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), 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 $X, Y \in L_p$,

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

Thus, for $a \in \mathbb{R}$ and $X, Y \in L_p$, $aX \in L_p$ and $X + Y \in L_p$. 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 finite means; square-integrable random variables have, in addition, finite variance.

By **Minkowski's Inequality** (see item (7) on page 6 below), the function

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

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 following three rules for every X, Y, Z:

1. $\mathrm{d}(X,Y) = \mathrm{d}(Y,X)$ Symmetric;
2. $\mathrm{d}(X,Y) = 0$ if and only if $X = Y$ Antireflexive[1];
3. $\mathrm{d}(X,Z) \le \mathrm{d}(X,Y) + \mathrm{d}(Y,Z)$ 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 every $p \ge 1$. For $0 < p < 1$ the space L_p is still a complete separable metric space but, because

[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.

$\varphi(x) := |x|^p$ isn't convex[2] 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) on page 6 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 and Y are regarded as "equivalent" (written $X \equiv Y$) if and only if $\mathsf{P}[X = Y] = 1$, in which case we treat them as the *same* element of L_p. 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) on page 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 set

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

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 that this limit is identical to $\|X\|_\infty$ (it's a good exercise, you should do it. Start with "Let $0 \le \lambda < \|X\|_\infty$ and set $\Lambda := \{\omega : \ |X| > \lambda\}$. Then what?), *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—that is, there is

[2] A real-valued function $\varphi : \mathcal{D} \to \mathbb{R}$ on a convex domain $\mathcal{D} \subset \mathbb{R}^n$ is called *convex* if, for each $x, y \in \mathcal{D}$ and each number $0 < p < 1$, it satisfies the inequality $\varphi\big(px + qy\big) \le p\varphi(x) + q\varphi(y)$, where $q := 1-p$. This is equivalent to the requirement that the set $E := \{(x, y) : \ x \in \mathcal{D}, \ y \ge \varphi(x)\}$ is a convex set, *i.e.*, contains the segment connecting any two points (x, y) and (x', y'). Any twice-differentiable function on the reals with $\varphi''(x) \ge 0$ is convex.

no countable set $\{\xi_j\} \subset L_\infty$ that is "dense" in the sense that, for every $\epsilon > 0$ and every $X \in L_\infty$, there is some j such that $\|X - \xi_j\|_\infty < \epsilon$. 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 equality is strict *if $\varphi(\cdot)$ is strictly convex and X has a non-degenerate distribution.*

One proof with a nice geometric feel relies on finding a tangent line 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

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

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 shorter proof that works for $\mathbb{R}^n$-valued random variables X begins by noting that $\varphi : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if its domain is a convex set in $\mathbb{R}^n$ and the "epigraph" $E := \{(x, y) : \ y \ge \varphi(x)\}$ is a convex set in $\mathbb{R}^{n+1}$. But that means any average of points

$\big(x, \varphi(x)\big) \in E$ must also lie in E (see Lemma 1). If we take such an average using the distribution measure μ_X of X, we have (all integrals are over $\mathbb{R}^n$):

$$\begin{aligned} \int \big(x, \varphi(x)\big)\, \mu_X(dx) &= \Big(\int x \mu_X(dx), \ \int \varphi(x) \mu_X(dx) \Big) \\ &\in E := \{(x,y) : \ \varphi(x) \le y\} \qquad \Rightarrow \\ \varphi\Big(\int x \mu_X(dx) \Big) &\le \int \varphi(x)\ \mu_X(dx) \qquad \text{or} \\ \varphi\big(\mathsf{E}X\big) &\le \mathsf{E}\varphi(X). \end{aligned}$$

Here's a short technical lemma justifying a claim made above:

Lemma 1 *Let E be a closed non-empty convex Borel set in $\mathbb{R}^n$, and let P be a Borel probability measure on $\mathbb{R}^n$ with $\mathsf{P}(E) = 1$. Then*

$$\mu := \int_E x\, \mathsf{P}(dx) \in E.$$

Proof. Suppose not— *i.e.*, suppose $\mu \notin E$. Let $x^* \in E$ be arbitrary, set $r := \|x^* - \mu\|$, and consider the compact set $E \cap B_r(\mu)$. The continuous function $d(x) := \|x - \mu\|$ attains a strictly positive minimum $d(\nu) = \|\nu - \mu\| = \epsilon > 0$ at some point ν of that compact set, which will also be the minimum distance from μ to the entire convex set E. The hyperplane $\mathcal{H} := \{x \in \mathbb{R}^n : \ (x - \nu) \cdot (\mu - \nu) = 0\}$ through ν and orthogonal to $(\mu - \nu)$ separates μ from E, and every point $x \in E$ satisfies

$$0 \ge (x - \nu) \cdot (\mu - \nu)$$

(else some point on $[x, \nu] \subset E$ is closer to μ than ν is). Integrating wrt P over E,

$$\begin{aligned} 0 &\ge \int_E (x - \nu) \cdot (\mu - \nu)\, \mathsf{P}(dx) \\ &= (\mu - \nu) \cdot (\mu - \nu) \\ &= \epsilon^2 > 0, \end{aligned}$$

a contradiction. Thus $\mu \in E$. □

.

A Note on Notation

The distribution μ_X (or "$\mu_X(dx)$") of a real-valued random variable X on $(\Omega, \mathcal{F}, \mathsf{P})$ can be specified by giving $\{\mu_X(B) = \mathsf{P}[X \in B]\}$ for all Borel sets $B \subset \mathbb{R}$ or, by Dynkin's

Theorem, just all sets B in a π-system generating the Borel sets. Since $\{(-\infty, x] : x \in \mathbb{R}\}$ is such a π-system, a distribution μ_X can be specified just by giving its Distribution Function $F(x) := \mathsf{P}[X \le x] = \mu_X(-\infty, x]$ for all x.

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 $\mathsf{P}[X \ge u] \le \mathsf{P}[\varphi(X) \ge \varphi(u)] \le \mathsf{E}[\varphi(X)]/\varphi(u)$. In particular, for any $u > 0$, $\mathsf{P}[|X| > u] \le \|X\|_p^p/u^p$, $\mathsf{P}[|X| > u] \le \frac{\sigma^2+\mu^2}{u^2}$, 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**[3]**:** 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:**[2] Let $1 < p < \infty$ 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:
$$\begin{aligned}
\mathsf{E}|X+Y|^p &\le \mathsf{E}\Big[\big(|X| + |Y|\big)\ |X+Y|^{p-1=p/q}\Big] \quad \text{(Triangle)}\\
&\le (\|X\|_p + \|Y\|_p)\,\|\ |X+Y|^{p/q}\ \|_q \quad \text{(Hölder)}\\
&= (\|X\|_p + \|Y\|_p)\,\big(\mathsf{E}\ |X+Y|^p\ \big)^{1/q=1-1/p}\\
\big(\mathsf{E}\ |X+Y|^p\ \big)^{1/p} &\le (\|X\|_p + \|Y\|_p).
\end{aligned}$$

[3]In HW07 you will show that Hölder's and Minkowski's Inequalities also hold for $p = 1$ and $p = \infty$.

6 Independence

Typical undergraduate probability courses present "independence" for finitely many events, discrete RVs, and absolutely continuous RVs. Here we present those as special cases of the concept of independence for any number (even uncountably many) σ-algebras .

6.1 Independent Events

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

$$\mathsf{P}\left[\cap_{i \in I} A_i\right] = \prod_{i \in I} \mathsf{P}[A_i] \tag{1}$$

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 $(1/2)^3 = 1/8$. Note (1) also holds for *countable* index sets I as well, by the monotone convergence theorem. Can you prove that?

6.2 The Borel-Cantelli Lemmas

Our proof below of the Strong Law of Large Numbers for iid bounded random variables relies on the almost-trivial but very useful:

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 ($\limsup_{n\to\infty} A_n$) has probability zero.

Proof.

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

□

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 $1 \le n \le N < \infty$, by independence,

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

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

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

□

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.*

6.2.1 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?

6.3 Independent Classes of Events

6.3.1 Arbitrary Independent Classes

Classes $\{\mathcal{C}_i\}$ of events (*e.g.*, π-systems or σ-algebras) in a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ are called *independent* if

$$\mathsf{P}\Big[\bigcap_{i\in I} A_i\Big] = \prod_{i\in I} \mathsf{P}[A_i]$$

for each finite I whenever each $A_i \in \mathcal{C}_i$. Note the requirement is only for *finite* intersections, and the definition still applies even for uncountable collections $\{\mathcal{C}_i\}$.

6.3.2 Independent σ-Algebras

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

- $\mathcal{C}_i \subset \mathcal{L}$, 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. ◻

6.4 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 $\mathcal{F}_i := \sigma(X_i) = X_i{}^{-1}(\mathcal{B})$ they generate are independent, *i.e.*, if

$$\mathsf{P}\Big(\bigcap_{i\in I}\ [X_i \in B_i]\Big) = \prod_{i\in I} \mathsf{P}[X_i \in B_i]$$

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

$$\mathsf{P}\Big(\bigcap_{i\in I}[X_i \le x_i]\Big) = \prod_{i\in I} F_i(x_i) \tag{2}$$

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 **absolutely continuous** random variables this is equivalent to requiring that the joint density function factor as the product of marginal density functions (proof: differentiate (2) 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.

Independence is a property of the probability measure and the *σ-algebras* $\{\sigma(X_i)\}$, not of the random variables $\{X_j\}$ themselves. 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.

6.4.1 Independent Events Revisited

Arbitrarily many events $\{E_i\}$, random variables $\{X_j\}$, and classes of events $\{\mathcal{C}_k\}$ are independent if and only if the σ-algebras they generate $\{\sigma(E_i),\ \sigma(X_j),\ \sigma(\mathcal{C}_k)\}$ are independent— we can treat all of these in the same unified way.