COUNTEREXAMPLES IN PROBABILITY - INDEPENDENCE

R. TRAYLOR, PH. D.

§1. Introduction

Counterexamples are the "dual" to theorems in terms of mathematical understanding. Theorems, propositions, lemmata, and the like help us to understand the conditions under which a concept or topic does work. Counterexamples show when concepts don't work, and assist in explaining why those hypotheses in theorems were needed.

Probability theory is a more subtle discipline than many give it credit for. Its nuance can cause misunderstandings that perpetuate through applications in statistics, data science, machine learning, and the other scientific and business fields that use these tools. The concept of independence is one often used but less understood, as is the difference between random events and random variables. This article will move through counterexamples involving the concept of independence. Many of these counterexamples can be found in Stoyanov [5]

§2. Definitions and Preliminaries

Definition 2.1 (Random Event)**.** A **random event** is a property that can be observed to either hold or not after an experiment is done or phenomenon observed.

This is a nice, elegant definition of a random event, from Jacod and Protter [4], and will be set apart nicely when we define a random variable, and does have a philosophical ring to it. We may not know the outcome of a phenomenon prior to observing it (or it occurring), but we can observe the result. Obvious examples include the result of a coin toss, the number of heads in 3 successive tosses, the sum of the faces on two die rolls, the actual lifetime of a lightbulb, and so forth.

We'll continue by defining a probability space on events.

Definition 2.2 (Probability space on events)**.** A **probability space** consists of a triple $(\Omega, \mathscr{F}, P)$ whose elements are the following:

- Ω is a space consisting of elements or "points" we call *elementary events* or outcomes. We put no restrictions on the structure of this space. This is commonly referred to as the *sample space*.
- $\mathscr{F}$ is a set of subsets of Ω with a particular structure we'll define below. Elements $A \in \mathscr{F}$ are called *events*.
- P is a *probability* on $\mathscr{F}$ that assigns a numerical value to events. This mapping (also called a measure) also must obey some rules we'll discuss below.

We'll discuss each of these pieces in turn. Ω we've already covered. An explicit example is the experiment of tossing two coins. Here Ω enumerates all possibilities for the results of two coin tosses. $\Omega = \{\{H,H\},\{H,T\},\{T,H\},\{T,T\}\}$.

$\mathscr{F}$ is a set of subsets of Ω with certain requirements:

Date: March 3, 2019.

(1) For every event $A \in \mathscr{F}$, its complementary event $A^c \in \mathscr{F}$. For $A = \{\text{heads on first toss}\}$, $A^c = \{\text{tails on first toss}\}$. $A \cup A^c = \Omega$ for every event A, so formally, we may say that $A^c = \Omega \setminus A$.
(2) For a countable sequence of events $(A_n)_{n\in\mathbb{N}}$, $\cup_n A_n \in \mathscr{F}$. That is, the set of subsets $\mathscr{F}$ should be closed under countable unions.
(3) $\emptyset \in \mathscr{F}$.

A set of subsets $\mathscr{F}$ that satisfies the above three requirements is called a *σ-field*. Note that $\mathscr{F} \subset 2^{\Omega}$. We don't need $\mathscr{F}$ to be the set of all possible subsets of Ω. There are intelligent choices we can make, which will lead us to the definition of a Borel σ-field, but we'll visit that later.

P is the probability on $\mathscr{F}$ that must satisfy the following:
(1) $P(A) \geq 0$ for every $A \in \mathscr{F}$ and $P(\Omega) = 1$.
(2) P is finitely additive. That is, for every finite number of pairwise disjoint events $A_1, A_2, \ldots, A_n \in \mathscr{F}$,

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i).$$

(The astute reader will notice that this is the definition of finite additivity for measures as well. Probabilities are measures.)
(3) P is continuous at $\emptyset$. That is, for every decreasing and countable sequence of events $A_1, A_2, \ldots \in \mathscr{F}$, $A_{n+1} \subset A_n$, and $\cap_n A_n = \emptyset$,

$$\lim_{n\to\infty} P(A_n) = 0.$$

Conditions (2) and (3) can be combined into one condition called σ-additivity:

For every countable set of pairwise disjoint events $A_1, A_2, \ldots \in \mathscr{F}$, $P\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.

Finally, we'll need the definition of independence for events.

Definition 2.3 (Independence of events)**.** Let $(\Omega, \mathscr{F}, P)$ be a probability space. Then events $A, B \in \mathscr{F}$ are **independent** if $P(A \wedge B) = P(A)P(B)$.

This is the definition we're all familiar with. If the probability of the two events occurring simultaneously (or the probability of the intersection of the two subsets A and B) is equal to the product of the individual probabilities, then the two events are considered independent. Of course, thanks to Bayes's formula, we can exhibit an equivalent definition based on conditional probabilities as long as the event on which we condition doesn't have probability 0. (See our article on Borel's paradox for a discussion on this.) We're interested in working only with independence at the moment, so we'll ignore the other definition for this article.

We can extend this definition to any finite number of events or classes and arrive at the concept of *mutual independence*:

Definition 2.4 (Mutual Independence)**.** Let $(\Omega, \mathscr{F}, P)$ be a probability space. Events $A_1, A_2, \ldots, A_n \in \mathscr{F}$ are **mutually independent** if for all $k = 2, 3, \ldots, n$ and for all $i_1, i_2, \ldots, i_k$ such that $1 \leq i_1 < i_2 < \ldots < i_k \leq n$,

$$P(A_{i_1} \wedge A_{i_2} \wedge \ldots \wedge A_{i_k}) = \prod_{j=1}^{k} A_{i_j}.$$

This innocuous-looking definition is actually extremely restrictive. It requires us to test the "product rule" for every possible subset of events in the desired set, all the way up to the entire set itself. For four events A_1, A_2, A_3, A_4 we must test the product rule for all pairs, all subsets of three events, and all four events together to determine mutual independence. We must satisfy

$$\begin{aligned} P(A_i \wedge A_j) &= P(A_i)P(A_j), & i \neq j;\, i,j = 1,2,3,4 \\ P(A_i \wedge A_j \wedge A_k) &= P(A_i)P(A_j)P(A_k), & i \neq j \neq k;\, i,j,k = 1,2,3,4 \\ P(A_1 \wedge A_2 \wedge A_3 \wedge A_4) &= P(A_1)P(A_2)P(A_3)P(A_4) \end{aligned}$$

simultaneously. For n events, there are $2^n - n - 1$ relations to satisfy, so the larger the set of events, the more relations we have to test, supposedly.

At this point, we'll proceed to discussing some examples and counterexamples involving questions on independence for events.

§3. COUNTEREXAMPLES INVOLVING EVENTS

3.1. A SIMPLE EXAMPLE. It is entirely possible to create a scenario under which two events are independent only for certain distributions of probability mass. This example comes from a Bernoulli scenario with parameter p, $0 < p < 1$. Assign the following probabilities to the sides of a coin: $P(H) = p$, $P(T) = 1 - p$. Toss this coin three times, and record the sequence of results.

$$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

is the set of elementary events. $\mathscr{F}$ is a set of subsets of Ω that matches the definition of a σ-field. Let $A \in \mathscr{F}$ be an event defined as $A = \{\text{at most one tail}\}$, and $B \in \mathscr{F}$ be $B = \{\text{all tosses are the same}\}$. Then in terms of subsets of Ω,

$$A = \{HHH, HTH, HHT, THH\} \quad \& \quad B = \{HHH, TTT\}$$

To determine independence, we need to also know what the subset $A \wedge B$ entails:

$$A \wedge B = \{HHH\}$$

Note: In calculating probabilities of events in Ω, we are implicitly assuming mutual independence of tosses in invoking the product rule on them. In addition, the elementary events in both A and B are disjoint, so we can obtain the probabilities of A and B by adding the probabilities of the individual elements of each subset.

Now,

$$\begin{aligned} P(A) &= p^3 + 3p^2(1-p) \\ P(B) &= p^3 + (1-p)^3 \\ P(A \wedge B) &= p^3. \end{aligned}$$

In order for the two events to be independent,

$$P(A \wedge B) = P(A)P(B)$$

so we must have that

$$p^3 = \left(p^3 + 3p^2(1-p)\right)\left(p^3 + (1-p)^3\right).$$

Solving this cubic equation in p with your favorite method, we get equality only when $p = 0, 1, \frac{1}{2}$.

Thus, for any values of p that are not the values above (of which $p = \frac{1}{2}$ is the only nontrivial probability), the two events will not be independent. The independence of the two events in this scenario depends on the way we allocate probability mass to the two atoms (H and T).

3.2. DOES PAIRWISE INDEPENDENCE IMPLY MUTUAL INDEPENDENCE?. For the next few examples/counterexamples, we'll explore if there is any "royal road" to determining mutual independence of a set of events that doesn't involve testing every single one of the $2^{n-1}-n-1$ relations involved in the definition of mutual independence. For this first question, we'll see that having pairwise independence in a set of three random variables will not net us the "level 3" product rule involving all three.

Example 1: Suppose an urn contains 16 capsules marked as follows:

3 marked 111	1 marked 110
3 marked 100	1 marked 101
3 marked 010	1 marked 011
3 marked 001	1 marked 000

Choose a capsule at random, with equal probability. The markings on the capsules provide our Ω. Now define events $A_j = \{1\text{in position } j\}$, $j = 1,2,3$. For example $A_1 = \{111, 100, 110, 101\}$ (with the appropriate repetitions, so $|A_1| = 8$ with 4 distinct elements).

We'll calculate the probabilities of the individual events, and the three pairwise intersections formed by A_1, A_2, A_3.

First,

$$P(A_1) = \frac{3}{16} + \frac{3}{16} + \frac{1}{16} + \frac{1}{16} = \frac{1}{2},$$

in accordance with the numbers of capsules each in A_1. Similar calculations show that $P(A_2) = P(A_3) = P(A_1) = \frac{1}{2}$.

Next, we have to test each pair of events for independence. As an example,

$$A_1 \wedge A_2 = \{1 \text{ in postion } 1 \text{ and } 1 \text{ in position } 2\} = \{111, 110\}$$

and

$$P(A_1 \wedge A_2) = \frac{3}{16} + \frac{1}{16} = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_1)P(A_2)$$

Performing this same exercise on the pairs A_1/A_3 and A_2/A_3 yield the same satisfaction of the pairwise product rule. However, $A_1 \wedge A_2 \wedge A_3 = \{111\}$, and

$$P(A_1 \wedge A_2 \wedge A_3) = \frac{3}{16} \neq \frac{1}{8} = P(A_1)P(A_2)P(A_3)$$

Since this product rule is not satisfied, the set of events A_1, A_2, A_3 are not mutually independent, showing that having all subsets of pairs of a set of events independence does not yield any higher levels of independence.

Example 2: We can amend the above example to also create a counterexample that involves the triple intersection being the empty set $\emptyset$:

Suppose a box contains four tickets marked 112, 121, 211, 222, drawn with equal probability and keep A_i, $i = 1,2,3$ defined the same way as above. We can verify pairwise independence, but when we look at $A_1 \wedge A_2 \wedge A_3$, we see there is no ticket marked 111, and thus $A_1 \wedge A_2 \wedge A_3 = \emptyset$ and $P(\emptyset) = 0$. Thus we cannot satisfy the product rule on the set of 3.

3.3. DOES SATISFACTION OF HIGHER LEVELS OF PRODUCT RULES IMPLY PAIRWISE INDEPENDENCE?. Next, we'll flip the question around. If we have higher levels of product rules satisfied (for subsets of size 3 or 4 or greater), can we get pairwise independence? The question is a valid one. Mutual independence involves testing $2^n - n - 1$ relations, the largest set of which is the $\binom{n}{2}$ pairwise relations. If we can find a way to declare pairwise independence without so much computation, that would be helpful. Alas, we get no such luck, as the below examples show:

Example 1: Let $\Omega = \{1,2,3,\ldots,8\}$, and $P(\omega \in \Omega) = \frac{1}{8}$. Define events $B_1 = \{1,2,3,4\}$ and $B_2 = \{1,5,6,7\} = B_3$. Here when we say $B_2 = B_3$ we mean that the two events are identical, not identically distributed. If we know B_2, we know B_3, so by intuition these two events should not be independent. (We can also verify this mathematically as well via the product rule. This is a nice illustration that our mathematical definitions should line up with intuition.) However, $B_1 \wedge B_2 \wedge B_3 = \{1\}$, with probability $\frac{1}{8}$, which indeed equals the product of probabilities $P(B_1)$, $P(B_2)$ and $P(B_3)$. Thus, we can create events that are independent at level 3, but clearly functionally dependent on level 2, which indicates in some ways that higher level product rule satisfaction can be weaker than pairwise.

Example 2: Let $\Omega = \{1,2,\ldots 16\}$ and draw a number from Ω with equal probability. We'll show that having independence at levels 3 *and* 4 don't imply pairwise independence. Define

$$A = \{2,3,4,5,6,9,13,16\} \qquad C = \{4,6,7,8,10,11,13,14\}$$
$$B = \{4,7,8,10,11,13,14,16\} \qquad D = \{3,4,5,6,9,10,15,16\}$$

$P(A) = P(B) = P(C) = P(D) = \frac{1}{2}$. $A \wedge B \wedge C \wedge D = \{1\}$ and has probability

$$P(A \wedge B \wedge C \wedge D) = \frac{1}{16} = P(A)P(B)P(C)P(D).$$

We can go through the same (albeit tedious) calculations for the 4 subsets of 3 events to see that every subset of three events of $\{A,B,C,D\}$ satisfies the product rule. At the pairwise level, we have $\binom{4}{2} = 6$ pairs to test, and they all must satisfy the product rule for the set to be pairwise independent. However, $C \wedge D = \{4,6,10\}$ and $P(C \wedge D) = \frac{3}{16}$ which is certainly not equal to $P(C)P(D)$.

It's worth pointing out that the specific numbers in the events A, B, C, D aren't really that important. What's important is the sizes of the intersections. We can design the events with different numbers, but the relation between them via intersection will remain the same to exhibit this counterexample. We could just as easily have substituted names or book titles for the numbers.

Lest the reader think these are just toy examples, misunderstanding of this concept has very applicable consequences. This leads to what's known as the multiple comparisons problem when testing multiple statistical hypotheses at once. The more attributes a researcher tests for, the more likely an erroneously significant difference will be found. In addition, if there is any dependence among the set of tests (that is, they are not mutually independent), then even fixes such as the family-wise error rate or the Bonferroni correction may not help all the way or only provide bounds on the tests. Ultimately, there is no royal road to determining mutual independence. We have to check all the relations.

3.4. INDEPENDENT CLASSES CAN PRODUCE DEPENDENT σ-FIELDS. Suppose we have two classes of random variables $\mathscr{A}$ and $\mathscr{B}$, each comprising a set of events from $\mathscr{F}$, that are independent. The σ-field generated by each class, denoted by $\sigma(\mathscr{A})$ and $\sigma(\mathscr{B})$ respectively, is the smallest σ-field that contains all the elements of the class. The σ-fields generated by classes or sets of events are bigger than the sets that generate them. We'd like to know if independence can "extend" to these generated σ-fields. Naturally, since this is a report on counterexamples, the answer is no.

Example 1: We'll take a look on a continuous space this time. Let $(\Omega, \mathscr{F}, P) = ([0,1], \mathscr{B}_{[0,1]}, \lambda)$, where $\mathscr{B}_{[0,1]}$ is the Borel σ-field over $[0,1]$. This σ-field consists of all open intervals on $[0,1]$, though we can also use half-open, half-closed intervals. $[0,\frac{1}{2}) \in \mathscr{B}_{[0,1]}$, as is $[\frac{1}{2},\frac{3}{4}) \cup [\frac{7}{8},1)$, and so forth. Finally λ is the Lebesgue measure, which for our purposes is just the length of the intervals (or the sum of the lengths of subintervals in the case of disjoint unions).

Define events $a_{11} = [0,\frac{1}{4}) \cup [\frac{1}{2},\frac{3}{4})$, $a_{12} = [0,\frac{1}{3}) \cup [\frac{2}{3},1)$, and $a_2 = [0,\frac{1}{2})$. Now define $\mathscr{A}_1 = \{a_{11}, a_{12}\}$, and $\mathscr{A}_2 = \{a_2\}$ as our classes. To test independence of the classes, we just need to check that all elements between classes are mutually independent. In this simple case, that means testing only the two pairs a_{11}, a_2 and a_{12}, a_2.

$$P(a_{11}) = \lambda\Big([0,\tfrac{1}{4}) \cup [\tfrac{1}{2},\tfrac{3}{4})\Big) = \frac{1}{2}$$

$$P(a_{12}) = \lambda\Big([0,\tfrac{1}{3}) \cup [\tfrac{2}{3},1)\Big) = \frac{2}{3}$$

$$P(a_2) = \lambda\Big([0,\tfrac{1}{2})\Big) = \frac{1}{2}$$

Then,

$$P(a_{11} \wedge a_2) = P([0,\tfrac{1}{4})) = \frac{1}{4} = \lambda(a_{11})\lambda(a_2)$$

$$P(a_{12} \wedge a_2) = P([0,\tfrac{1}{3})) = \frac{1}{3} = \lambda(a_{12})\lambda(a_2)$$

and we see our classes $\mathscr{A}_1$ and $\mathscr{A}_2$ are independent. Now we consider the σ-fields $\sigma(\mathscr{A}_1)$ and $\sigma(\mathscr{A}_2)$. To show that these two $\sigma-$ fields are not independent, we only must exhibit two events, one event from each σ-field which are not independent.

The event $B = a_{11} \cap a_{12}$ is in $\sigma(\mathscr{A}_1)$, and $B = [0,\frac{1}{4}) \cup [\frac{2}{3},\frac{3}{4})$ and has $\lambda(B) = \frac{1}{3}$.

Next, simply take $a_2 \in \sigma(\mathscr{A}_2)$. Then

$$B \wedge a_2 = [0,\tfrac{1}{4}) \text{ and } \lambda(B \wedge a_2) = \frac{1}{4}$$

This is certainly not the product of measures of B and a_2, and thus the two $\sigma-$fields are not independent. Independence doesn't "extend" necessarily.

Example 2: We can amend this example to show that a discrete space can have independent classes events that produce dependent σ-fields. Take a small sample space with equally likely outcomes, and form two classes $\mathscr{A}_1$ and $\mathscr{A}_2$ where $\mathscr{A}_1 = \{\omega\}$ for some $\omega \in \Omega$, and $\mathscr{A}_2 = \{\omega_1, \omega_2\}$ for some other $\omega_1, \omega_2 \in \Omega$. The same argument can be used above to show that the σ-fields generated by these two classes are not independent.

A natural question that arises after finding a counterexample is to ask how we can "fix" the statement such that it is true for everything? What conditions do we need to add so that independent classes so produce independent σ-fields?

We can show that if $\mathscr{A}_1$ and $\mathscr{A}_2$ were π-systems, that is, for each class $\mathscr{A}_i, i = 1,2$, $\Omega \in \mathscr{A}_i$ and the A_i's are closed under intersections, in addition to being independent of each other, then these classes will produce independent σ-fields.

3.5. IS IT POSSIBLE TO HAVE A PROBABILITY SPACE WITH NO NON-TRIVIAL INDEPENDENT EVENTS?. As the subsection title indicates, we wish here to explore the idea of *totally dependent spaces* wherein we cannot find even a pair of independent events that are nontrivial. In fact, the answer is that we can certainly construct such spaces, but *only if the probability measure is purely discrete*. We'll explore two examples to get an idea, then prove why such a space cannot exist if P has any continuous part.

Example 1: Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$. We don't particularly care what these ω's are; they can be numbers, die rolls, objects, functions, anything. Take an irrational number $\epsilon \notin \mathbb{Q}$, and set $P(\omega_i) = \epsilon, i = 2,3,\ldots,n$, and $P(\omega_1) = 1-(n-1)\epsilon$. Naturally, we also need $0 < \epsilon < \frac{1}{n-1}$ to ensure $P(\Omega) = 1$.

We'll assume there exist generic nontrivial events $A, B \in \mathscr{F}$ that are independent, and arrive at a fundamental contradiction. Note that this is sufficient to show that no set of nontrivial independent events exists in this space, because lacking pairwise independence implies we cannot hope to reach mutual independence for larger sets, as our first sets of examples in sections 3.2 and 3.3 showed.

There are three distinct possibilities for ω_1 regarding events A and B:

(1) $\omega_1 \notin A$ and $\omega_1 \notin B$
(2) $\omega_1 \notin A$ and $\omega_1 \in B$ (and vice versa)
(3) $\omega_1 \in A$ and $\omega_1 \in B$

We can check each one of these cases and see that satisfying the product rule (as assumed) will lead to a contradiction of our initial definition of ϵ as irrational. The argument is very similar for all three cases, so we'll just exhibit it for one case.

Suppose $\omega_1 \notin A$ and $\omega_1 \in B$. Then $A = \{\omega_{i_1}, \ldots, \omega_{i_k}\}$, where $i_1, \ldots, i_k \in \{\omega_2, \ldots, \omega_k\}$. The elements of A can only be pulled from the elements of Ω besides ω_1. Therefore, $P(A) = k\epsilon$. Now, since $\omega_1 \in B$, $B = \{\omega_1, \omega_{l_1}, \ldots, \omega_{l_m}\}$, where $l_1, \ldots, l_m \in \{\omega_2, \ldots, \omega_n\}$ Then $P(B) = m\epsilon + 1 - \epsilon(n-1)$. Finally, $A \wedge B$ can only consist of some number u of events from $\{\omega_2, \omega_3, \ldots, \omega_n\}$, so $P(A \wedge B) = u\epsilon$. If A and B are independent, then

$$u\epsilon = k\epsilon(m\epsilon + 1 - \epsilon(n-1))$$

Solving for ϵ

$$\epsilon = \frac{u-k}{m-k(n-1)}.$$

Since u, k, m, and n are all integers, $\epsilon \in \mathbb{Q}$, which contradicts our construction of ϵ as irrational.

A similar argument for each of the other cases yields the same conclusion–that the independence of the events A and B contradicts our definition of ϵ as irrational. Thus, in the space we constructed, there are no two events that are independent. Since we've constructed such a space, we've shown it is possible to have a completely dependent space.

Example 2: What if the space Ω were countably infinite? Let $\Omega = \{\omega_1, \omega_2, \ldots\}$ be a countably infinite space. Define $P(\omega_k) = \frac{1}{2^{k!}}$ for $k \neq 1$, and $P(\omega_1) = 1 - \sum_{k=2}^{\infty} P(\omega_k)$.

The series $\epsilon = \sum_{k=2}^{\infty} \frac{1}{2^{k!}}$ does indeed converge, and it converges to a special type of irrational number called a *Liouville number*. We can (a bit tediously) repeat similar arguments as the finite case in Example 1 to show that any finite or infinite collection of arbitrarily composed random events cannot be independent without contradicting the irrationality of ϵ, thus showing that this space is indeed another totally dependent space.

In both of these cases, the probability measure P was purely discrete. What if P is not?

By a *non-purely discrete space*, we mean that P isn't just a "sum of atoms". That is, for some $\Omega_c \subset \Omega$ such that $P(\Omega_c) = c > 0$, $P(\omega) = 0$ for all $\omega \in \Omega_c$. The probability of the subset Ω_c is greater than 0, but the probability of any individual element of Ω_c has zero probability, so the sum of the probabilities of elements will not yield the probability of the subset. This is a *nonatomic* space, and implies P has a continuous component. For an ease of example, a probability space using the normal distribution would be an example here, though it is absolutely continuous, and more than we technically need.

We can show that it is only for purely discrete probability spaces that we can construct totally dependent spaces, and everything hinges on one theorem:

Theorem 3.1 (Lyapunov)**.** *For any real number b, $0 \leq b \leq c$, there exists some subset $D \subset \Omega_c$ such that $P(D) = b$.*

This isn't the original text of the theorem, which is more encompassing than the version written above, but it's been modified it to fit our context [5]. Essentially, for any real number b between 0 and the probability of the entire subset Ω_c, I can find a subset D in Ω_c that has that probability b. This theorem does not hold for discrete probability measures.

Continuing with our informal proof that no non-purely discrete probability space can be totally dependent, fix some number p such that $0 < p < c$. Then by Lyapunov's theorem, there exist three events D_1, D_2, D_3 that are pairwise disjoint such that $P(D_1) = p^2$, and $P(D_2) = P(D_3) = p(1-p)$. These events are all nontrivial. Additionally, note that $P(D_1 \cup D_2 \cup D_3) = p - p^2 < c$. Now define events $A = D_1 \cup D_2$, and $B = D_1 \cup D_3$. Then

$$P(A \wedge B) = P(D_1) = p^2 = p \cdot p = P(A)P(B).$$

We've thus shown that the nonatomic nature of P makes it impossible to construct a totally dependent space.

§4. COUNTEREXAMPLES INVOLVING RANDOM VARIABLES

4.1. DEFINITIONS AND PRELIMINARIES. The term "random variable" is often used, and rarely formally defined. Here we'll present two definitions, one more restrictive, and the other more abstract.

Definition 4.1 (Random variable (Tucker, 2014) citeTucker)**.** A **random variable** X is a real valued function with domain Ω such that for every real number x, $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathscr{F}$.

This condition is called $\mathscr{F}$-measurable. We don't actually have to require that a random variable map our event space Ω to the real numbers $\mathbb{R}$. Jacod and Protter [4] defines a random variable as a mapping to any other space.

Definition 4.2 (Random variable (Jacod and Protter, 2004))**.** Let $(\Omega, \mathscr{F}, P)$ be a probability space, and $(E, \mathscr{E})$ be another measurable space. Then a measurable function $X : \Omega \to E$ with respect to $\mathscr{F}, \mathscr{E}$ is called a **random variable**

There's admittedly a lot of formality in these definitions, and shouldn't be taken lightly. They weren't developed quickly. In fact, 215 years elapsed between De Moivre's

The Doctrine of Chances (1718) and Kolmogorov's *Foundations on the Theory of Probability* (1933) that formalized the intuitive notions of probability theory using measure theory.

Effectively, we may think colloquially of a random variable as a function that maps our event space onto another space on which we can do some "real math".

Definition 4.3 (Probability Distribution)**.** The **probability distribution** of X is defined on $X(\Omega)$ by

$$P^X(A) = P(\{\omega : X(\omega) \in A\}) = P(X^{-1}(A))$$

We define the probability distribution or law of X using the event probability of the inverse image of $A \in \mathscr{E}$. In this way, we can connect the probability on the event space to a probability distribution on X. The goal of this lecture isn't to get into the weeds of the measure theory, though it is worth consideration. This will be sufficient for our purposes.

Remark: Often the literature will use $(\Omega, \mathscr{F}, P)$ for both random events and random variables. The simplest random variable we can have is called the ***indicator function****, denoted either $\mathbb{I}(\cdot)$ or $\mathbb{1}(\cdot)$. The indicator function is defined on some set or subset A and takes value 1 if the argument is in A and 0 otherwise. This simple mapping turns events into random variables, so the two are often interchanged. From now on, we'll use $(\Omega, \mathscr{F}, P)$ for a probability space over random variables in keeping with convention. In addition, oftentimes $\mathscr{F}$ will be replaced with $\mathscr{B} \subset \mathscr{F}$, where $\mathscr{B}$ is the Borel σ-field over Ω. Since a Borel σ-field is the smallest σ-field containing Ω, anything in $\mathscr{B}$ is also in $\mathscr{F}$.*

We can now define pairwise and mutual independence for random variables.

Definition 4.4 (Pairwise Independence)**.** Let X_1, X_2 be random variables on a probability space $(\Omega, \mathscr{F}, P)$. We say X_1 and X_2 are **independent** if any of the following hold

(1) $P(X_1 \in B_1, X_2 \in B_2) = P(X_1 \in B_1)P(X_2 \in B_2)$ for al $B_1, B_2 \in \mathscr{B}$
(2) $F(x_1, x_2) = F_1(x_1)F_2(x_2)$ for all $(x_1, x_2) \in \mathbb{R}^2$, $x_1 \in \mathbb{R}, x_2 \in \mathbb{R}$, where F is the joint distribution of X_1, X_2, and F_1, F_2 are their respective marginals.
(3) If F, F_1, F_2 are absolutely continuous, then the densities f, f_1, f_2 exist and

$$f(x_1, x_2) = f_1(x_1)f_2(x_2),$$

for all $x_1, x_2 \in \mathbb{R}$.

All three above are equivalent to say X_1 and X_2 are independent, though showing the equivalence is nontrivial and requires some more advanced measure theory.

For discrete random variables, if $P(X_1 = x_{1i}) = p_i$, $i \in \mathbb{N}$, $\sum_i p_i = 1$ and $P(X_2 = x_{2j}) = p_j$, $j \in \mathbb{N}$, $\sum_j p_j = 1$, then X_1 and X_2 are independent if and only if

$$P(X_1 = x_{1i}, X_2 = x_{2j}) = P(X_1 = x_{1i})P(X_2 = x_{2j})$$

for all i, j.

Similarly, we can define mutual independence for random variables:

Definition 4.5 (Mutual Independence)**.** $X_1, X_2, \ldots, X_n$ are mutually independent random variables on $(\Omega, \mathscr{F}, P)$ if for all $2 \leq k \leq n$, $1 \leq i_1 < i_2 < \ldots < i_k \leq n$

$$P(X_{i_1} \in B_{i_1}, X_{i_2} \in B_{i_2}, \ldots, X_{i_k} \in B_{i_n}) = \prod_{l=1}^{k} P(X_{i_l} \in B_{i_l})$$

for arbitrary Borel sets $B_{i_l} \in \mathscr{B}$, $l = 1, 2, \ldots, k$.

We can also create similar definitions using distributions and densities modeling after pairwise independence and extending. The main point to note is that, just like for random events, we have to satisfy this new product rule at every single level.

At this point, we can begin by exploring the same sorts of questions we did for random variables.

4.2. PAIRWISE INDEPENDENCE DOES NOT IMPLY MUTUAL INDEPENDENCE. We'll show that pairwise independence implies mutual independence in neither the discrete case nor the continuous case.

Example 1: This example is similar to Example 2 in Section 3.2. Take

$$\Omega = \{(1,0,0),(0,1,0),(0,0,1),(1,1,1)\}.$$

(X,Y,Z) is a tuple of random variables corresponding to the value in each position. We assume each tuple in Ω is drawn with equal probability. Thus

$$P(X=1) = P\left((1,0,0)\vee(1,1,1)\right) = \tfrac{1}{2},$$

and so forth. We can verify that X, Y, Z are pairwise independent by testing all 12 combinations $(X=0,Y=1),(Y=0,Z=0)$, and so forth. As an illustration,

$$P(X=1,Z=0) = P((1,0,0)) = \frac{1}{4} = P(X=1)P(Z=0).$$

We test all combinations and see that the product rule is satisfied for all outcomes of pairs for X,Y,Z and thus the set of variables is pairwise independent. However,

$$P(X=1,Y=1,Z=1) = \tfrac{1}{4} \neq \tfrac{1}{8} = P(X=1)P(Y=1)P(Z=1),$$

and thus the three events cannot be mutually independent.

Example 2: In this example, we can show that the functional dependence of one variable in the set on the other two can yield pairwise independence of the set, but fail the higher level product rule and thus mutual independence.

Let S_3 denote the symmetric group on three elements (here we'll take the numbers 1, 2, and 3 as our elements, though it really doesn't matter much). This group is the set of all possible permutations of $(1,2,3)$. The elements of S_3 (using the notation $(1,2,3)$ for the elements) are

$$S_3 = \left\{\begin{pmatrix}123\\123\end{pmatrix}, \begin{pmatrix}123\\132\end{pmatrix}, \begin{pmatrix}123\\213\end{pmatrix}, \begin{pmatrix}123\\231\end{pmatrix}, \begin{pmatrix}123\\312\end{pmatrix}, \begin{pmatrix}123\\321\end{pmatrix}\right\}.$$

Let $\Omega = \{S_3,(1,1,1),(2,2,2),(3,3,3)\}$. Draw an element with equal probability. Define X_k as the number appearing in the kth place, $k=1,2,3$. That is X_1 is the number that appears in the first place of the tuple when one is randomly drawn. We can calculate that $P(X_i) = \frac{1}{3}$ for each $i=1,2,3$. In addition, due to the symmetry of Ω,

$$P(X_k = i, X_l = j) = \frac{1}{9},$$

for $k,l = 1,2,3$ and $i,j = 1,2,3$, so we have pairwise independence of the positions. However, any two variables uniquely determine the third. For example, if we know X_1 and X_3, we know X_2, because each tuple is unique in its pairwise combinations. Therefore, each variable is a function of the other two, which means they cannot be independent at level 3. We can also mathematically verify this by noting that the probability of any unique tuple is $\frac{1}{9}$, but the product of probabilities of the individual events is $\frac{1}{27}$.

Example 3: We move to an absolutely continuous example to show that pairwise independence doesn't imply mutual independence in this space either. Let ξ, η be continuous uniform random variables on $[0, \pi]$. Then

$$F_\xi(x) = F_\eta(x) = \begin{cases} \frac{x}{\pi}, & 0 \leq x \leq \pi \\ 0, & \text{otherwise} \end{cases}$$

gives the distributions of ξ and η. Define $X_1 = \tan(\xi)$, $X_2 = \tan(\eta)$ and $X_3 = -\tan(\xi + \eta)$ as random variables.

With a tedious amount of trig simplification, we can show that

$$X_1 X_2 X_3 = X_1 + X_2 + X_3$$

and thus

$$X_i = \frac{X_j + X_k}{X_j X_k - 1}, \quad i, j, k = 1, 2, 3; i \neq j \neq k.$$

We have functional dependence of each variable on the other two, so we know that we will not be able to satisfy the product rule at level 3. (Leveraging this fact saves us a lot of computation.)

To show that the set $\{X_1, X_2, X_3\}$ is pairwise independent, we make use of the following theorem:

Theorem 4.1 ((Tucker, 2014)). *If X, Y are independent random variables, and f, g are Borel-measurable functions, then $f(X)$ and $g(Y)$ are independent random variables.*

The Borel-measurability here is a bit more than we need, but $\tan(\cdot)$ is certainly a Borel-measurable function. Thus, we can conclude that X_1 and X_2 are independent. To verify that X_1, X_3 and X_2, X_3 are pairwise independent, we can derive the distributions F_{X_1}, F_{X_2}, and F_{X_3} and show that the product rule holds for the pairs. (Note that $F_{X_1} = F_{X_2}$. In general we find these distributions by Jacobian transformations (also seen in mulivariate calculus to change coordinate systems). We'll give a brief aside to explain the concept, and direct the interested reader to [3] for further details.

For the univariate case (which applies to X_1 and X_2), suppose X is a continuous random variable with density f_X and support[1] S_X. Let $Y = g(X)$, where $g(\cdot)$ is an injective differentiable function on S_X. Denote the inverse of g by $x = g^{-1}(y)$ and let $\frac{dx}{dy} = \frac{dg^{-1}(y)}{dy}$. Then the density of Y is given by

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dx}{dy} \right|,$$

where the support of Y is $S_Y = \{y = g(x) : x \in S_X\}$.

For the bivariate random variable, please see [**p.86**, 3]. This technique is how one can obtain the distribution for X_3. Once these distributions are obtained, we can see that X_1, X_3 and X_2, X_3 are independent pairs using the distributional definition of independence.

[1]The space on which X has positive probability is the support. For example, ξ has support $[0, \pi]$.

4.3. IF $f(X)$ AND $g(Y)$ ARE INDEPENDENT, ARE X AND Y?. We saw in Example 3 of section 4.2 that Borel measurable functions (or to be more strict, continuous) functions of independent random variables are themselves independent random variables. Now we'd like to see if the converse holds.

We can prove that if $f,g:\mathbb{R}\to\mathbb{R}$, then the question above answers "yes".[FIND CITATION] But without this condition, we'll see that the statement doesn't hold.

Example 1: For the discrete case, define a pair (X,Y) of random variables with $p_{i,j} = P(X=i, Y=j)$, $i,j\in\{-1,0,1\}$. We'll assign probability mass as follows:

$$p_{1,-1} = p_{1,1} = \tfrac{1}{32}, \qquad p_{-1,0} = p_{0,-1} = \tfrac{5}{32}$$
$$p_{-1,-1} = p_{1,-1} = p_{1,0} = p_{0,1} = \tfrac{3}{32}, \qquad p_{0,0} = \tfrac{8}{32}$$

Then the distributions of X and Y are as follows:

$$P(X) = \begin{cases} \frac{1}{2}, & X=0 \\ \frac{7}{32}, & X=1 \\ \frac{9}{32}, & X=-1 \end{cases} \text{ and } P(Y) = \begin{cases} \frac{1}{2}, & Y=0 \\ \frac{5}{32}, & Y=1 \\ \frac{11}{32}, & Y=-1 \end{cases}$$

To show X and Y are not independent, we simply need to find a combination for which the product rule doesn't hold. For instance,

$$P(X=0,Y=1) = \frac{3}{32} \neq \frac{1}{2}\cdot\frac{5}{32} = P(X=0)P(Y=1)$$

However, let's take $g(X)=X^2$ and $h(Y)=Y^2$. Then

$$P(X^2) = \begin{cases} \frac{1}{2}, & X=0 \\ \frac{1}{2}, & X=1 \end{cases} \text{ and } P(Y^2) = \begin{cases} \frac{1}{2}, & Y=0 \\ \frac{1}{2}, & Y=1 \end{cases}$$

We can check here that X^2 and Y^2 are indeed independent, yet X and Y are not.

Example 2: For the absolutely continuous case, take any two independent random variables X_1, X_2. Let $Y\in\{-1,1\}$ with equal probability, and assume Y is independent of both X_1 and X_2. Now define random variables $Z_1 = YX_1$ and $Z_2 = YX_2$. Due to the absolute continuity of X_1 and X_2, Z_1 and Z_2 are absolutely continuous. However, both depend on Y, so they have a functional dependence on each other as well and are thus not independent.

$Z_1^2 = X_1^2$ and $Z_2^2 = X_2^2$, so we know the squares are independent because X_1 and X_2 are independent.

[DO I WANT TO EXPAND ON THE WHY BITS HERE?]

§5. CHARACTERISTIC FUNCTIONS

5.1. DEFINITIONS AND PRELIMINARIES. Transforms of distributions of random variables are valuable in studying and solving problems that would otherwise be computationally intractable or just plain annoying on the distribution or density. In particular, the *characteristic function* of a distribution is one such transformation.

Definition 5.1 (Characteristic Function (Tucker, 2014)). Let X be a random variable on $(\Omega,\mathscr{F},P)$. Then its **characteristic function** is given by

$$\phi_X(u) = E\left[e^{iuX}\right].$$

Here $E[\cdot]$ is the *expectation operator*, given by the Lebesgue integral of X with respect to the probability measure P: $E[X]=\int X\,dP$. If X has distribution F, then

$$\phi_X(u)=\int_{-\infty}^{\infty} e^{iux}\,dF(x).$$

The characteristic function of a random variable is simply its Fourier transformation, and is unique. That is, if the characteristic functions of two random variables X and Y are equal, then $X=Y$.

We have the following theorem:

Theorem 5.1. *If X and Y are independent random variables, then $\phi_{X+Y}=\phi_X\phi_Y$.*

That is, for two independent random variables, the characteristic function of the sum is the product of the individual characteristic functions. This can be extended to mutual independence as well.

This convenient expression for the characteristic function of the sum of independent random variables is desired for all sorts of applications. For example, if we want to find the moments of the sampling distribution of a mean $\bar{X}$ of data, and we know the data are independent (less likely in practice, but that's a different conversation), then we don't have to use the more tedious Jacobian transformation to find this distribution. We can simply use the individual characteristic functions.

5.2. IF $\phi_{X+Y}=\phi(X)\phi(Y)$, ARE X AND Y INDEPENDENT?. Here we'll explore the converse to the above theorem. Without any other conditions, the answer to this question is "no."

Example 1: Let (X_1,X_2) have joint density

$$f(x_1,x_2)=\begin{cases}\frac{1}{4}\left[1+x_1x_2(x_1^2+x_2^2\right], & |x_1|\leq 1,|x_2|\leq 1\\ 0, & \text{otherwise.}\end{cases}$$

We can calculate the marginal densities of X_1 and X_2 respectively by integrating out the other variable.

$$f_1(x_1)=f_2(x_2)=f(x)=\begin{cases}\frac{1}{2}, & |x|\leq 1\\ 0, & \text{otherwise.}\end{cases}$$

Then clearly $f(x_1,x_2)\neq f_1(x_1)f_2(x_2)$ so X_1 and X_2 are not independent.

We can use the definition of the characteristic function to calculate $\phi_1(u)$ and $\phi_2(u)$:

$$\begin{aligned}\phi_1(u)&=\int_{-1}^{1}\frac{1}{2}e^{-iux}\,dx\\ &=\frac{\sin(u)}{u}\\ &=\phi_2(u).\end{aligned}$$

The distribution $g(x)$ of $X=X_1+X_2$ can be found by transformation and is given by

$$g(x)=\begin{cases}\frac{1}{4}(2+x), & -2\leq x\leq 0\\ \frac{1}{4}(2-x), & 0<x\leq 2.\end{cases}$$

Calculating the characteristic function of X in the same way,

$$\phi_{X_1+X_2}=\frac{\sin^2(u)}{u^2}=\phi_1\phi_2,$$

so we can indeed have a scenario in which the characteristic function of a sum of dependent random variables is the product of individual characteristic functions.

Example 2: Let $X_1 = X_2 = X \sim \text{Cauchy}(1,x)$. These random variables are identical, not just separate Cauchy random variables. The density for a Cauchy$(1,x)$ distribution is

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

We designed X_1 and X_2 to be identical, so clearly they aren't independent. However, a Cauchy family of random variables is closed under sums. Thus

$$\phi_{X_1+X_2}(u) = e^{-2|u|} = e^{-|u|}e^{-|u|} = \phi_1\phi_2.$$

To be honest, this one felt like a bit of a cheap shot. However, the way to "fix" this issue so that the converse to our theorem is true is to add the following condition that $u_1, u_2 \in \mathbb{R}$ to get the following theorem:

Theorem 5.2. *Let $X_1, X_2, \ldots, X_n$ be random variables with characteristic functions $\phi_1(u_1)$, $\phi_2(u_2), \ldots, \phi_n(u_n)$ respectively. Then $X_1, X_2, \ldots, X_n$ are independent random variables if and only if for all $u_1, u_2, u_n \in \mathbb{R}$,*

$$\phi_{X_1+X_2+\ldots+X_n} = E\left[\exp\left(i\sum_{j=1}^{n} X_j u_j\right)\right] = \prod_{j=1}^{n}\phi_j(u_j).$$

The issue in both of our examples is that the u's were not defined on all of $\mathbb{R}$.

§6. CONVOLUTIONS AND SUB-INDEPENDENCE

6.1. DEFINITIONS AND PRELIMINARIES.

Definition 6.1 (Convolution)**.** The **convolution** of two distributions F_X and F_Y, denoted $F_X \star F_Y$ is given by

$$F_X \star F_Y = \int_{\mathbb{R}} F_X(z-y)\,dF_Y(y), \quad z \in \mathbb{R}.$$

We have the following theorem:

Theorem 6.1. *If X_1, X_2 are independent random variables with distributions F_1 and F_2 respectively, then*

$$F_{X_1+X_2} = F_1 \star F_2.$$

We can get the distribution of the sum of independent random variables by taking the convolution of their distributions. Is the converse true?

6.2. IF $F_{X_1+X_2} = F_1 \star F_2$, ARE X_1 AND X_2 INDEPENDENT?.

Unfortunately, like we saw for characteristic functions, the answer is "no."

Example: Let $f_a(x) = \frac{a}{\pi(a^2+x^2)}$. This family of functions for $a \in \mathbb{R}$ is the Cauchy family of distributions. We saw an element of this family used in Example 2 of Section 5.2. We can show that the family of Cauchy densities is closed under convolution. That is, the convolution of two Cauchy density functions is another Cauchy density function.

Now let $\xi, \eta \sim f_a$ be two independent Cauchy random variables. (Not identical, two separate and independent ones.) Define $X = \alpha\xi + \beta\eta$ and $Y = \gamma\xi + \delta\eta$. Since X and Y

are functionally related, they are not independent random variables. However, because the Cauchy family is closed under convolutions,

$$X+Y \sim f_{(\alpha+\beta+\delta+\gamma)a} = f_{(\alpha+\beta)a} \star f_{(\gamma+\delta)a} = f_X \star f_Y,$$

and so the converse does not hold.

6.3. SUB-INDEPENDENCE. Studying counterexamples and the conditions under which certain theorems fail can lead to entire new subtopics of research in mathematics. One such example is born directly from the counterexample shown above. We may not get independence if the convolution relation

$$F_{X_1+X_2} = F_{X_1} \star F_{X_2}$$

holds, but what do we still have?

Durairajan [1] in 1979 defined two random variables to be **sub-independent** if

$$F_{X_1+X_2}(z) = F_{X_1} \star F_{X_2}(z)$$

for $z \in \mathbb{R}$. Hamedani [2] in 2013 gave an equivalent definition at the "probability level". For the discrete case,

Definition 6.2 (Sub-independent (Hamedani)–discrete)**.** The discrete random variables X and Y are sub-independent if for every $z \in \mathscr{S}_{X+Y}$,

$$P(A^z) = \sum_{i,j} \sum_{x_i+x_j=z} P(A_i)P(B_j)$$

where $A_i = \{\omega \in \Omega : X(\omega) = x_i\}$, $B_j = \{\omega \in \Omega : Y(\omega) = y_j\}$, and $A^z = \{\omega \in \Omega : X(\omega)+Y(\omega) = z\}$.

Definition 6.3 (Sub-independent (Hamedani)–continuous)**.** Continuous random variables X and Y if for every $c \in \mathbb{R}$,

$$P(A_c) = \sum_{i=1}^{\infty} P\left(A_i^{(c)}\right) P\left(B_i^{(c)}\right),$$

where $A_c = \{\omega \in \Omega : X(\omega)+Y(\omega) < c\}$, $A_i^{(c)} = \{\omega \in \Omega : X(\omega) - \frac{c}{2} \in E_i\}$ and $B_i^{(c)} = \{\omega \in \Omega : X(\omega) - \frac{c}{2} \in F_i\}$, where E_i and F_i are intervals.

There's a lot of formality in these definitions, so we won't worry too much about them for the purpose of this article. In his 2015 monograph, Hamedani and Maadooliat [2] study random variables that are "not quite independent", but rather sub-independent, with an aim of determining what sorts of properties and theorems we still have or can modify. In particular, many limit theorems in probability such as the famous Central Limit Theorem rely on independence of the random variables involved. Hamedani has given a formulation of the Central Limit Theorem that does not require full independence, but rather only sub-independence, among other properties and characteristics studied. Exploration of this topic will be deferred to a separate article.

§7. CONCLUSION

This lecture explored the concept of independence in probability theory and some of the limitations. We did not cover conditional independence, as we feel this deserves its own discussion following this one. We gave relevant definitions of independence (pairwise and mutual) for random events and random variables and explored the relationship between pairwise and mutual independence. We also explored the notion of totally dependent spaces, and found conditions under which they can and cannot be constructed. We found that independence has limitations, namely that independence of classes doesn't guarantee independent σ-fields generated by such classes without additional conditions, and that independence of functions of random variables do not imply independence of the underlying variables without additional conditions. Characteristic functions and convolutions of distributions and their uses in characterizing independence were also discussed. Finally, we gave a very brief overview of the work in sub-independence by Hamedani [2].

References

[1] Durairajan, T. "A classroom note on subindependence". In: *Gujarat Statistical Review* 6 (1979), pp. 17–18.
[2] Hamedani, G. and Maadooliat, M. *Sub-Independence: A Useful Concept*. Nova Science Publishers, 2015.
[3] Hogg, R., McKean, J., and Craig, A. *Introduction to Mathematical Statistics, 6th ed.* Pearson, 2005.
[4] Jacod, J. and Protter, P. *Probability Essentials*. Springer, 2004.
[5] Stoyanov, J. *Counterexamples in Probability, 3rd ed.* Dover, 2013.

