# STA 711: Probability \& Measure Theory 

Robert L. Wolpert

## 5 Expectation Inequalities and $L_{p}$ Spaces

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

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

$$
\mathrm{E}|a X|^{p}=|a|^{p} \mathrm{E}|X|^{p}<\infty
$$

- For any $X, Y \in L_{p}$,

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

and hence, for $a \in \mathbb{R}$ and $X, Y \in L_{p}, a X \in L_{p}$ 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 $\mathrm{E}|X|<\infty$ or, equivalently, if $X \in L_{1}$; it is called "square integrable" if $\mathrm{E}|X|^{2}<\infty$ 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) below), the function

$$
\|X\|_{p}:=\left\{\mathrm{E}|X|^{p}\right\}^{1 / p}
$$

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

$$
\begin{array}{lll}
\text { 1. } & \mathrm{d}(X, Y)=\mathrm{d}(Y, X) & \text { Symmetric; } \\
\text { 2. } & \mathrm{d}(X, Y)=0 \text { if and only if } X=Y & \text { Antireflexive }^{1} ; \\
3 . & \mathrm{d}(X, Z) \leq \mathrm{d}(X, Y)+\mathrm{d}(Y, Z) & \text { Triangle inequality. }
\end{array}
$$

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 $p \geq 1$. For $0<p<1$ the space $L_{p}$ is still a complete separable metric space but, because

[^0]$\varphi(x):=|x|^{p}$ isn't convex for $p<1$, " $\|X-Y\|_{p}$ " doesn't satisfy the triangle inequality and so isn't a metric- but $\|X-Y\|_{p}^{p}=\mathrm{E}|X-Y|^{p}$ 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 $\varphi(x)=|x|^{q / p}$,
$$
0<p<q<\infty \Rightarrow\|X\|_{p}=\left\{\mathrm{E}|X|^{p}\right\}^{1 / p} \leq\left\{\mathrm{E}|X|^{q}\right\}^{1 / q}=\|X\|_{q}
$$
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 $\mathrm{P}[X=Y]$ 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 $\mathrm{P}[X=Y]=1$, 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) below), for each $Y \in L_{q}$ the linear functional $\ell_{Y}$ defined on $L_{p}$ by

$$
X \mapsto \ell_{Y}[X]:=\mathrm{E}[X Y]
$$

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 $\langle X, Y\rangle=\mathrm{E}[X Y]$.

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

$$
\|X\|_{\infty}:=\inf \{B \geq 0: \mathrm{P}[|X| \leq B]=1\}
$$

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 $p \rightarrow \infty$ always exists, and is identical to the supremum $\sup _{p<\infty}\|X\|_{p}=\lim _{p \rightarrow \infty}\|X\|_{p}$. One can show that this limit is identical to $\|X\|_{\infty}$ (it's a good exercise, you should do it. Start with "Let $0 \leq \lambda<\|X\|_{\infty}$ and set $\Lambda:=\{\omega:|X|>\lambda\}$. Then what?), i.e., that

$$
\sup _{p<\infty}\|X\|_{p}=\lim _{p \rightarrow \infty}\|X\|_{p}=\|X\|_{\infty}
$$

The space $L_{\infty}:=\left\{X:\|X\|_{\infty}<\infty\right\}$ 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 no countable set $\left\{\xi_{j}\right\} \subset L_{\infty}$ that is "dense" in the sense that, for every $\epsilon>0$ and every $X \in L_{\infty}$, there is some $j$ such that $\left\|X-\xi_{j}\right\|_{\infty}<\epsilon$. Can you prove $L_{\infty}(\Omega, \mathcal{F}, \mathrm{P})$ isn't separable for $\Omega=(0,1], \mathcal{F}=\mathcal{B}$, and $\mathrm{P}=\lambda$ ? What if instead P has finite or countable support $\left\{\omega_{j}\right\}$, with $\mathrm{P}\left[\left\{\omega_{j}\right\}\right]=p_{j}>0, \sum p_{j}=1$ ? For $X \sim \operatorname{No}(0,1)$, what is $\|X\|_{\infty}$ ? How about $X \sim \operatorname{Bi}(n, p)$ ? Or $X \sim \operatorname{Un}(a, b)$ ?

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

$$
\varphi(\mathrm{E}[X]) \leq \mathrm{E}[\varphi(X)]
$$

One proof with a nice geometric feel relies on finding a tangent line to the graph of $\varphi$ at the point $\mu=\mathrm{E}[X]$. 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 $x=c$ :

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

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

$$
\begin{gathered}
0 \leq \frac{c-b}{c-a}[\varphi(a)-\varphi(b)]+\frac{b-a}{c-a}[\varphi(c)-\varphi(b)] \\
\frac{\varphi(b)-\varphi(a)}{b-a} \leq \frac{\varphi(c)-\varphi(b)}{c-b}
\end{gathered}
$$

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

$$
\phi^{\prime}(b-):=\sup _{a<b} \frac{\varphi(b)-\varphi(a)}{b-a} \leq \lambda \leq \inf _{c>b} \frac{\varphi(c)-\varphi(b)}{c-b}=: \phi^{\prime}(b+)
$$

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

$$
\varphi(\mu)=\mathrm{E}[\varphi(\mu)+\lambda(X-\mu)] \leq \mathrm{E}[\varphi(X)]
$$

as claimed. Notice we didn't have to bound $\varphi$ 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 $\phi: \mathbb{R}^{n} \rightarrow \mathbb{R}$ is convex if and only if its domain is a convex set in $\mathbb{R}^{n}$ and the "epigraph" $E:=\{(x, y): y \geq \phi(x)\}$ is a convex set in $\mathbb{R}^{n+1}$. But that means any average of points $(x, \phi(x)) \in E$ must also lie in $E$ (see Lemma 1). If we take such an average using the distribution measure $\mu_{X}$ of $X$, we have (all integrals are over $\mathbb{R}^{n}$ ):

$$
\begin{aligned}
\int(x, \phi(x)) \mu_{X}(d x) & =\left(\int x \mu_{X}(d x), \quad \int \phi(x) \mu_{X}(d x)\right) \\
& \in E:=\{(x, y): y \geq \phi(x)\} \quad \Rightarrow \\
\phi\left(\int x \mu_{X}(d x)\right) & \leq \int \phi(x) \mu_{X}(d x) \quad \text { or } \\
\phi(\mathrm{E} X) & \leq \mathrm{E} \phi(X) .
\end{aligned}
$$

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 $\mathrm{P}(E)=1$. Then

$$
\mu:=\int_{E} x \mathrm{P}(d x) \in E .
$$

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

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

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

$$
\begin{aligned}
0 & \geq \int_{E}(x-\nu) \cdot(\mu-\nu) \mathrm{P}(d x) \\
& =(\mu-\nu) \cdot(\mu-\nu) \\
& =\epsilon^{2}>0,
\end{aligned}
$$

a contradiction. Thus $\mu \in E$.

## A Note on Notation

The distribution $\mu_{X}$ (or " $\mu_{X}(d x)$ ") of a real-valued random variable $X$ on $(\Omega, \mathcal{F}, \mathrm{P})$ can be specified by giving $\left\{\mu_{X}(B)=\mathrm{P}[X \in B]\right\}$ for all Borel sets $B \subset \mathbb{R}$ or, by Dynkin's Theorem, just all sets $B$ in a $\pi$-system generating the Borel sets. Since $\{(-\infty, x]: x \in \mathbb{R}\}$ is such a $\pi$-system, a distribution $\mu_{X}$ can be specified just by giving its Distribution Function $F(x):=\mathrm{P}[X \leq x]=\mu_{X}(-\infty, x]$ for all $x$.

The expectation $\mathrm{E}[g(X)]$ for Borel functions $g: \mathbb{R} \rightarrow \mathbb{R}$ has been written in many different ways over the centuries. Some of these include:

$$
\left.\begin{array}{rl}
\mathrm{E}[g(X)] & =\int_{\Omega} g(X(\omega)) \mathrm{P}(d \omega) \\
=\int_{\Omega} g(X) d \mathrm{P} \\
& =\int_{\mathbb{R}} g(x) \mu_{X}(d x) \\
=\int_{\mathbb{R}} g d \mu_{X} \\
& =\int_{\mathbb{R}} g(x) F_{X}(d x)
\end{array}=\int_{\mathbb{R}} g d F_{X} \quad=\int_{\mathbb{R}} g(x) d F_{X}(x)\right)
$$

This last one is "Stieltjes" notation, from an early definition of the Riemann integral of a continuous func. $g$ as $\int_{a}^{b} g(x) d F_{X}(x)=\lim _{n \rightarrow \infty} \sum_{0 \leq i<n} g\left(x_{i}\right)\left[F_{X}\left(x_{i+1}\right)-F_{X}\left(x_{i}\right)\right]$, with $x_{i}=$ $a+i(b-a) / n$. All reduce to $\int g(x) f_{X}(x) d x$ for AC $F_{X}$, with $f_{X}(x):=d F_{X}(x) / d x=F_{X}^{\prime}(x)$.

## Miscellaneous Integral Identities and Inequalities

1. If $\mu_{X}$ is the distribution of $X$, and if $g$ is a measurable real-valued function on $\mathbb{R}$, then $\mathrm{E} g(X):=\int_{\Omega} g(X(\omega)) \mathrm{P}(d \omega)=\int_{\mathbb{R}} g(x) \mu_{X}(d x)$ if either side exists. In particular, $\mu:=\mathrm{E} X=\int x \mu_{X}(d x)$ and $\sigma^{2}:=\mathrm{E}(X-\mu)^{2}=\int(x-\mu)^{2} \mu_{X}(d x)$ can be calculated using sums and PMFs if $X$ is discrete, or integrals and pdfs if it's absolutely continuous.
2. For any $p>0, \mathrm{E}|X|^{p}=\int_{0}^{\infty} p x^{p-1} \mathrm{P}[|X|>x] d x$ and $\mathrm{E}|X|^{p}<\infty \Leftrightarrow \sum_{n=1}^{\infty} n^{p-1} \mathrm{P}[|X|>$ $n]<\infty$. The case $p=1$ is easiest and most important: if $S:=\sum_{n=0}^{\infty} \mathrm{P}[|X|>n]<\infty$, then $\mathrm{E}|X| \leq S<\mathrm{E}|X|+1$. If $X$ takes on only nonnegative integer values then $\mathrm{E} X=S$.
3. Markov's Inequality: If $\varphi$ is positive and nondecreasing, then $\mathrm{P}[X \geq u] \leq \mathrm{P}[\varphi(X) \geq \varphi(u)] \leq \mathrm{E}[\varphi(X)] / \varphi(u)$. In particular, for any $u>0$, $\mathrm{P}[|X|>u] \leq\|X\|_{p}^{p} / u^{p}, \mathrm{P}[|X|>u] \leq \frac{\sigma^{2}+\mu^{2}}{u^{2}}$, and $(\forall t>0), \mathrm{P}[X>u] \leq M(t) e^{-t u}$ for the MGF $M(t):=\mathrm{E} \exp (t X)$.
4. Chebychev's Inequality: Applying Markov's inequality to $|x-\mu|^{2}$ gives Chebychev's Inequality, $\mathrm{P}[|X-\mu|>k \sigma] \leq \frac{1}{k^{2}}$. A one-sided version is also available: $\mathrm{P}[X>u] \leq$ $\frac{\sigma^{2}}{\sigma^{2}+(u-\mu)^{2}}$ (Pf: $\mathrm{P}[(X-\mu+t)>(u-\mu+t)] \leq$ ?; optimize over $\left.t \geq \mu-u\right)$.
5. Jensen's Inequality: Let $\varphi(x)$ be a convex function on $\mathbb{R}$, and $X \in L_{1}(\Omega, \mathcal{F}, \mathbf{P})$. Then $\varphi(\mathrm{E}[X]) \leq \mathrm{E}[\varphi(X)]$. Examples: $\varphi(x)=|x|^{p}, p \geq 1 ; \varphi(x)=e^{x} ; \varphi(x)=[0 \vee x]$. (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 ${ }^{2}$ : Let $r \geq 1$ and $p, q>1$ with $\frac{1}{p}+\frac{1}{q}=\frac{1}{r}$. Then $\|X Y\|_{r} \leq$ $\|X\|_{\tilde{p}}\|Y\|_{q}$. (Pf: If $\|\tilde{X}\|_{p}=\|\tilde{Y}\|_{q}=1$, then $|\tilde{X} \tilde{Y}|^{r}=\exp \left\{\frac{r}{p} \log |\tilde{X}|^{p}+\frac{r}{q} \log |\tilde{Y}|^{q}\right\} \leq$ $\left.\left\{\frac{r}{p}|\tilde{X}|^{p}+\frac{r}{q}|\tilde{Y}|^{q}\right\}\right)$. The special case of $p=q=2, r=1$ is the famous:
Cauchy-Schwartz Inequality: $\mathrm{E} X Y \leq \mathrm{E}|X Y| \leq \sqrt{\mathrm{E}\left[X^{2}\right] \mathrm{E}\left[Y^{2}\right]}$.
7. Minkowski's Inequality: ${ }^{2}$ Let $1<p<\infty$ and let $X, Y \in L_{p}(\Omega, \mathcal{F}, \mathrm{P})$. Then the norm $\|X\|_{p}:=\left(\mathrm{E}|X|^{p}\right)^{\frac{1}{p}}$ obeys the triangle inequality on $L_{p}(\Omega, \mathcal{F}, \mathrm{P})$ :

$$
\|X+Y\|_{p} \leq\|X\|_{p}+\|Y\|_{p}
$$

Pf:

$$
\begin{aligned}
\mathrm{E}|X+Y|^{p} & \leq \mathrm{E}\left[(|X|+|Y|)|X+Y|^{p-1=p / q}\right] \quad \text { (Triangle) } \\
& \leq\left(\|X\|_{p}+\|Y\|_{p}\right)\left\||X+Y|^{p / q}\right\|_{q} \quad \text { (Hölder) } \\
& =\left(\|X\|_{p}+\|Y\|_{p}\right)\left(\mathrm{E}|X+Y|^{p}\right)^{1 / q=1-1 / p} \\
\left(\mathrm{E}|X+Y|^{p}\right)^{1 / p} & \leq\left(\|X\|_{p}+\|Y\|_{p}\right) .
\end{aligned}
$$

[^1]
## 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) $\sigma$-algebras .

### 6.1 Independent Events

A collection (finite, countable, or uncountable) of events $\left\{A_{i}\right\} \subset \mathcal{F}$ in a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ is called independent if

$$
\mathrm{P}\left[\cap_{i \in I} A_{i}\right]=\prod_{i \in I} \mathrm{P}\left[A_{i}\right]
$$

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

$$
\mathrm{P}\left[A_{i} \cap A_{j}\right]=\mathrm{P}\left[A_{i}\right] \mathrm{P}\left[A_{j}\right]
$$

for each $i \neq j$. For a simple counter-example, toss two fair coins and let $H_{n}$ be the event "Heads on the $n$th 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 $\mathrm{P}\left[A_{i}\right]=1 / 2$ and each pair has $\mathrm{P}\left[A_{i} \cap A_{j}\right]=(1 / 2)^{2}=1 / 4$ for $i \neq j$, but $\cap A_{i}=\emptyset$ has probability zero and not $(1 / 2)^{3}=1 / 8$.

### 6.2 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 $\left\{A_{n}\right\}$ be events on some probability space $(\Omega, \mathcal{F}, \mathrm{P})$ that satisfy

$$
\sum_{n=1}^{\infty} \mathrm{P}\left[A_{n}\right]<\infty
$$

Then the event that infinitely-many of the $\left\{A_{n}\right\}$ occur $\left(\limsup _{n \rightarrow \infty} A_{n}\right)$ has probability zero.
Proof.

$$
\mathrm{P}\left[\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_{m}\right] \leq \mathrm{P}\left[\bigcup_{m=n}^{\infty} A_{m}\right] \leq \sum_{m=n}^{\infty} \mathrm{P}\left[A_{m}\right] \rightarrow 0 \quad \text { as } n \rightarrow \infty
$$

This result does not require independence of the $\left\{A_{n}\right\}$, but its partial converse does:

Lemma 2 (Second Borel-Cantelli) Let $\left\{A_{n}\right\}$ be independent events on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ that satisfy

$$
\sum_{n=1}^{\infty} \mathrm{P}\left[A_{n}\right]=\infty
$$

Then the event that infinitely-many of the $\left\{A_{n}\right\}$ occur (the limsup) has probability one.
Proof. First recall that $1+x \leq e^{x}$ for all real $x \in \mathbb{R}$, positive or not (draw a graph). For each pair of integers $1 \leq n \leq N<\infty$, by independence,

$$
\begin{aligned}
\mathrm{P}\left[\bigcap_{m=n}^{N} A_{m}^{c}\right] & =\prod_{m=n}^{N}\left(1-\mathrm{P}\left[A_{m}\right]\right) \\
& \leq \prod_{m=n}^{N} e^{-\mathrm{P}\left[A_{m}\right]}=\exp \left(-\sum_{m=n}^{N} \mathrm{P}\left[A_{m}\right]\right) \\
& \rightarrow \exp \left(-\sum_{m=n}^{\infty} \mathrm{P}\left[A_{m}\right]\right)=e^{-\infty}=0
\end{aligned}
$$

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

$$
\begin{aligned}
\mathrm{P}\left[\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_{m}\right] & =1-\mathrm{P}\left[\bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_{m}^{c}\right] \\
& \geq 1-\sum_{n=1}^{\infty} \mathrm{P}\left[\bigcap_{m=n}^{\infty} A_{m}^{c}\right]=1-0=1 .
\end{aligned}
$$

Together these two results comprise the
Proposition 1 (Borel's Zero-One Law) For independent events $\left\{A_{n}\right\}$, the event $A:=$ $\lim \sup A_{n}$ has probability $\mathrm{P}(A)=0$ or $\mathrm{P}(A)=1$, depending on whether the sum $\sum \mathrm{P}\left(A_{n}\right)$ 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 $n$th 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 $n$th turn, add $n$ silver coins.

Let $\gamma$ 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 $\gamma$ and numerical estimate to four decimals. Why doesn't $0<\gamma<1$ violate Borel's zero-one law (Prop. 1 below)? Can you find $\gamma$ exactly, perhaps with the help of Mathematica or Maple?

### 6.3 Independent Classes of Events

### 6.3.1 Arbitrary Independent Classes

Classes $\left\{\mathcal{C}_{i}\right\}$ of events (e.g., $\pi$-systems or $\sigma$-algebras) in a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ are called independent if

$$
\mathrm{P}\left[\bigcap_{i \in I} A_{i}\right]=\prod_{i \in I} \mathrm{P}\left[A_{i}\right]
$$

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 $\left\{\mathcal{C}_{i}\right\}$.

### 6.3.2 Independent $\sigma$-Algebras

An important tool for simplifying proofs of independence is
Theorem 2 (Basic Criterion) If classes $\left\{\mathcal{C}_{i}\right\}$ of events are independent and if each $\mathcal{C}_{i}$ is $a \pi$-system, then $\left\{\sigma\left(\mathcal{C}_{i}\right)\right\}$ are independent too.

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

Then

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

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

Thus $\mathcal{L}$ is a $\lambda$-system containing $\mathcal{C}_{i}$, and so by Dynkin's $\pi$ - $\lambda$ theorem it contains $\sigma\left(\mathcal{C}_{i}\right)$. Thus $\sigma\left(\mathcal{C}_{i}\right)$ and $\left\{A_{j}\right\}_{j \in J}$ are independent for each $\left\{A_{j} \in \mathcal{C}_{j}\right\}$, so $\left\{\sigma\left(\mathcal{C}_{i}\right),\left\{\mathcal{C}_{j}\right\}_{j \in J}\right\}$ are independent $\pi$-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 $\left\{X_{i}\right\}$ on some probability space $(\Omega, \mathcal{F}, \mathrm{P})$ are called independent if the $\sigma$-algebras $\mathcal{F}_{i}:=\sigma\left(X_{i}\right)=X_{i}^{-1}(\mathcal{B})$ they generate are independent, i.e., if

$$
\mathrm{P}\left(\bigcap_{i \in I}\left[X_{i} \in B_{i}\right]\right)=\prod_{i \in I} \mathrm{P}\left[X_{i} \in B_{i}\right]
$$

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

$$
\begin{equation*}
\mathrm{P}\left(\bigcap_{i \in I}\left[X_{i} \leq x_{i}\right]\right)=\prod_{i \in I} F_{i}\left(x_{i}\right) \tag{1}
\end{equation*}
$$

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 (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 $X \sim \operatorname{Bi}(7,0.3)$, absolutely continuous $Y \sim \operatorname{Ex}(2.0)$, mixed $Z=(\zeta \wedge 0)$ for $\zeta \sim$ 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.

Indepenence is a property of the probability measure and the $\sigma$-algebras $\left\{\sigma\left(X_{i}\right)\right\}$, not of the random variables $\left\{X_{j}\right\}$ themselves. Since $\sigma(g(X)) \subseteq \sigma(X)$ for any random variable $X$ and Borel function $g(\cdot)$, if $\left\{X_{i}\right\}$ are independent and if $g_{i}(\cdot)$ are arbitrary Borel functions, it follows that $\left\{g_{i}\left(X_{i}\right)\right\}$ are independent too- and, in particular, that if $X \Perp Y$ then $X \Perp g(Y)$ 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 $\left\{E_{i}\right\}$, random variables $\left\{X_{j}\right\}$, and classes of events $\left\{\mathcal{C}_{k}\right\}$ are independent if and only if the $\sigma$-algebras they generate $\left\{\sigma\left(E_{i}\right), \sigma\left(X_{j}\right), \sigma\left(\mathcal{C}_{k}\right)\right\}$ are independentwe can treat all of these in the same unified way.


[^0]:    ${ }^{1}$ 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 $[X]=\{Y: \Omega \rightarrow \mathbb{R}: \mathrm{P}[X \neq Y]=0\}$ of $p$-integrable random variables; see paragraph below.

[^1]:    ${ }^{2}$ In HW07 you will show that Hölder's and Minkowski's Inequalities also hold for $p=1$ and $p=\infty$.

