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

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

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 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 equality in Eqn (**??**). 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 nonzero, 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:

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

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.

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 $\mathsf{E}Y$. Fix $\epsilon > 0$ and, by the definition of $\mathsf{E}Y$, find $X_* \in \mathcal{E}_+$ with $X_* \le Y$ and $\mathsf{E}X_* \ge \mathsf{E}Y - \epsilon$. Since $X_* \in \mathcal{E}$ takes only finitely many values, it must be bounded for all ω 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 $\cap A_n = \emptyset$, so $\mathsf{P}[A_n] \to 0$. Fix N_ϵ 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 $X_n + \epsilon \ge 0$,

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

Since $X_* \le B$,

$$\mathsf{E}X_n \ge \mathsf{E}X_* - \epsilon - B\,\mathsf{P}[A_n] \ge \mathsf{E}X_* - 2\epsilon \ge \mathsf{E}Y - 3\epsilon.$$

Thus, since $X_n \le Y$ and $X_n \in \mathcal{E}_+$,

$$\mathsf{E}X_n \le \mathsf{E}Y \le \mathsf{E}X_n + 3\epsilon$$

for every $\epsilon > 0$ and $n \ge N_\epsilon$, so $\mathsf{E}X_n \to \mathsf{E}Y$ as claimed. $\square$

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 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

$$\mathsf{E}[a_1X_1 + a_2X_2] = a_1\mathsf{E}[X_1] + a_2\mathsf{E}[X_2]$$

provided the right-hand side is well-defined (not of the form $\infty - \infty$). It follows that it 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 (**??**).

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?

4.1.2 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 $\{X_n\}$.

4.1.3 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, for any $p \in \mathbb{R}$ and $n \in \mathbb{N}$ define

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

For $p < 0$ the function x^{-p} is increasing on $\mathbb{R}_+$, 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 on $\mathbb{R}_+$, 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 $p \neq 1$. 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

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

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

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

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

[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 $(\forall \epsilon > 0)(\exists N_\epsilon < \infty)(\forall x > N_\epsilon)\ |f(x)| \le \epsilon g(x)$— roughly, that $f(x)/g(x) \to 0$.

each approach ∞ 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 $\omega \in \Omega$. 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 (**??**) is satisfied.*[2] *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

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

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 $Z_m \nearrow X$ 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 $Y_n^{(m)}(\omega) \geq X_n(\omega) - \epsilon$, and verify that $Z_m(\omega) \geq X(\omega) - 2\epsilon$), and verify that

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

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

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

[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.

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 (**??**). 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*

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

i.e., *Eqn (**??**) 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 **??** and **??** with $Z = -Y$ and $Z = Y$, respectively:

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

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 pointwise convergence condition "$(\forall \omega \in \Omega)\ X_n(\omega) \to X(\omega)$" in the statements of both Theorems **??** and **??** can be weakened to *convergence in probability*, "$(\forall \epsilon > 0)\ \mathsf{P}[|X_n - X| > \epsilon] \to 0$."

5 L_p Spaces and some Expectation Inequalities

Fix a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ and, for any number $p > 0$, 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 $X, Y \in L_p$,

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

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 (**??**) 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 $X = Y$;[3]
3. $\mathrm{d}(X,Z) \le \mathrm{d}(X,Y) + \mathrm{d}(Y,Z)$.

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 (**??**) or Theorem **??** 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 (**??**) 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 $\mathsf{P}[|X| \le B] = 1$, 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 infinity 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

[3]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 functions; see paragraph below.

$L_\infty(\Omega,\mathcal{F},\mathsf{P})$ 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 4 (Jensen's Inequality) *Let φ be convex and $X\in L_1$ 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 $\mathsf{P}[|X| > u] \le \|X\|_p/u^p$, $\mathsf{P}[|X| > u] \le (\sigma^2 + \mu^2)u^2$, 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 **??** on $p.$ **??** 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 $p < 1$?

8. **Hoeffding's Inequality:** If $\{X_j\}$ are independent and bounded above and below individually by $(\exists\ \{a_j, b_j\})$ s.t. $\mathsf{P}[a_j \le X_j \le b_j] = 1$, then $(\forall c > 0)$, $S_n := \sum_{j=1}^n X_j$ satisfies $\mathsf{P}[S_n - \mathsf{E}S_n \ge c] \le \exp\big(-2c^2/\sum_1^n |b_j - a_j|^2\big)$. If X_j are iid with $\|X_j\|_\infty \le B$, then $\mathsf{P}[(\bar{X}_n - \mu) \ge \epsilon] \le e^{-n\epsilon^2/2B^2}$. What bound do you get for Bernoulli RVs on $\mathsf{P}[|\bar{X}_n - p| > \epsilon]$? The proof requires independence, so we'll look at it next week.

Last edited: November 20, 2015