Probability Theory

Robert L. Wolpert
Institute of Statistics and Decision Sciences
Duke University, Durham, NC, USA

1. Lebesgue Theorems

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

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

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

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

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 ω, 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 an equality in Equation (1). The two most famous of these conditions are both attributed to 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.

1.1. Expectation

Let $\mathcal{E}$ be the linear space of finite-valued $\mathcal{F}$-measureable 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 only unique if we insist that the $\{a_j\}$ be distinct and that the $\{A_j\}$ be disjoint, 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 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 $X \geq 0$ on $(\Omega, \mathcal{F}, \mathsf{P})$ is*

$$\mathsf{E}X = \lim \mathsf{E}X_n$$

for any simple sequence $X_n \in \mathcal{E}_+$ such that $X_n(\omega) \nearrow X(\omega)$ for each $\omega \in \Omega$.

For this definition to make sense we have three things to check:

1. For any $X \geq 0$ there exists at least one sequence $\{X_n\} \subset \mathcal{E}_+$ for which $X_n(\omega) \nearrow X(\omega)$ for all $\omega \in \Omega$;

2. For any sequence $\{X_n\} \subset \mathcal{E}_+$ for which $m < n \Rightarrow X_m(\omega) \leq X_n(\omega)$ for all $\omega \in \Omega$, the limit $\lim \mathsf{E}[X_n]$ exists as an extended-real-valued number (*i.e.*, the limit takes values in $\overline{\mathbb{R}}_+ = [0, \infty]$ and in particular might be ∞);

3. If $\{X_n\} \subset \mathcal{E}_+$ and $\{Y_m\} \subset \mathcal{E}_+$ are two sequences that *both* satisfy $X_n \nearrow X$ and $Y_m \nearrow X$, then the limits

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

   coincide.

For the third of these it's useful to first prove

Lemma 1 *Let $Z_n \in \mathcal{E}_+$ for $n \in \mathbb{N}$ and suppose $Z_n \searrow 0$. Then $\mathsf{E}Z_n \to 0$.*

Proof. Since each element in $\mathcal{E}$ is bounded, find K with $Z_1 \le K < \infty$; then for any $\epsilon > 0$, $0 \le \mathsf{E}[Z_n] \le \epsilon \mathsf{P}[Z_n \le \epsilon] + K\mathsf{P}[Z_n > \epsilon] \to 0$. ∎

Item (3) above now follows by symmetry from:

Lemma 2 *If $X_n, Y_m \in \mathcal{E}_+$ are two increasing sequences and if $\lim_{n\to\infty} X_n(\omega) \le \lim_{m\to\infty} Y_m(\omega)$ for each $\omega \in \Omega$, then $\lim_{n\to\infty} \mathsf{E}[X_n] \le \lim_{m\to\infty} \mathsf{E}[Y_m]$.*

Proof. Fix any $n \in \mathbb{N}$; then $(X_n \wedge Y_m) \nearrow X_n$ as $m \to \infty$, since (by hypothesis) $\lim Y_m \ge \lim X_m \ge X_n$. By monotonicity on $\mathcal{E}_+$, using Lemma(1),

$$\mathsf{E}(X_n) = \lim_{m\to\infty} \mathsf{E}(X_n \wedge Y_m) \le \lim_{m\to\infty} \mathsf{E}(Y_m).$$

Now take $n \to \infty$ to find

$$\lim_{n\to\infty} \mathsf{E}(X_n) \le \lim_{m\to\infty} \mathsf{E}(Y_m)$$

as required. ∎

Now that we have $\mathsf{E}X$ well-defined for random variabls $X \ge 0$ we may extend the definition by

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

to all random variables for which *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.

1.1.1. Example

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. First, for any $p \in \mathbb{R}$ define

$$S(n) = \sum_{k=1}^{n} k^{-p} \qquad I(n) = \int_1^n x^{-p}\, dx = \begin{cases} \frac{n^{1-p}-1}{1-p} & p \ne 1 \\ \log n & p = 1 \end{cases}.$$

For $p < 0$ the function x^{-p} is increasing, so $I(n)+1 \le S(n) < I(n+1)$ and so

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

and $S(n) \propto n^{1-p} \to \infty$ as $n \to \infty$.

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

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

for $p \ne 1$. 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

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

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

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

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

each approach ∞ as $n \to \infty$. Notice that the partial sum, the difference

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

does converge (to $\log 2$) as $n \to \infty$... making the example interesting.

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 $X_n \nearrow X$. Then*

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

i.e., Equation (1) is satisfied.

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

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

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

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

Theorem 2 (Fatou) *Let $X_n \geq 0$ be random variables. Then*

$$\mathsf{E}[\liminf_{n\to\infty} X_n] \leq \liminf_{n\to\infty} \mathsf{E}[X_n].$$

Notice that *equality* may fail, as in the example of Equation (2). To prove this, just set $Y_n = \inf_{m\geq n} X_n$ and apply MCT to Y_n. 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. Finally we have

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 $\mathsf{E}Y < \infty$. Then*

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

i.e., Equation (1) is satisfied; moreover, we have $\mathsf{E}|X_n - X| \to 0$.

To show this we 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 $(X_n - X)_+ = 0 \vee (X_n - X)$ and $(X_n - X)_- = 0 \vee (X - X_n)$; each is dominated by $2Y$ and converges to zero. Then use

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

2. Product Spaces

Let $(\Omega_j, \mathcal{F}_j, \mathsf{P}_j)$ be a probability space for $j = 1, 2$ and set

$$\begin{aligned}
\Omega &= \Omega_1 \times \Omega_2 \\
&\equiv \{(\omega_1, \omega_2) : \ \omega_j \in \Omega_j\} \\
\mathcal{F} &= \mathcal{F}_1 \times \mathcal{F}_2 \\
&\equiv \sigma\{A_1 \times A_2 : \ A_j \in \mathcal{F}_j\} \\
\mathsf{P} &= \mathsf{P}_1 \times \mathsf{P}_2, \text{ the unique extension satisfying} \\
\mathsf{P}(A_1 \times A_2) &= P_1(A_1) \cdot P_2(A_2).
\end{aligned}$$

For any $A \in \mathcal{F}$ and $\omega_2 \in \Omega_2$ the (second) *section* of A is

$$A_{\omega_2} = \{\omega_1 : (\omega_1, \omega_2) \in A\} \subset \Omega_1.$$

It's not completely obvious, but one can verify that $A_{\omega_2} \in \mathcal{F}_1$— it's trivial for product sets $A = A_1 \times A_2$, but we need a $\pi - \lambda$ argument to conclude it for all of $\mathcal{F}$. What happens for sets $A \subset \mathcal{F}^{\mathsf{P}}$ in the P-completion of $\mathcal{F}_1 \times \mathcal{F}_2$? Similarly, for any $\mathcal{F}$-measurable random variable $X : \Omega_1 \times \Omega_2 \to \mathcal{S}$ ($\mathcal{S}$ would be $\mathbb{R}$, for real-valued RV's, but could also be $\mathbb{R}^n$ or any metric space), and for any $\omega_2 \in \Omega_2$, the *section* of X is $X_{\omega_2} : \Omega_1 \to \mathcal{S}$ defined by

$$X_{\omega_2}(\omega_1) = X(\omega_1, \omega_2).$$

If $X = 1_A$ is the indicator function of some set $A \in \mathcal{F}$, then the section X_{ω_2} is the indicator function $X_{\omega_2} = 1_{A_{\omega_2}}$ of the section A_{ω_2}. It is (again) perhaps not *quite* obvious, but true, that X_{ω_2} is $\mathcal{F}_1$-measurable. It follows most easily from the same result for sets, upon looking at the set $A = X^{-1}(B) = \{\omega : X(\omega) \in B\}$ for arbitrary $B \in \sigma(\mathcal{S})$ and checking that $A_{\omega_2} = X_{\omega_2}^{-1}(B) = \{\omega : X(\omega) \in B\}$. Is it still true if X is only $\mathcal{F}^{\mathsf{P}}$-measurable?

Finally,

2.1. Fubini

Fubini's theorem gives conditions (namely, that either $X \geq 0$ or $\mathsf{E}|X| < \infty$) to guarantee that these three integrals are meaningful and equal:

$$\int_{\Omega_2} \left\{ \int_{\Omega_1} X_{\omega_2} d\mathsf{P}_1 \right\} d\mathsf{P}_2 \stackrel{?}{=} \iint_{\Omega} X \, d\mathsf{P} \stackrel{?}{=} \int_{\Omega_1} \left\{ \int_{\Omega_2} X_{\omega_1} d\mathsf{P}_2 \right\} d\mathsf{P}_1$$