STA 711: Probability & Measure Theory
Robert L. Wolpert

5 Independence

5.1 Independent Events

A collection of events $\{A_i\} \subset \mathcal{F}$ in some probability space $(\Omega, \mathcal{F}, \mathsf{P})$ 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 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 $A_1 = \{$ first coin shows Heads $\}$, $A_2 = \{$ second coin shows Heads $\}$, $A_3 = \{$ coins disagree $\}$; then each $\mathsf{P}[A_i] = 1/2$ and each $\mathsf{P}[A_i \cap A_j] = 1/4$ for $i \neq j$, but $\cap A_i = \emptyset$ has probability zero.

5.2 Independent Classes of Events

Classes $\{\mathcal{C}_i\}$ of events are called *independent* if

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

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

Theorem 1 (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 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:

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

Then

- $\Omega \in \mathcal{L}$, obviously;
- $B \in \mathcal{L} \Rightarrow B^c \in \mathcal{L}$, quick computation;

- ~~$B_n \in \mathcal{L}$ and $\{B_n\}$ disjoint $\Rightarrow \cup B_n \in \mathcal{L}$, 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 mathematical induction) completes the proof. ∎

5.3 Independent Random Variables

A collection of random variables $\{X_i\}$ on some probability space $(\Omega, \mathcal{F}, \mathsf{P})$ are called *independent* 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 $\{B_i \in \mathcal{B}(\mathbb{R})\}$. This is just the same as the requirement that the σ-algebras $\mathcal{F}_i := \sigma(X_i) = X_i^{-1}(\mathcal{B})$ be independent; by the Basic Criterion it is enough to check that the joint CDFs factor, *i.e.*, that

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

for each $x \in \mathbb{R}^I$ (or just a dense set of them).

For jointly continuous random variables this is equivalent to requiring that the joint density function factor as the product of marginal density functions, while for 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 random variables $X \sim \mathsf{Bi}(7, 0.3)$, $Y \sim \mathsf{Ex}(2.0)$, $Z = \zeta \wedge 0$ for $\zeta \sim \mathsf{No}(0,1)$, and C with the Cantor distribution. It also applies to infinite (even uncountable) collections of random variables.

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 $g(\cdot)$.

5.4 Two Zero-One Laws

First, a little toy example to motivate. 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 probablity 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?

5.4.1 The Borel-Cantelli Lemmas and Borel's Zero-One Law

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

Lemma 2 (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.

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

□

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

Lemma 3 (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 limit supremum) has probability one.

Proof. First recall that $1 + x \le e^x$ for all real $x \in \mathbb{R}$, positive or not (draw 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 $N \to \infty$ for any fixed $n \in \mathbb{N}$; thus each $\cap_{m=n}^{\infty} A_m^c$ is a null set, and so is their union, so

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

□

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.*

5.4.2 Kolomogorov's Zero-One Law

For any collection $\{X_n\}$ of random variables on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$, define two sequences of σ-algabras by:

$$\mathcal{F}_n := \sigma\{X_i : \ i \leq n\} \qquad \mathcal{T}_n := \sigma\{X_i : \ i \geq n+1\}$$

and, from them, construct the π-system $\mathcal{P}$ and σ-algebra $\mathcal{T}$ by

$$\mathcal{P} := \bigcup_{n=1}^{\infty}\mathcal{F}_n \qquad \mathcal{T} := \bigcap_{n=1}^{\infty}\mathcal{T}_n.$$

In general $\mathcal{P}$ will not be a σ-algebra, because it will not be closed under countable unions or intersections, but it is a field and hence a π-system, and generates the σ-algebra $\vee\mathcal{F}_n \equiv \sigma(\mathcal{P}) \subseteq \mathcal{F}$.

The class $\mathcal{T}$, called the *tail* σ-field, includes such events as "X_n converges" or "$\limsup X_n \leq 1$" or, with $S_n := \sum_1^n X_j$, "$\frac{1}{n}S_n$ Converges" or "$\frac{1}{n}S_n \to 0$."

Theorem 4 (Kolmogorov's Zero-One Law) *For independent random variables X_n, the tail σ-field $\mathcal{T}$ is "almost trivial" in the sense that every event $\Lambda \in \mathcal{T}$ has probability $\mathsf{P}[\Lambda]=0$ or $\mathsf{P}[\Lambda]=1$.*

Proof. Let $A \in \mathcal{P} = \cup\mathcal{F}_n$, and $\Lambda \in \mathcal{T}$. Then for some $n \in \mathbb{N}$, $A \in \mathcal{F}_n$ and $\Lambda \in \mathcal{T}_n$, so $A \perp\!\!\!\perp \Lambda$. Thus $\mathcal{P}$ and $\mathcal{T}$ are independent; since $\mathcal{P}$ is a π-system, it follows from the Basic Criterion that $\sigma(\mathcal{P})$ and $\mathcal{T}$ are independent. But each X_n is $\sigma(\mathcal{P})$-measurable, so $\mathcal{T} \subset \sigma(\mathcal{P})$ and each $\Lambda \in \mathcal{T}$ must also be in $\sigma(\mathcal{P}) \perp\!\!\!\perp \mathcal{T}$; thus

$$\mathsf{P}[\Lambda] = \mathsf{P}[\Lambda \cap \Lambda] = \mathsf{P}[\Lambda]\mathsf{P}[\Lambda] = \mathsf{P}[\Lambda]^2,$$

so $0 = \mathsf{P}[\Lambda]\big(1-\mathsf{P}[\Lambda]\big)$ proving the theorem. □

5.5 Product Spaces

Do independent random variables exist, with arbitrary (marginal) distributions? How can they be constructed? One way is to build product probability spaces; let's see how to do that.

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

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

Any random variables X_1 on $(\Omega_1, \mathcal{F}_1, \mathsf{P}_1)$ and X_2 on $(\Omega_2, \mathcal{F}_2, \mathsf{P}_2)$ can be extended to the common space $(\Omega, \mathcal{F}, \mathsf{P})$ by defining $X_1^*(\omega_1, \omega_2) := X_1(\omega_1)$ and $X_2^*(\omega_1, \omega_2) := X_2(\omega_2)$; it's easy to show that $\{X_j^*\}$ are independent and have the same marginal distributions as $\{X_j\}$. Thus, independent random variables do exist with arbitrary distributions. The same construction extends to countable families.

5.6 Fubini's Theorem

We now consider how to evaluate probabilities and integrals on product spaces.

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

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

It's not completely obvious, but one can verify that $A_{\omega_2} \in \mathcal{F}_1$— it's trivial for product sets $A = A_1 \times A_2$, but we need a π-λ argument to conclude it for all of $\mathcal{F}$. What happens for sets $A \subset \mathcal{F}^{\mathsf{P}}$ in the P-completion of $\mathcal{F}_1 \times \mathcal{F}_2$?

Similarly, for any $\mathcal{F}$-measurable random variable $X : \Omega_1 \times \Omega_2 \to \mathcal{S}$ ($\mathcal{S}$ would be $\mathbb{R}$, for real-valued RVs, but could also be $\mathbb{R}^n$ or any metric space), and for any $\omega_2 \in \Omega_2$, the *section* of X is $X_{\omega_2} : \Omega_1 \to \mathcal{S}$ defined by

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

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

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

$$\int_{\Omega_2}\left\{\int_{\Omega_1} X_{\omega_2} d\mathsf{P}_1\right\} d\mathsf{P}_2 \stackrel{?}{=} \iint_{\Omega} X\, d\mathsf{P} \stackrel{?}{=} \int_{\Omega_1}\left\{\int_{\Omega_2} X_{\omega_1} d\mathsf{P}_2\right\} d\mathsf{P}_1 \tag{1}$$

To prove this, first note that it's true for indicators $X = \mathbf{1}_{A_1\times A_2}$ of measurable rectangles with each $A_j \in \mathcal{F}_j$; then verify that the class $\mathcal{C}$ of events $B \in \mathcal{F}$ for which it hold for $X = \mathbf{1}_B$ is a λ-system. Since the rectangles form a π-system it follows that $\mathcal{F} \subset \mathcal{C}$ so (1) holds for all indicators $X = \mathbf{1}_B$ of events $B \in \mathcal{F}$, hence for all nonnegative simple functions in $\mathcal{E}_+$, and finally for all $\mathcal{F}$-measurable $X \geq 0$ by the MCT. For $X \in L_1$, apply this result separately to X_+ and X_-.

Fubini's theorem applies more generally. Each probability measure P_j may be replaced by an arbitrary σ-finite measure; or one of them (say, P_2) may be replaced by a measurable kernel $K(\omega_1, d\omega_2)$ that is a σ-finite measure $K(\omega_1, \cdot)$ on $\mathcal{F}_2$ in its second variable for each fixed ω_1, and an $\mathcal{F}_1$-measurable function $K(\cdot, B)$ in its first variable for each fixed $B \in \mathcal{F}_2$.

As an easy consequence (why?), for any sequence of random variables we may exchange summation and expectation and conclude

$$\mathsf{E}\left\{\sum_{n=1}^{\infty} X_n\right\} \stackrel{?}{=} \sum_{n=1}^{\infty}\{\mathsf{E}X_n\}$$

whenever each $X_n \geq 0$ or when $\sum_{n=1}^{\infty}\mathsf{E}|X_n| < \infty$, but otherwise equality may fail. For an example, integrate by parts to verify:

$$\int_0^1\left\{\int_0^1 \frac{y^2-x^2}{(x^2+y^2)^2}\,dx\right\} dy = \int_0^1\left\{\frac{-1}{1+y^2}\right\} dy = \frac{-\pi}{4}$$
$$\int_0^1\left\{\int_0^1 \frac{y^2-x^2}{(x^2+y^2)^2}\,dy\right\} dx = \int_0^1\left\{\frac{+1}{1+x^2}\right\} dx = \frac{+\pi}{4}$$

and, as expected in light of Fubini's Theorem,

$$\begin{aligned}\iint_{0,0}^{1,1}\left|\frac{y^2-x^2}{(x^2+y^2)^2}\right| dx\,dy &\geq \int_0^{\pi/2}\int_0^1 \frac{r^2|\sin^2\theta-\cos^2\theta|}{r^4}\, r\,dr\,d\theta \\ &= \left(\int_0^{\pi/2}\sin^2(2\theta)\,d\theta\right)\left(\int_0^1 r^{-1}\,dr\right) \\ &= (\pi/4)\,(\infty)\end{aligned}$$

5.7 Hoeffding's Inequality

If $\{X_j\}$ are independent and (individually) bounded, so $(\forall\ j \in \mathbb{N})$ $(\exists\ \{a_j, b_j\})$ for which $\mathsf{P}[a_j \le X_j \le b_j] = 1$, then $(\forall c > 0)$, $S_n := \sum_{j=1}^n X_j$ satisfies

$$\mathsf{P}\big[(S_n - \mathsf{E}S_n) \ge c\big] \le \exp\Big(-2c^2 \Big/ \sum_1^n |b_j - a_j|^2 \Big).$$

If X_j are iid and bounded by $\|X_j\|_\infty \le 1$, e.g., then

$$\mathsf{P}[(\bar{X}_n - \mu) \ge \epsilon] \le e^{-n\epsilon^2/2}.$$

Wassily Hoeffding proved this improvement on Chebychev's inequality for L_∞ random variables in 1963 at UNC. It follows from Hoeffding's Lemma:

$$\mathsf{E}[e^{\lambda(X_j - \mu_j)}] \le \exp\big(\lambda^2 (b_j - a_j)^2/8\big),$$

proved in turn from Jensen's ineq and Taylor's theorem (with remainder). The importance is that the bound decreases *exponentially*, while Chebychev only decreases like a power; the price is that $\{X_j\}$ must be bounded in L_∞, not merely in L_2. See also related and earlier **Bernstein**'s inequality (1937), **Chernoff** bounds (1952), and **Azuma**'s inequality (1967).

Here's a proof for the important special case of $X_j = \pm 1$ with probability 1/2 each (and hence $\mu = 0$):

$$\begin{aligned}
\mathsf{P}[\bar{X}_n \ge \epsilon] &= \mathsf{P}[S_n \ge n\epsilon] \\
&= \mathsf{P}\left[e^{\lambda S_n} \ge e^{n\lambda\epsilon}\right] && \text{for any } \lambda > 0 \\
&\le \mathsf{E}\ e^{\lambda S_n}\ e^{-n\lambda\epsilon} && \text{by Markov's inequality} \\
&= \left\{\tfrac{1}{2}e^{\lambda} + \tfrac{1}{2}e^{-\lambda}\right\}^n\ e^{-n\lambda\epsilon} && \text{by independence} \\
&\le \left\{e^{\lambda^2/2}\right\}^n\ e^{-n\lambda\epsilon} && \text{see footnote}^1 \\
&= \exp\big(n\lambda^2/2 - n\lambda\epsilon\big)
\end{aligned}$$

The exponent is minimized at $\lambda = \epsilon$, so

$$\mathsf{P}[\bar{X}_n \ge \epsilon] \le \exp\big(n\epsilon^2/2 - n\epsilon\epsilon\big) = e^{-n\epsilon^2/2}.$$

The general case isn't much harder, but proving that $\mathsf{E}e^{\lambda X} \le e^{\lambda^2/2}$ is a bit more delicate.

By Borel/Cantelli it follows from Hoeffding's inequality that $(\bar{X}_n - \mu) > \epsilon$ only finitely-many times for each $\epsilon > 0$, if $\{X_n\} \subset L_\infty$ are iid, leading to our first **Strong Law of Large Numbers**: $\mathsf{P}[\bar{X}_n \to \mu] = 1$ (why does this follow?).

Note that Chebychev's inequality only guarantees the algebraic bound $\mathsf{P}[\bar{X}_n \ge \epsilon] \le 1/n\epsilon^2$, instead of Hoeffding's exponential bound. Since $1/n\epsilon^2$ isn't summable in n, Chebychev's bound isn't strong enough to prove a strong LLN.

Last edited: November 6, 2012

[1] $\cosh(\lambda) = \sum \frac{\lambda^{2k}}{(2k)!} \le \sum \frac{(\lambda^2)^k}{2^k(k!)} = e^{\lambda^2/2}$