STA 711: Probability & Measure Theory
Robert L. Wolpert

5 Expectation Inequalities and L_p Spaces

Fix a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ and, for any real number $p > 0$ (not necessarily an integer) and let "L_p" or "$L_p(\Omega, \mathcal{F}, \mathsf{P})$", pronounced "ell pee", denote the vector space of real-valued (or sometimes complex-valued) random variables X for which $\mathsf{E}|X|^p < \infty$. Note that this *is* a vector space, since

- For any $X \in L_p$ and $a \in \mathbb{R}$,

$$\mathsf{E}\big|aX\big|^p = |a|^p \; \mathsf{E}|X|^p < \infty.$$

- For any $X, Y \in L_p$,

$$\begin{aligned} \mathsf{E}|X+Y|^p &\le \mathsf{E}\left\{(|X|+|Y|)^p\right\} \\ &\le \mathsf{E}\left\{2\max(|X|,|Y|)^p\right\} = 2^p \; \mathsf{E}\left\{\max(|X|^p,|Y|^p)\right\} \\ &\le 2^p \; \mathsf{E}\left\{|X|^p + |Y|^p\right\} \quad = 2^p \; \left\{\mathsf{E}|X|^p + \mathsf{E}|Y|^p\right\} < \infty. \end{aligned}$$

and hence $aX \in L_p$ and $X + Y \in L_p$ if $X, Y \in L_p$. By far the two most important cases are $p = 1$ and $p = 2$. A random variable X is called "integrable" if $\mathsf{E}|X| < \infty$ or, equivalently, if $X \in L_1$; it is called "square integrable" if $\mathsf{E}|X|^2 < \infty$ or, equivalently, if $X \in L_2$. Integrable random variables have well-defined means; square-integrable random variables have, in addition, finite variance.

By **Minkowski's Inequality** (see item (7) below), the function

$$\|X\|_p := \left\{\mathsf{E}|X|^p\right\}^{1/p}$$

is a *norm* on the space L_p for $p \ge 1$, inducing a *metric* $\mathrm{d}(X,Y) = \|X - Y\|_p$ that obeys the three rules (for every X, Y, Z):

1. $\mathrm{d}(X,Y) = \mathrm{d}(Y,X)$;
2. $\mathrm{d}(X,Y) = 0$ if and only if $X = Y$;[1]
3. $\mathrm{d}(X,Z) \le \mathrm{d}(X,Y) + \mathrm{d}(Y,Z)$.

[1] Strictly speaking, d is only a *metric* if we identify any two random variables X, Y with $\mathrm{d}(X,Y) = 0$, *i.e.*, if we regard L_p as a space of *equivalence classes* $[X] = \{Y : \Omega \to \mathbb{R} : \; \mathsf{P}[X \ne Y] = 0\}$ of p-integrable random variables; see paragraph below.

including the triangle inequality. 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 $\|X - Y\|_p^p = \mathsf{E}|X - Y|^p$ *is* a metric for $0 < p < 1$, under which L_p is a complete separable metric space. By **Jensen's Inequality** (see item (5) or Theorem 1 below) for the convex function $\varphi(x) = |x|^{q/p}$,

$$0 < p < q < \infty \Rightarrow \|X\|_p = \{\mathsf{E}|X|^p\}^{1/p} \le \{\mathsf{E}|X|^q\}^{1/q} = \|X\|_q$$

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 $\mathsf{P}[X = Y]$ 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 $\mathsf{P}[X = Y] = 1$. 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; what we mean then is uniqueness *up to equivalence*.

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

$$X \mapsto \ell_Y[X] := \mathsf{E}[XY]$$

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) real Hilbert space with inner product $\langle X, Y\rangle = \mathsf{E}[XY]$.

Call a random variable X "essentially bounded" if there exists a finite number $0 \le B < \infty$ such that $\mathsf{P}\big[|X| \le B\big] = 1$, and in that case let

$$\|X\|_\infty := \inf\big\{B \ge 0 :\ \mathsf{P}\big[|X| \le B\big] = 1\big\}$$

denote the *infimum* of the constants B with this property (or $+\infty$ if no such B exists). Since $\|X\|_p$ is non-decreasing in $p \in (0, \infty)$ for each random variable X, the limit of $\|X\|_p$ as $p \to \infty$ always exists, and is identical to the supremum $\sup_{p<\infty} \|X\|_p = \lim_{p\to\infty} \|X\|_p$. One can show (it's a good exercise, you should do it) that this limit is identical to $\|X\|_\infty$, *i.e.*, that

$$\sup_{p<\infty} \|X\|_p = \lim_{p\to\infty} \|X\|_p = \|X\|_\infty$$

The space $L_\infty = \{X :\ \|X\|_\infty < \infty\}$ of essentially bounded random variables is also a complete metric space but, except in some trivial cases, it isn't separable. Can you prove $L_\infty(\Omega, \mathcal{F}, \mathsf{P})$ isn't separable for $\Omega = (0, 1]$, $\mathcal{F} = \mathcal{B}$, and $\mathsf{P} = \lambda$? What if instead P has finite or countable support $\{\omega_j\}$, with $\mathsf{P}[\{\omega_j\}] = p_j > 0$, $\sum p_j = 1$? For $X \sim \mathsf{No}(0, 1)$, what is $\|X\|_\infty$? How about $X \sim \mathsf{Bi}(n, p)$? Or $X \sim \mathsf{Un}(a, b)$?

Theorem 1 (Jensen's Inequality) *Let φ be a convex function on $\mathbb{R}$ and let $X \in L_1$ be integrable. Then*

$$\varphi\big(\mathsf{E}[X]\big) \le \mathsf{E}\big[\varphi(X)\big].$$

The cleanest proof I know of this relies on finding a tangent to the graph of φ at the point $\mu = \mathsf{E}[X]$. To start, note by convexity that for any $a < b < c$, $\varphi(b)$ lies below the value at $x = b$ of the linear function taking the same values as $\varphi(x)$ at $x = a$ and $x = c$:

$$\varphi(b) \le \frac{c-b}{c-a}\varphi(a) + \frac{b-a}{c-a}\varphi(c)$$

Subtracting $\varphi(b)$ and then rearranging terms,

$$0 \le \frac{c-b}{c-a}[\varphi(a) - \varphi(b)] + \frac{b-a}{c-a}[\varphi(c) - \varphi(b)]$$

$$\frac{\varphi(b) - \varphi(a)}{b-a} \le \frac{\varphi(c) - \varphi(b)}{c-b}$$

so any line through $\big(b, \varphi(b)\big)$ with slope λ in the range

$$\phi'(b-) := \sup_{a<b} \frac{\varphi(b) - \varphi(a)}{b-a} \le \lambda \le \inf_{c>b} \frac{\varphi(c) - \varphi(b)}{c-b} =: \phi'(b+)$$

lies below the graph of $\varphi(x)$ (draw a picture). Now let $b = \mu$ and let λ be any number in that interval (this will be the derivative $\lambda = \varphi'(\mu)$ if φ is differentiable at μ, but φ might have a "corner" at μ like $|x|$ does at zero). The line $x \rightsquigarrow \varphi(\mu) + \lambda(x-\mu)$ through $(\mu, \varphi(\mu))$ with slope λ lies below the graph of $\varphi(x)$ and touches the graph at $x = \mu$ (draw it!), so

$$\varphi(\mu) = \mathsf{E}\big[\varphi(\mu) + \lambda(X-\mu)\big] \le \mathsf{E}\big[\varphi(X)\big]$$

as claimed. Notice we didn't have to bound φ above or below, or insist that $\varphi(X) \in L_1$.

A Note on Notation

The distribution μ_X of a real-valued random variable X on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ is completely determined by the Distribution Function $F(x) = \mu_X(-\infty, x] = \mathsf{P}[X \le x]$, and the expectation $\mathsf{E}[g(X)]$ for Borel functions $g : \mathbb{R} \to \mathbb{R}$ has been written in many different ways over the centuries. Some of these include:

$$\begin{aligned}
\mathsf{E}[g(X)] &= \int_\Omega g\big(X(\omega)\big)\,\mathsf{P}(d\omega) = \int_\Omega g\big(X\big)\,d\mathsf{P} \\
&= \int_{\mathbb{R}} g(x)\,\mu_X(dx) \qquad = \int_{\mathbb{R}} g\,d\mu_X \\
&= \int_{\mathbb{R}} g(x)\,F_X(dx) \qquad = \int_{\mathbb{R}} g\,dF_X \qquad = \int_{\mathbb{R}} g(x)\,dF_X(x)
\end{aligned}$$

This last one is "Stieltjes" notation, from an early definition of the Riemann integral of a continuous func. g as $\int_a^b g(x)\,dF_X(x) = \lim_{n\to\infty} \sum_{0\le i<n} g(x_i)[F_X(x_{i+1}) - F_X(x_i)]$, with $x_i = a + i(b-a)/n$. All reduce to $\int g(x) f_X(x)\,dx$ for AC F_X, with $f_X(x) := dF_X(x)/dx = F_X'(x)$.

Miscellaneous Integral Identities and Inequalities

1. If μ_X is the distribution of X, and if g is a measurable real-valued function on $\mathbb{R}$, then $\mathsf{E}g(X) := \int_\Omega g(X(\omega))\,\mathsf{P}(d\omega) = \int_{\mathbb{R}} g(x)\,\mu_X(dx)$ if either side exists. In particular, $\mu := \mathsf{E}X = \int x\,\mu_X(dx)$ and $\sigma^2 := \mathsf{E}(X-\mu)^2 = \int (x-\mu)^2\,\mu_X(dx)$ can be calculated using sums and PMFs if X is discrete, or integrals and pdfs if it's absolutely continuous.

2. For any $p > 0$, $\mathsf{E}|X|^p = \int_0^\infty p\,x^{p-1}\mathsf{P}[|X| > x]\,dx$ and $\mathsf{E}|X|^p < \infty \Leftrightarrow \sum_{n=1}^\infty n^{p-1}\mathsf{P}[|X| > n] < \infty$. The case $p = 1$ is easiest and most important: if $S := \sum_{n=0}^\infty \mathsf{P}[|X| > n] < \infty$, then $\mathsf{E}|X| \le S < \mathsf{E}|X|+1$. If X takes on only nonnegative integer values then $\mathsf{E}X = S$.

3. **Markov's Inequality:** If φ is positive and nondecreasing, then $\mathsf{P}[X \ge u] \le \mathsf{E}[\varphi(X)]/\varphi(u)$. In particular $\mathsf{P}[|X| > u] \le \|X\|_p^p/u^p$, $\mathsf{P}[|X| > u] \le \frac{\sigma^2+\mu^2}{u^2}$, and $(\forall t > 0)$, $\mathsf{P}[X > u] \le M(t)\,e^{-tu}$ for the MGF $M(t) := \mathsf{E}\exp(tX)$.

4. **Chebychev's Inequality:** Applying Markov's inequality to $|x-\mu|^2$ gives Chebychev's Inequality, $\mathsf{P}[|X-\mu| > k\sigma] \le \frac{1}{k^2}$. A one-sided version is also available: $\mathsf{P}[X > u] \le \frac{\sigma^2}{\sigma^2+(u-\mu)^2}$ (Pf: $\mathsf{P}[(X-\mu+t) > (u-\mu+t)] \le ?$; optimize over $t \ge \mu - u$).

5. **Jensen's Inequality:** Let $\varphi(x)$ be a convex function on $\mathbb{R}$, and $X \in L_1(\Omega, \mathcal{F}, \mathsf{P})$. Then $\varphi(\mathsf{E}[X]) \le \mathsf{E}[\varphi(X)]$. Examples: $\varphi(x) = |x|^p$, $p \ge 1$; $\varphi(x) = e^x$; $\varphi(x) = [0 \vee x]$. (Introduce $L_p \supset L_q$). The equality is *strict* if $\varphi(\cdot)$ is strictly convex and X has a non-degenerate distribution. See Theorem 1 on $p.3$ for a proof.

6. **Hölder's Inequality:** Let $r \ge 1$ and $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$. Then $\|XY\|_r \le \|X\|_p\,\|Y\|_q$. (Pf: If $\|\tilde{X}\|_p = \|\tilde{Y}\|_q = 1$, then $|\tilde{X}\tilde{Y}|^r = \exp\{\frac{r}{p}\log|\tilde{X}|^p + \frac{r}{q}\log|\tilde{Y}|^q\} \le \{\frac{r}{p}|\tilde{X}|^p + \frac{r}{q}|\tilde{Y}|^q\}$). The special case of $p = q = 2$, $r = 1$ is the famous:
   Cauchy-Schwartz Inequality: $\mathsf{E}\,XY \le \mathsf{E}\,|XY| \le \sqrt{\mathsf{E}[X^2]\;\mathsf{E}[Y^2]}$.

7. **Minkowski's Inequality:** Let $1 \le p \le \infty$ and let $X, Y \in L_p(\Omega, \mathcal{F}, \mathsf{P})$. Then the norm $\|X\|_p := (\mathsf{E}|X|^p)^{\frac{1}{p}}$ obeys the triangle inequality on $L_p(\Omega, \mathcal{F}, \mathsf{P})$:

$$\|X+Y\|_p \le \|X\|_p + \|Y\|_p$$

   Pf: $\mathsf{E}|X+Y|^p \le \mathsf{E}\big(|X|+|Y|\big)|X+Y|^{p/q}$, then apply Hölder. What if $p < 1$?

6 Independence

6.1 Independent Events

A collection of events $\{A_i\} \subset \mathcal{F}$ in a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ is called *independent* if

$$\mathsf{P}\left[\cap_{i\in I} A_i\right] = \prod_{i\in I} \mathsf{P}[A_i]$$

for each finite set I of indices. This is a stronger requirement than "pairwise independence," the requirement merely that

$$\mathsf{P}[A_i \cap A_j] = \mathsf{P}[A_i]\mathsf{P}[A_j]$$

for each $i \neq j$. For a simple counter-example, toss two fair coins and let H_n be the event "Heads on the nth toss" for $n = 1, 2$. Then the three events $A_1 := H_1$, $A_2 := H_2$, and $A_3 := H_1 \Delta H_2$ (the event that the coins disagree) each have $\mathsf{P}[A_i] = 1/2$ and each pair has $\mathsf{P}[A_i \cap A_j] = (1/2)^2 = 1/4$ for $i \neq j$, but $\cap A_i = \emptyset$ has probability zero and not $(1/2)^3 = 1/8$.

6.1.1 The Borel-Cantelli Lemmas

Lemma 1 (Borel-Cantelli) *Let $\{A_n\}$ be events on some probability space $(\Omega, \mathcal{F}, \mathsf{P})$ that satisfy*

$$\sum_{n=1}^{\infty} \mathsf{P}[A_n] < \infty.$$

Then the event that infinitely-many of the $\{A_n\}$ occur (the lim sup*) has probability zero.*

Proof.

$$\mathsf{P}\Big[\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m\Big] \le \mathsf{P}\Big[\bigcup_{m=n}^{\infty} A_m\Big] \le \sum_{m=n}^{\infty} \mathsf{P}[A_m] \to 0 \quad \text{as } n \to \infty.$$

□

This result does *not* require independence of the $\{A_n\}$, but its partial converse does:

Lemma 2 (Second Borel-Cantelli) *Let $\{A_n\}$ be* ***independent*** *events on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ that satisfy*

$$\sum_{n=1}^{\infty} \mathsf{P}[A_n] = \infty.$$

Then the event that infinitely-many of the $\{A_n\}$ occur (the lim sup*) has probability one.*

Proof. First recall that $1+x \le e^x$ for all real $x \in \mathbb{R}$, positive or not (draw a graph). For each pair of integers $1 \le n \le N < \infty$,

$$\begin{aligned}\mathsf{P}\Big[\bigcap_{m=n}^{N} A_m^c\Big] &= \prod_{m=n}^{N}\big(1-\mathsf{P}[A_m]\big)\\ &\le \prod_{m=n}^{N} e^{-\mathsf{P}[A_m]} = \exp\left(-\sum_{m=n}^{N}\mathsf{P}[A_m]\right)\\ &\to \exp\left(-\sum_{m=n}^{\infty}\mathsf{P}[A_m]\right) = e^{-\infty} = 0\end{aligned}$$

as $N \to \infty$. Thus each $\cap_{m=n}^{\infty} A_m^c$ is a null set, hence so is their union, so

$$\begin{aligned}\mathsf{P}\Big[\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m\Big] &= 1-\mathsf{P}\Big[\bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_m^c\Big]\\ &\ge 1-\sum_{n=1}^{\infty}\mathsf{P}\left(\bigcap_{m=n}^{\infty} A_m^c\right) = 1-0 = 1.\end{aligned}$$

□

Together these two results comprise the

Proposition 1 (Borel's Zero-One Law) *For independent events $\{A_n\}$, the event $A := \limsup A_n$ has probability $\mathsf{P}(A) = 0$ or $\mathsf{P}(A) = 1$, depending on whether the sum $\sum \mathsf{P}(A_n)$ is finite or not.*

6.1.2 B/C Illustration

Here's a little toy example to illustrate the Borel-Cantelli lemmas. Begin with a leather bag containing one gold coin, and $n = 1$.

(a) At nth turn, first add one additional *silver* coin to the bag, then draw one coin at random. Let A_n be the event

$$A_n = \{\text{Draw gold coin on } n\text{th draw}\}.$$

Whichever coin you draw, replace it; increment n; and repeat.

(b) As above— but at nth turn, add n silver coins.

Let γ be the probability that you *ever* draw the gold coin. In each case, is $\gamma = 0$? $\gamma = 1$? or $0 < \gamma < 1$? In latter case, give exact asymptotic expression for γ and numerical estimate to four decimals. Why doesn't $0 < \gamma < 1$ violate Borel's zero-one law (Prop. 1 below)? Can you find γ exactly, perhaps with the help of Mathematica or Maple?

6.2 Independent Classes of Events

Classes $\{\mathcal{C}_i\}$ of events (*e.g.*, π-systems or σ-algebras) are called *independent* if

$$\mathsf{P}\left[\cap_{i\in I} A_i\right] = \prod_{i\in I} \mathsf{P}[A_i]$$

for each finite I whenever each $A_i \in \mathcal{C}_i$. An important tool for simplifying proofs of independence is

Theorem 2 (Basic Criterion) *If classes $\{\mathcal{C}_i\}$ of events are independent and if each $\mathcal{C}_i$ is a π-system, then $\{\sigma(\mathcal{C}_i)\}$ are independent too.*

Proof. Let I be a finite index set with at least $|I| \geq 2$ elements and $\{\mathcal{C}_i\}_{i\in I}$ an independent collection of π-systems. Fix $i \in I$, set $J := I \setminus \{i\}$, and fix $A_j \in \mathcal{C}_j$ for each $j \in J$. Set:

Then
$$\mathcal{L} := \left\{B \in \mathcal{F} : \quad \mathsf{P}\Big[B \cap \bigcap_{j\in J} A_j\Big] = \mathsf{P}[B] \cdot \prod_{j\in J} \mathsf{P}[A_j]\right\}.$$

- $\mathcal{C}_i \subset \mathcal{L}$, by the hypothesis that $\{\mathcal{C}_i\}$ are independent;
- $\Omega \in \mathcal{L}$, by the independence of $\{\mathcal{C}_j\}_{j\in J}$;
- $B \in \mathcal{L} \Rightarrow B^c \in \mathcal{L}$, by a quick computation; and
- $B_n \in \mathcal{L}$ and $\{B_n\}$ disjoint $\Rightarrow \cup B_n \in \mathcal{L}$, another quick computation.

Thus $\mathcal{L}$ is a λ-system containing $\mathcal{C}_i$, and so by Dynkin's π-λ theorem it contains $\sigma(\mathcal{C}_i)$. Thus $\sigma(\mathcal{C}_i)$ and $\{A_j\}_{j\in J}$ are independent for each $\{A_j \in \mathcal{C}_j\}$, so $\{\sigma(\mathcal{C}_i), \{\mathcal{C}_j\}_{j\in J}\}$ are independent π-systems. Repeating the same argument $|I| - 1$ times (or, more elegantly, mathematical induction on the cardinality $|I|$) completes the proof. ∎

6.3 Independent Random Variables

A collection of random variables $\{X_i\}$ on some probability space $(\Omega, \mathcal{F}, \mathsf{P})$ are called *independent* if the σ-algebras $\mathcal{F}_i := \sigma(X_i) = {X_i}^{-1}(\mathcal{B})$ are independent, *i.e.*, if

$$\mathsf{P}\big(\cap_{i\in I}\,[X_i \in B_i]\big) = \prod_{i\in I} \mathsf{P}[X_i \in B_i]$$

for each finite set I of indices and each collection of Borel sets $\{B_i \in \mathcal{B}(\mathbb{R})\}$. By the Basic Criterion it is enough to check that the joint CDFs factor, *i.e.*, that

$$\mathsf{P}\big(\cap_{i\in I}\,[X_i \leq x_i]\big) = \prod_{i\in I} F_i(x_i) \tag{1}$$

for each finite index set I and each $x \in \mathbb{R}^I$, or just for a dense set of such x (Why?).

For finitely-many jointly continuous random variables this is equivalent to requiring that the joint density function factor as the product of marginal density functions (proof: differentiate (1) w.r.t. each x_i), while for finitely-many discrete random variables it's equivalent to the usual factorization criterion for the joint pmf. The present definition goes beyond those two cases, however— for example, it includes the case of a discrete random variable $X \sim \mathsf{Bi}(7, 0.3)$, absolutely continuous $Y \sim \mathsf{Ex}(2.0)$, mixed $Z = (\zeta \wedge 0)$ for $\zeta \sim \mathsf{No}(0,1)$, and discrete continuous C with the Cantor distribution. It also applies to infinite (even uncountable) collections of random variables, where no joint pdf or pmf can exist.

Since $\sigma\big(g(X)\big) \subseteq \sigma(X)$ for any random variable X and Borel function $g(\cdot)$, if $\{X_i\}$ are independent and if $g_i(\cdot)$ are arbitrary Borel functions, it follows that $\{g_i(X_i)\}$ are independent too— and, in particular, that if $X \perp\!\!\!\perp Y$ then $X \perp\!\!\!\perp g(Y)$ for all Borel functions $g(\cdot)$. If X and Y are independent, then so are X^2 and $(Y \vee 0)$, for example, with no need to compute joint pdfs or pmfs or the like.

6.3.1 B/C + Independence Illustration

Let $X \in L_1(\Omega, \mathcal{F}, \mathsf{P})$ and let $\{A_n\} \subset \mathcal{F}$ be events that satisfy $\mathsf{P}[A_n] \to 0$. Does it follow that $X_n := X\mathbf{1}_{A_n}$ converges almost-surely to 0?

If $\sum_n \mathsf{P}[A_n] < 0$, then *yes*— by the Borel-Cantelli lemma,

$$\mathsf{P}[X_n \not\to 0] \le \mathsf{P}[\limsup A_n] = 0,$$

so $X_n \to 0$ *a.s.*

BUT, if $\{\mathsf{P}[A_n]\}$ is not summable, then *a.s.* convergence can fail. For example, if $\{A_n\}$ are independent and $X \equiv 1$, then

$$\mathsf{P}[X_n \not\to 0] = \mathsf{P}[\limsup A_n] = 1,$$

so $\mathsf{P}[X_n \to 0] = 0$. In Week 7 we will find a new sense of convergence called "convergence in probability" and show that $X_n \to 0$ *pr.*