le \epsilon$ and, for all $n \ge N_\epsilon$,

$$\mathsf{E}X_n \ge \mathsf{E}X_* - \epsilon - B\,\mathsf{P}[A_n] \ge \mathsf{E}X_* - 2\epsilon \ge \lambda - 2\epsilon.$$

Thus $\sup_n \mathsf{E}X_n \ge \lambda - 2\epsilon$.

Since this is true for all $\epsilon > 0$, also $\sup_n \mathsf{E}X_n \ge \lambda$; since that holds for all $\lambda < \mathsf{E}Y$, $\lim_n \mathsf{E}X_n = \sup_n \mathsf{E}X_n \ge \mathsf{E}Y$ as claimed. □

Now that we have $\mathsf{E}X$ well-defined for random variables $X \ge 0$ we may extend the definition of expectation to all (not necessarily non-negative) RVs $X$ by

$$\mathsf{E}X := \mathsf{E}X_+ - \mathsf{E}X_-$$

as long as *either* of the nonnegative random variables $X_+ := (X \vee 0)$, $X_- := (-X \vee 0)$ has finite expectation. If both $\mathsf{E}X_+$ and $\mathsf{E}X_-$ are infinite, we must leave $\mathsf{E}X$ undefined. If both are finite, call $X$ *integrable* and note that

$$\big|\mathsf{E}X\big| \le \mathsf{E}X_+ + \mathsf{E}X_- = \mathsf{E}|X|.$$

### 4.1.1 Examples

Let $\Omega = \mathbb{N}_0 := \{0, 1, \dots\}$ with $\mathcal{F} = 2^\Omega$ and probability measure determined by

$$\mathsf{P}[\{\omega\}] = 2^{-\omega-1}, \qquad \omega \in \Omega.$$

The random variable $\zeta(\omega) = \omega$ has the geometric distribution $\zeta \sim \mathsf{Ge}(1/2)$ with $\mathsf{P}[\zeta = n] = 2^{-n-1}$, but we'll be interested in the random variables

$$Y(\omega) := 2^\omega \qquad \text{and} \qquad X_n := Y\mathbf{1}_{\{\omega<n\}}.$$

Then $Y \ge 0$ and $X_n \in \mathcal{E}_+$ with $X_n \nearrow Y$ as $n \to \infty$, so

$$\mathsf{E}X_n = \sum_{\omega=0}^{n-1} 2^\omega \mathsf{P}[\{\omega\}] = n/2,$$

and, by Prop. 1,

$$\mathsf{E}Y = \lim_{n\to\infty} \mathsf{E}X_n \qquad = \infty.$$

The distribution of $Y$ has a colorful history. Known as the *St. Petersburg Paradox*, it led to the invention of idea of "utility" in decision theory. The random variable $Z(\omega) := (-2)^\omega$ is well-defined and finite, but does not have an expectation because $\mathsf{E}Z_+ = \mathsf{E}Z_- = \infty$.

### 4.1.2 Properties of Expectation

Expectation is a *linear operation* in the sense that, if $a_1, a_2 \in \mathbb{R}$ are two constants and $X_1, X_2$ are two random variables on $(\Omega, \mathcal{F}, \mathsf{P})$, then

$$\mathsf{E}[a_1 X_1 + a_2 X_2] = a_1 \mathsf{E}[X_1] + a_2 \mathsf{E}[X_2]$$

provided the right-hand side is well-defined (not of the form $\infty - \infty$). It follows that expectation respects monotonicity, in the sense that $X_1 \le X_2 \Rightarrow \mathsf{E}[X_1] \le \mathsf{E}[X_2]$ and, as special cases, that $\big|\mathsf{E}[X]\big| \le \mathsf{E}\big[|X|\big]$ and $X \ge 0 \Rightarrow \mathsf{E}[X] \ge 0$. We will encounter many more identities and inequalities for expectations in Section (5).

Expectation is unaffected by changes on null-sets— if $\mathsf{P}[X \ne Y] = 0$, then $\mathsf{E}X = \mathsf{E}Y$. How would you prove this?

### 4.1.3 A Small Extension

The definition of expectation extends without change to random variables $X$ that take values in the *extended* real numbers $\bar{\mathbb{R}} := [-\infty, \infty]$. Obviously $\mathsf{E}X = +\infty$ if $\mathsf{P}[X = +\infty] > 0$ and $\mathsf{P}[X = -\infty] = 0$, $\mathsf{E}X = -\infty$ if $\mathsf{P}[X = +\infty] = 0$ and $\mathsf{P}[X = -\infty] > 0$, and $\mathsf{E}X$ is undefined if both $\mathsf{P}[X = +\infty] > 0$ and $\mathsf{P}[X = -\infty] > 0$. Otherwise, if $\mathsf{P}[|X| = \infty] = 0$, then $X$ (and any function of $X$) have the same expectation as if $X$ were replaced by the real-valued RV $X^*$ defined to be $X(\omega)$ when $|X(\omega)| < \infty$ and otherwise zero, since then $\mathsf{P}[X \ne X^*] = 0$.

With this extension, we can always consider the expectations of quantities like $\limsup X_n$ and $\liminf X_n$, which might take on the values $\pm\infty$ for some RV sequences $\{X_n\}$.

### 4.1.4 Lebesgue Summability Counterexample

Does the alternating sum

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots = \sum_{k \in \mathbb{N}} \frac{(-1)^{k+1}}{k} \tag{3}$$

converge? Let's look closely— the answer depends on what you mean by "converge." First, define

$$S(n) = \sum_{k=1}^{n} k^{-1} \qquad \log(n) = \int_1^n x^{-1}\, dx.$$

By summing

$$k < x < k+1 \Rightarrow \frac{1}{k+1} < \frac{1}{x} < \frac{1}{k} \Rightarrow \frac{1}{k+1} < \int_k^{k+1} x^{-1} dx < \frac{1}{k},$$

from $k = 1$ to $n-1$, and from $k = 2$ to $n$, note that for all $n \in \mathbb{N}$

$$\log(n+1) < S(n) \le \log(n) + 1.$$

Thus the harmonic series $S(n) = \sum_{k=1}^n k^{-1} \asymp \log n$. In fact $[S(n) - \log n]$ converges as $n \to \infty$ to a finite limit, the Euler-Mascheroni constant $\gamma_e \approx 0.577215665$.

Thus *in the Lebesgue sense*, the alternating series of Eqn (3) does *not* converge, since its negative and positive parts[1]

$$\begin{aligned} S_-(n) &:= \sum_{j=1}^{n/2} \frac{1}{2j} & S_+(n) &:= \sum_{j=1}^{n/2} \frac{1}{2j-1} \\ &= \frac{1}{2} S(n/2) & &= S(n) - \frac{1}{2} S(n/2) \\ &= \frac{1}{2}[\log(n/2) + \gamma_e] + o(1) & &= \frac{1}{2}[\log(2\,n) + \gamma_e] + o(1) \end{aligned}$$

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

$$\sum_{k=1}^{2n} \frac{(-1)^{k+1}}{k} = \left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \left(\frac{1}{5} - \frac{1}{6}\right) + \cdots = \sum_{j=1}^{n} \frac{1}{(2j-1)(2j)},$$

bounded above by $\pi^2/8$ for all $n$ (why?), making the example interesting. More precisely, the difference

$$\sum_{k=1}^{n} \frac{(-1)^{k+1}}{k} = S_+(n) - S_-(n) = \frac{1}{2}\big[\log(2\,n) - \log(n/2)\big] + o(1)$$

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?

## 4.2 Lebesgue's Convergence Theorems

**Theorem 1 (MCT)** *Let $X$ and $X_n \geq 0$ be random variables (not necessarily simple) for which $X_n(\omega) \nearrow X(\omega)$ for each[2] $\omega \in \Omega$. Then*

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

i.e., *Eqn (1) is satisfied. If $\mathsf{E}|X| < \infty$, then also $\mathsf{E}|X_n - X| \to 0$.*

For the proof we must find for each $n$ an approximating sequence $Y_n^{(m)} \subset \mathcal{E}_+$ such that $Y_n^{(m)} \nearrow X_n$ as $m \to \infty$ and, from it, construct a single sequence

$$Z_m := \max_{1 \leq n \leq m} Y_n^{(m)} \in \mathcal{E}_+$$

[1] The "little oh" notation "$o(1)$" means that any remaining terms converge to zero as $n \to \infty$. More generally, "$f = o(g)$" means that $(\forall \epsilon > 0)(\exists N_\epsilon < \infty)(\forall x > N_\epsilon)\ |f(x)| \leq \epsilon g(x)$— roughly, that $f(x)/g(x) \to 0$.

[2] In fact it is enough to assume that $\mathsf{P}[X_n \geq 0] = 1$ and $\mathsf{P}[X_n \nearrow X] = 1$, *i.e.*, that $X_n$ are nonnegative and increase to $X$ outside of a null set $N \in \mathcal{F}$, since $X_n \mathbf{1}_{N^c}$ and $X\mathbf{1}_{N^c}$ have the same expectations as $X_n$ and $X$.

that satisfies $Z_m \le X_m$ for each $m$ (this is true because, for each $n \le m$, $Y_n^{(m)} \le X_n \le X_m$) and $Z_m \nearrow X$ as $m \to \infty$ (to see this, take $\omega \in \Omega$ and $\epsilon > 0$; first find $n$ such that $X_n(\omega) \ge X(\omega) - \epsilon$, then find $m \ge n$ such that $Y_n^{(m)}(\omega) \ge X_n(\omega) - \epsilon$, and verify that $Z_m(\omega) \ge X(\omega) - 2\epsilon$), and verify that

$$\lim_{n\to\infty} \mathsf{E}[X_n] \ge \lim_{m\to\infty} \mathsf{E}[Z_m] = \mathsf{E}X \ge \lim_{n\to\infty} \mathsf{E}[X_n].$$

**Theorem 2 (Fatou's Lemma)** *Let $X_n \ge 0$ be random variables. Then*

$$\mathsf{E}\left[\liminf_{n\to\infty} X_n\right] \le \liminf_{n\to\infty} \mathsf{E}\left[X_n\right].$$

To prove this, just set $Y_n := \inf_{m\ge n} X_m$. Then $Y_n \to Y := \liminf X_n$ by definition, and $\{Y_n\}$ is increasing, so the MCT and the inequality $Y_n \le X_n$ give

$$\mathsf{E}\left[\liminf_{n\to\infty} X_n\right] := \mathsf{E}\left[\lim_{n\to\infty} Y_n\right] = \mathsf{E}\left[Y\right] = \liminf_{n\to\infty} \mathsf{E}\left[Y_n\right] \le \liminf_{n\to\infty} \mathsf{E}\left[X_n\right]$$

Notice that *equality* may fail, as in the example of Eqn (2). The condition $X_n \ge 0$ isn't entirely superfluous, but it can be weakened to $X_n \ge Z$ for any integrable random variable $Z$ (*i.e.*, one with $\mathsf{E}|Z| < \infty$).

For indicator random variables $X_n := \mathbf{1}_{A_n}$ of events $\{A_n\}$, since $\mathsf{E}X_n = \mathsf{P}[A_n]$, Fatou's lemma asserts that

$$\mathsf{P}\left(\liminf_{n\to\infty} A_n\right) \le \liminf_{n\to\infty} \mathsf{P}[A_n] \le \limsup_{n\to\infty} \mathsf{P}[A_n] \le \mathsf{P}\left(\limsup_{n\to\infty} A_n\right)$$

**Corollary 1** *Let $\{X_n\}$, $Z$ be random variables on $(\Omega, \mathcal{F}, \mathsf{P})$ with $X_n \ge Z$ and $\mathsf{E}|Z| < \infty$. Then*

$$\mathsf{E}\left[\liminf_{n\to\infty} X_n\right] \le \liminf_{n\to\infty} \mathsf{E}\left[X_n\right].$$

That is, we may weaken the condition "$X_n \ge 0$" to "$X_n \ge Z \in L_1$" in the statement of Fatou's lemma. To prove this, apply Fatou to $(X_n - Z)$ and add $\mathsf{E}Z$ to both sides.

**Corollary 2** *Let $\{X_n\}$, $Z$ be random variables on $(\Omega, \mathcal{F}, \mathsf{P})$ with $X_n \le Z$ and $\mathsf{E}|Z| < \infty$. Then*

$$\mathsf{E}\left[\limsup_{n\to\infty} X_n\right] \ge \limsup_{n\to\infty} \mathsf{E}\left[X_n\right].$$

To prove this, use the identity $-(\limsup a_n) = \liminf(-a_n)$ (true for any real numbers $\{a_n\}$) and apply Fatou's lemma to the nonnegative sequence $(Z - X_n)$.

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 $\mathsf{P}[X_n \to X] = 1$, and suppose that $\mathsf{P}\big[|X_n| \le Y\big] = 1$ for some integrable random variable $Y$ with $\mathsf{E}Y < \infty$. Then*

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

i.e., *Eqn (1) is satisfied if $\{X_n\}$ is* "dominated" *by $Y \in L_1$. Moreover, $\mathsf{E}|X_n - X| \to 0$.*

**Proof.** To show this just apply Fatou Corollaries 1 and 2 with $Z = -Y$ and $Z = Y$, respectively:

$$\begin{aligned}\mathsf{E}X = \mathsf{E}\,[\liminf X_n] &\le \liminf \mathsf{E}\,[X_n] \\ &\le \limsup \mathsf{E}\,[X_n] \le \mathsf{E}\,[\limsup X_n] = \mathsf{E}X\end{aligned}$$

For the "moreover" part, apply DCT separately to the positive and negative parts of $X$, $(X_n - X)_+ := 0 \vee (X_n - X)$ and $(X_n - X)_- := 0 \vee (X - X_n)$; each is dominated by $2Y$ and converges to zero as $n \to \infty$. Then use

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

□

We will see later that the condition "$\mathsf{P}[X_n \to X] = 1$", known as "almost sure" convergence, can be weakened to *convergence in probability*: "$(\forall \epsilon > 0)\ \mathsf{P}[|X_n - X| > \epsilon] \to 0$." The domination condition in the DCT can be weakened too, and the MCT positivity condition $X_n \ge 0$ can be weakened to $X_n \ge Z$ for some RV $Z$ with $\mathsf{E}|Z| < \infty$.

## Counter-examples

For both examples, let $(\Omega, \mathcal{F}, \mathsf{P})$ be the unit interval $\Omega = (0, 1]$ with the Borel sets and Lebesgue measure.

### Undominated, No convergence

The sequence $\{X_n(\omega) := 2^n \mathbf{1}_{(0,2^{-n}]}(\omega)\}$ of Eqn (2) does not satisfy equation Eqn (1), so there must *not* exist a dominating $Y$ with $|X_n| \le Y$ and $\mathsf{E}|Y| < \infty$. The smallest dominating function is

$$Y := \sup_{n\ge 0} X_n = \sum_{n\ge 0} 2^n \mathbf{1}_{(2^{-n-1},2^{-n}]}$$

whose expectation is

$$\mathsf{E}Y = \sum_{n\ge 0} 2^n (2^{-n} - 2^{-n-1}) = \sum_{n\ge 0} 2^{-1} = \infty.$$

This can also be seen from the relation

$$\frac{1}{2\omega} \le Y < \frac{1}{\omega},$$

so $\mathsf{E}Y \ge \int_0^1 \frac{1}{2\omega}\,d\omega = \infty$.

**Undominated, Convergence**

Now consider the sequence $\{Y_n(\omega) := n\mathbf{1}_{(\frac{1}{n+1},\frac{1}{n}]}$ on the same $(\Omega, \mathcal{F}, \mathsf{P})$. Again there is no dominatation by an integrable RV, since the smallest dominating RV

$$Y := \sup_{n\in\mathbb{N}} Y_n = \sum_{n\in\mathbb{N}} Y_n = \sum_{n\in\mathbb{N}} n\mathbf{1}_{(\frac{1}{n+1},\frac{1}{n}]},$$

has expectation

$$\mathsf{E}Y := \sum_{n\in\mathbb{N}} n\left[\frac{1}{n} - \frac{1}{n+1}\right] = \sum_{n\in\mathbb{N}} \frac{1}{n+1} = \infty.$$

Still, $Y_n(\omega) \to 0$ for every $\omega \in \Omega$ and $\mathsf{E}Y_n = \frac{1}{n+1} \to 0$. This shows that domination is sufficient but not necessary to ensure that equality holds in Eqn (1).