

2.9: Probability Spaces Revisited

In this section we discuss probability spaces from the more advanced point of view of measure theory. The previous two sections on positive measures and existence and uniqueness are prerequisites. The discussion is divided into two parts: first those concepts that are shared rather equally between probability theory and general measure theory, and second those concepts that are for the most part unique to probability theory. In particular, it's a mistake to think of probability theory as a mere branch of measure theory. Probability has its own notation, terminology, point of view, and applications that makes it an incredibly rich subject on its own.

Basic Concepts

Our first discussion concerns topics that were discussed in the section on positive measures. So no proofs are necessary, but you will notice that the notation, and in some cases the terminology, is very different.

Definitions

We can now give a precise definition of the probability space, the mathematical model of a random experiment.

A *probability space* $(S, \mathcal{S}, \mathbb{P})$, consists of three essential parts:

1. A set of outcomes S .
2. A σ -algebra of events \mathcal{S} .
3. A probability measure \mathbb{P} on the *sample space* (S, \mathcal{S}) .

Often the special notation $(\Omega, \mathcal{F}, \mathbb{P})$ is used for a probability space in the literature—the symbol Ω for the set of outcomes is intended to remind us that these are *all* possible outcomes. However in this text, we don't insist on the special notation, and use whatever notation seems most appropriate in a given context.

In probability, σ -algebras are not just important for theoretical and foundational purposes, but are important for practical purposes as well. A σ -algebra can be used to specify *partial information* about an experiment—a concept of fundamental importance. Specifically, suppose that \mathcal{A} is a collection of events in the experiment, and that we know whether or not A occurred for each $A \in \mathcal{A}$. Then in fact, we can determine whether or not A occurred for each $A \in \sigma(\mathcal{A})$, the σ -algebra generated by \mathcal{A} .

Technically, a random variable for our experiment is a *measurable* function from the sample space into another measurable space.

Suppose that $(S, \mathcal{S}, \mathbb{P})$ is a probability space and that (T, \mathcal{T}) is another measurable space. A *random variable* X with values in T is a measurable function from S into T .

1. The *probability distribution* of X is the mapping on \mathcal{T} given by $B \mapsto \mathbb{P}(X \in B)$.
2. The collection of events $\{X \in B : B \in \mathcal{T}\}$ is a sub σ -algebra of \mathcal{S} , and is the σ -algebra generated by X , denoted $\sigma(X)$.

Details

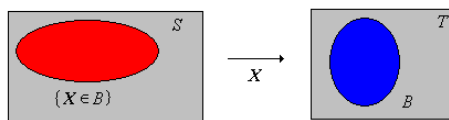


Figure 2.9.1: The event $\{X \in B\}$ associated with $B \in \mathcal{T}$

If we observe the value of X , then we know whether or not each event in $\sigma(X)$ has occurred. More generally, we can construct the σ -algebra associated with any collection of random variables.

suppose that (T_i, \mathcal{T}_i) is a measurable space for each i in an index set I , and that X_i is a random variable taking values in T_i for each $i \in I$. The σ -algebra generated by $\{X_i : i \in I\}$ is

$$\sigma\{X_i : i \in I\} = \sigma\{X \in B_i : B_i \in \mathcal{T}_i, i \in I\} \quad (2.9.1)$$

If we observe the value of X_i for each $i \in I$ then we know whether or not each event in $\sigma\{X_i : i \in I\}$ has occurred. This idea is very important in the study of stochastic processes.

Null Events, Almost Sure Events, and Equivalence

Suppose that $(S, \mathcal{S}, \mathbb{P})$ is a probability space.

Define the following collections of events:

1. $\mathcal{N} = \{A \in \mathcal{S} : \mathbb{P}(A) = 0\}$, the collection of *null* events
2. $\mathcal{M} = \{A \in \mathcal{S} : \mathbb{P}(A) = 1\}$, the collection of *almost sure* events
3. $\mathcal{D} = \mathcal{N} \cup \mathcal{M} = \{A \in \mathcal{S} : \mathbb{P}(A) = 0 \text{ or } \mathbb{P}(A) = 1\}$, the collection of *essentially deterministic* events

The collection of essentially deterministic events \mathcal{D} is a sub σ -algebra of \mathcal{S} .

In the section on independence, we showed that \mathcal{D} is also a collection of independent events.

Intuitively, equivalent events or random variables are those that are indistinguishable from a probabilistic point of view. Recall first that the *symmetric difference* between events A and B is $A \triangle B = (A \setminus B) \cup (B \setminus A)$; it is the event that occurs if and only if one of the events occurs, but not the other, and corresponds to *exclusive or*. Here is the definition for events:

Events A and B are *equivalent* if $A \triangle B \in \mathcal{N}$, and we denote this by $A \equiv B$. The relation \equiv is an equivalence relation on \mathcal{S} . That is, for $A, B, C \in \mathcal{S}$,

1. $A \equiv A$ (the *reflexive property*).
2. If $A \equiv B$ then $B \equiv A$ (the *symmetric property*).
3. If $A \equiv B$ and $B \equiv C$ then $A \equiv C$ (the *transitive property*).

Thus $A \equiv B$ if and only if $\mathbb{P}(A \triangle B) = \mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A) = 0$ if and only if $\mathbb{P}(A \setminus B) = \mathbb{P}(B \setminus A) = 0$. The equivalence relation \equiv partitions \mathcal{S} into disjoint classes of mutually equivalent events. Equivalence is preserved under the set operations.

Suppose that $A, B \in \mathcal{S}$. If $A \equiv B$ then $A^c \equiv B^c$.

Suppose that $A_i, B_i \in \mathcal{S}$ for i in a countable index set I . If $A_i \equiv B_i$ for $i \in I$ then

1. $\bigcup_{i \in I} A_i \equiv \bigcup_{i \in I} B_i$
2. $\bigcap_{i \in I} A_i \equiv \bigcap_{i \in I} B_i$

Equivalent events have the same probability.

If $A, B \in \mathcal{S}$ and $A \equiv B$ then $\mathbb{P}(A) = \mathbb{P}(B)$.

The converse trivially fails, and a [counterexample](#) is given below. However, the null and almost sure events do form equivalence classes.

Suppose that $A \in \mathcal{S}$.

1. If $A \in \mathcal{N}$ then $A \equiv B$ if and only if $B \in \mathcal{N}$.
2. If $A \in \mathcal{M}$ then $A \equiv B$ if and only if $B \in \mathcal{M}$.

We can extend the notion of equivalence to random variables taking values in the same space. Thus suppose that (T, \mathcal{T}) is another measurable space. If X and Y are random variables with values in T , then (X, Y) is a random variable with values in $T \times T$, which is given the usual product σ -algebra $\mathcal{T} \otimes \mathcal{T}$. We assume that the diagonal set $D = \{(x, x) : x \in T\} \in \mathcal{T} \otimes \mathcal{T}$, which is almost always true in applications.

Random variables X and Y taking values in T are *equivalent* if $\mathbb{P}(X = Y) = 1$. Again we write $X \equiv Y$. The relation \equiv is an equivalence relation on the collection of random variables that take values in T . That is, for random variables X, Y , and Z with values in T ,

1. $X \equiv X$ (the *reflexive property*).
2. If $X \equiv Y$ then $Y \equiv X$ (the *symmetric property*).
3. If $X \equiv Y$ and $Y \equiv Z$ then $X \equiv Z$ (the *transitive property*).

So the collection of random variables with values in T is partitioned into disjoint classes of mutually equivalent variables.

Suppose that X and Y are random variables taking values in T and that $X \equiv Y$. Then

1. $\{X \in B\} \equiv \{Y \in B\}$ for every $B \in \mathcal{T}$.
2. X and Y have the same probability distribution on (T, \mathcal{T}) .

Again, the converse to part (b) fails with a passion, and a [counterexample](#) is given below. It often happens that a definition for random variables subsumes the corresponding definition for events, by considering the indicator variables of the events. So it is with equivalence.

Suppose that $A, B \in \mathcal{S}$. Then $A \equiv B$ if and only if $\mathbf{1}_A \equiv \mathbf{1}_B$.

Equivalence is preserved under a deterministic transformation of the variables. For the next result, suppose that (U, \mathcal{U}) is yet another measurable space, along with (T, \mathcal{T}) .

Suppose X, Y are random variables with values in T and that $g: T \rightarrow U$ is measurable. If $X \equiv Y$ then $g(X) \equiv g(Y)$.

Suppose again that $(S, \mathcal{S}, \mathbb{P})$ is a probability space corresponding to a random experiment. Let \mathcal{V} denote the collection of all real-valued random variables for the experiment, that is, all measurable functions from S into \mathbb{R} . With the usual definitions of addition and scalar multiplication, $(\mathcal{V}, +, \cdot)$ is a vector space. However, in probability theory, we often do not want to distinguish between random variables that are equivalent, so it's nice to know that the vector space structure is preserved when we *identify* equivalent random variables. Formally, let $[X]$ denote the equivalence class generated by a real-valued random variable $X \in \mathcal{V}$, and let \mathcal{W} denote the collection of all such equivalence classes. In modular notation, \mathcal{W} is the set \mathcal{V} / \equiv . We define addition and scalar multiplication on \mathcal{W} by

$$[X] + [Y] = [X + Y], \quad c[X] = [cX]; \quad [X], [Y] \in \mathcal{W}, \quad c \in \mathbb{R} \quad (2.9.2)$$

$(\mathcal{W}, +, \cdot)$ is a vector space.

Often we don't bother to use the special notation for the equivalence class associated with a random variable. Rather, it's understood that equivalent random variables represent the same object. Spaces of functions in a general measure space are studied in the chapter on Distributions, and spaces of random variables are studied in more detail in the chapter on Expected Value.

Completion

Suppose again that $(S, \mathcal{S}, \mathbb{P})$ is a probability space, and that \mathcal{N} denotes the collection of null events, as above. Suppose that $A \in \mathcal{N}$ so that $\mathbb{P}(A) = 0$. If $B \subseteq A$ and $B \in \mathcal{S}$, then we know that $\mathbb{P}(B) = 0$ so $B \in \mathcal{N}$ also. However, in general there might be subsets of A that are not in \mathcal{S} . This leads naturally to the following definition.

The probability space $(S, \mathcal{S}, \mathbb{P})$ is *complete* if $A \in \mathcal{N}$ and $B \subseteq A$ imply $B \in \mathcal{S}$ (and hence $B \in \mathcal{N}$).

So the probability space is complete if every subset of an event with probability 0 is also an event (and hence also has probability 0). We know from our work on positive measures that every σ -finite measure space that is not complete can be completed. So in particular a probability space that is not complete can be completed. To review the construction, recall that the equivalence relation \equiv that we used above on \mathcal{S} is extended to $\mathcal{P}(S)$ (the power set of S).

For $A, B \subseteq S$, define $A \equiv B$ if and only if there exists $N \in \mathcal{N}$ such that $A \triangle B \subseteq N$. The relation \equiv is an equivalence relation on $\mathcal{P}(S)$.

Here is how the probability space is completed:

Let $\mathcal{S}_0 = \{A \subseteq S : A \equiv B \text{ for some } B \in \mathcal{S}\}$. For $A \in \mathcal{S}_0$, define $\mathbb{P}_0(A) = \mathbb{P}(B)$ where $B \in \mathcal{S}$ and $A \equiv B$. Then

1. \mathcal{S}_0 is a σ -algebra of subsets of S and $\mathcal{S} \subseteq \mathcal{S}_0$.
2. \mathbb{P}_0 is a probability measure on (S, \mathcal{S}_0) .
3. $(S, \mathcal{S}_0, \mathbb{P}_0)$ is complete, and is the *completion* of $(S, \mathcal{S}, \mathbb{P})$.

Product Spaces

Our next discussion concerns the construction of probability spaces that correspond to specified distributions. To set the stage, suppose that $(S, \mathcal{S}, \mathbb{P})$ is a probability space. If we let X denote the identity function on S , so that $X(x) = x$ for $x \in S$, then $\{X \in A\} = A$ for $A \in \mathcal{S}$ and hence $\mathbb{P}(X \in A) = \mathbb{P}(A)$. That is, \mathbb{P} is the probability distribution of X . We have seen this before—every probability measure can be thought of as the distribution of a random variable. The next result shows how to construct a probability space that corresponds to a sequence of independent random variables with specified distributions.

Suppose $n \in \mathbb{N}_+$ and that $(S_i, \mathcal{S}_i, \mathbb{P}_i)$ is a probability space for $i \in \{1, 2, \dots, n\}$. The corresponding product measure space $(S, \mathcal{S}, \mathbb{P})$ is a probability space. If $X_i: S \rightarrow S_i$ is the i th coordinate function on S so that $X_i(\mathbf{x}) = x_i$ for $\mathbf{x} = (x_1, x_2, \dots, x_n) \in S$ then (X_1, X_2, \dots, X_n) is a sequence of independent random variables on $(S, \mathcal{S}, \mathbb{P})$, and X_i has distribution \mathbb{P}_i on (S_i, \mathcal{S}_i) for each $i \in \{1, 2, \dots, n\}$.

Proof

Of course, the existence of the product space $(S, \mathcal{S}, \mathbb{P})$ follows immediately from the more general result for products of positive measure spaces. Recall that $S = \prod_{i=1}^n S_i$ and that \mathcal{S} is the σ -algebra generated by sets of the form $\prod_{i=1}^n A_i$ where $A_i \in \mathcal{S}_i$ for each $i \in \{1, 2, \dots, n\}$. Finally, \mathbb{P} is the unique positive measure on (S, \mathcal{S}) satisfying

$$\mathbb{P}\left(\prod_{i=1}^n A_i\right) = \prod_{i=1}^n \mathbb{P}_i(A_i) \quad (2.9.3)$$

where again, $A_i \in \mathcal{S}_i$ for each $i \in \{1, 2, \dots, n\}$. Clearly \mathbb{P} is a probability measure since $\mathbb{P}(S) = \prod_{i=1}^n \mathbb{P}_i(S_i) = 1$. Suppose that $A_i \in \mathcal{S}_i$ for $i \in \{1, 2, \dots, n\}$. Then $\{X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n\} = \prod_{i=1}^n A_i \in \mathcal{S}$. Hence

$$\mathbb{P}(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) = \prod_{i=1}^n \mathbb{P}_i(A_i) \quad (2.9.4)$$

If we fix $i \in \{1, 2, \dots, n\}$ and let $A_j = S_j$ for $j \neq i$, then the displayed equation give $\mathbb{P}(X_i \in A_i) = \mathbb{P}_i(A_i)$, so X_i has distribution \mathbb{P}_i on (S_i, \mathcal{S}_i) . Returning to the displayed equation we have

$$\mathbb{P}(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i) \quad (2.9.5)$$

so (X_1, X_2, \dots, X_n) are independent.

Intuitively, the given probability spaces correspond to n random experiments. The product space then is the probability space that corresponds to the experiments performed independently. When modeling a random experiment, if we say that we have a finite sequence of independent random variables with specified distributions, we can rest assured that there actually is a probability space that supports this statement

We can extend the last result to an infinite sequence of probability spaces. Suppose that (S_i, \mathcal{S}_i) is a measurable space for each $i \in \mathbb{N}_+$. Recall that the product space $\prod_{i=1}^{\infty} S_i$ consists of all sequences $\mathbf{x} = (x_1, x_2, \dots)$ such that $x_i \in S_i$ for each $i \in \mathbb{N}_+$. The corresponding product σ -algebra \mathcal{S} is generated by the collection of *cylinder sets*. That is, $\mathcal{S} = \sigma(\mathcal{B})$ where

$$\mathcal{B} = \left\{ \prod_{i=1}^{\infty} A_i : A_i \in \mathcal{S}_i \text{ for each } i \in \mathbb{N}_+ \text{ and } A_i = S_i \text{ for all but finitely many } i \in \mathbb{N}_+ \right\} \quad (2.9.6)$$

Suppose that $(S_i, \mathcal{S}_i, \mathbb{P}_i)$ is a probability space for $i \in \mathbb{N}_+$. Let (S, \mathcal{S}) denote the product measurable space so that $\mathcal{S} = \sigma(\mathcal{B})$ where \mathcal{B} is the collection of cylinder sets. Then there exists a unique probability measure \mathbb{P} on (S, \mathcal{S}) that satisfies

$$\mathbb{P}\left(\prod_{i=1}^{\infty} A_i\right) = \prod_{i=1}^{\infty} \mathbb{P}_i(A_i), \quad \prod_{i=1}^{\infty} A_i \in \mathcal{B} \quad (2.9.7)$$

If $X_i : S \rightarrow S_i$ is the i th coordinate function on S for $i \in \mathbb{N}_+$, so that $X_i(\mathbf{x}) = x_i$ for $\mathbf{x} = (x_1, x_2, \dots) \in S$, then (X_1, X_2, \dots) is a sequence of independent random variables on $(S, \mathcal{S}, \mathbb{P})$, and X_i has distribution \mathbb{P}_i on (S_i, \mathcal{S}_i) for each $i \in \mathbb{N}_+$.

Proof

The proof is similar to the one in for positive measure spaces in the section on existence and uniqueness. First recall that the collection of cylinder sets \mathcal{B} is a semi-algebra. We define $\mathbb{P} : \mathcal{B} \rightarrow [0, 1]$ as in the statement of the theorem. Note that all but finitely many factors are 1. The consistency conditions are satisfied, so \mathbb{P} can be extended to a probability measure on the algebra \mathcal{A} generated by \mathcal{B} . That is, \mathcal{A} is the collection of all finite, disjoint unions of cylinder sets. The standard extension theorem and uniqueness theorem now apply, so \mathbb{P} can be extended uniquely to a measure on $\mathcal{S} = \sigma(\mathcal{A})$. The proof that (X_1, X_2, \dots) are independent and that X_i has distribution \mathbb{P}_i for each $i \in \mathbb{N}_+$ is just as in the previous theorem.

Once again, if we model a random process by starting with an infinite sequence of independent random variables, we can be sure that there exists a probability space that supports this sequence. The particular probability space constructed in the last theorem is called the *canonical probability space* associated with the sequence of random variables. Note also that it was important that we had *probability measures* rather than just general positive measures in the construction, since the infinite product $\prod_{i=1}^{\infty} \mathbb{P}_i(A_i)$ is always well defined. The next section on Stochastic Processes continues the discussion of how to construct probability spaces that correspond to a collection of random variables with specified distributional properties.

Probability Concepts

Our next discussion concerns topics that are unique to probability theory and do not have simple analogies in general measure theory.

Independence

As usual, suppose that $(S, \mathcal{S}, \mathbb{P})$ is a probability space. We have already studied the independence of collections of events and the independence of collections of random variables. A more complete and general treatment results if we define the independence of collections of *collections* of events, and most importantly, the independence of collections of σ -algebras. This extension actually occurred already, when we went from independence of a collection of events to independence of a collection of random variables, but we did not note it at the time. In spite of the layers of set theory, the basic idea is the same.

Suppose that \mathcal{A}_i is a collection of events for each i in an index set I . Then $\mathcal{A} = \{\mathcal{A}_i : i \in I\}$ is independent if and only if for every choice of $A_i \in \mathcal{A}_i$ for $i \in I$, the collection of events $\{A_i : i \in I\}$ is independent. That is, for every finite $J \subseteq I$,

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j) \quad (2.9.8)$$

As noted above, independence of random variables, as we defined previously, is a special case of our new definition.

Suppose that (T_i, \mathcal{T}_i) is a measurable space for each i in an index set I , and that X_i is a random variable taking values in a set T_i for each $i \in I$. The independence of $\{X_i : i \in I\}$ is equivalent to the independence of $\{\sigma(X_i) : i \in I\}$.

Independence of events is also a special case of the new definition, and thus our new definition really does subsume our old one.

Suppose that A_i is an event for each $i \in I$. The independence of $\{A_i : i \in I\}$ is equivalent to the independence of $\{\mathcal{A}_i : i \in I\}$ where $\mathcal{A}_i = \sigma\{A_i\} = \{S, \emptyset, A_i, A_i^c\}$ for each $i \in I$.

For every collection of objects that we have considered (collections of events, collections of random variables, collections of collections of events), the notion of independence has the basic inheritance property.

Suppose that \mathcal{A} is a collection of collections of events.

1. If \mathcal{A} is independent then \mathcal{B} is independent for every $\mathcal{B} \subseteq \mathcal{A}$.
2. If \mathcal{B} is independent for every finite $\mathcal{B} \subseteq \mathcal{A}$ then \mathcal{A} is independent.

Our most important collections are σ -algebras, and so we are most interested in the independence of a collection of σ -algebras. The next result allows us to go from the independence of certain types of collections to the independence of the σ -algebras generated by these collections. To understand the result, you will need to review the definitions and theorems concerning π -systems and λ -systems. The proof uses Dynkin's π - λ theorem, named for Eugene Dynkin.

Suppose that \mathcal{A}_i is a collection of events for each i in an index set I , and that \mathcal{A}_i is a π -system for each $i \in I$. If $\{\mathcal{A}_i : i \in I\}$ is independent, then $\{\sigma(\mathcal{A}_i) : i \in I\}$ is independent.

Proof

In light of the previous result, it suffices to consider a finite set of collections. Thus, suppose that $\{\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_n\}$ is independent. Now, fix $A_i \in \mathcal{A}_i$ for $i \in \{2, 3, \dots, n\}$ and let $E = \bigcap_{i=2}^n A_i$. Let $\mathcal{L} = \{B \in \mathcal{S} : \mathbb{P}(B \cap E) = \mathbb{P}(B)\mathbb{P}(E)\}$. Trivially $S \in \mathcal{L}$ since $\mathbb{P}(S \cap E) = \mathbb{P}(E) = \mathbb{P}(S)\mathbb{P}(E)$. Next suppose that $A \in \mathcal{L}$. Then

$$\mathbb{P}(A^c \cap E) = \mathbb{P}(E) - \mathbb{P}(A \cap E) = \mathbb{P}(E) - \mathbb{P}(A)\mathbb{P}(E) = [1 - \mathbb{P}(A)]\mathbb{P}(E) = \mathbb{P}(A^c)\mathbb{P}(E) \quad (2.9.9)$$

Thus $A^c \in \mathcal{L}$. Finally, suppose that $\{A_j : j \in J\}$ is a countable collection of disjoint sets in \mathcal{L} . Then

$$\mathbb{P}\left[\left(\bigcup_{j \in J} A_j\right) \cap E\right] = \mathbb{P}\left[\bigcup_{j \in J} (A_j \cap E)\right] = \sum_{j \in J} \mathbb{P}(A_j \cap E) = \sum_{j \in J} \mathbb{P}(A_j)\mathbb{P}(E) = \mathbb{P}(E) \sum_{j \in J} \mathbb{P}(A_j) = \mathbb{P}(E)\mathbb{P}\left(\bigcup_{j \in J} A_j\right) \quad (2.9.10)$$

Therefore $\bigcup_{j \in J} A_j \in \mathcal{L}$ and so \mathcal{L} is a λ -system. Trivially $\mathcal{A}_1 \subseteq \mathcal{L}$ by the original independence assumption, so by the π - λ theorem, $\sigma(\mathcal{A}_1) \subseteq \mathcal{L}$. Thus, we have that for every $A_1 \in \sigma(\mathcal{A}_1)$ and $A_i \in \mathcal{A}_i$ for $i \in \{2, 3, \dots, n\}$,

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n \mathbb{P}(A_i) \quad (2.9.11)$$

Thus we have shown that $\{\sigma(\mathcal{A}_1), \mathcal{A}_2, \dots, \mathcal{A}_n\}$ is independent. Repeating the argument $n-1$ additional times, we get that $\{\sigma(\mathcal{A}_1), \sigma(\mathcal{A}_2), \dots, \sigma(\mathcal{A}_n)\}$ is independent.

The next result is a rigorous statement of the strong independence that is implied the independence of a collection of events.

Suppose that \mathcal{A} is an independent collection of events, and that $\{\mathcal{B}_j : j \in J\}$ is a partition of \mathcal{A} . That is, $\mathcal{B}_j \cap \mathcal{B}_k = \emptyset$ for $j \neq k$ and $\bigcup_{j \in J} \mathcal{B}_j = \mathcal{A}$. Then $\{\sigma(\mathcal{B}_j) : j \in J\}$ is independent.

Proof

Let \mathcal{B}_j^* denote the set of all finite intersections of sets in \mathcal{B}_j , for each $j \in J$. Then clearly \mathcal{B}_j^* is a π -system for each j , and $\{\mathcal{B}_j^* : j \in J\}$ is independent. By the previous theorem, $\{\sigma(\mathcal{B}_j^*) : j \in J\}$ is independent. But clearly $\sigma(\mathcal{B}_j^*) = \sigma(\mathcal{B}_j)$ for $j \in J$.

Let's bring the result down to earth. Suppose that A, B, C, D are independent events. In our elementary discussion of independence, you were asked to show, for example, that $A \cup B^c$ and $C^c \cup D^c$ are independent. This is a consequence of the much stronger statement that the σ -algebras $\sigma\{A, B\}$ and $\sigma\{C, D\}$ are independent.

Exchangeability

As usual, suppose that $(S, \mathcal{S}, \mathbb{P})$ is a probability space corresponding to a random experiment. Roughly speaking, a sequence of events or a sequence of random variables is *exchangeable* if the probability law that governs the sequence is unchanged when the order of the events or variables is changed. Exchangeable variables arise naturally in sampling experiments and many other settings, and are a natural generalization of a sequence of independent, identically distributed (IID) variables. Conversely, it turns out that any exchangeable sequence of variables can be constructed from an IID sequence. First we give the definition for events:

Suppose that $\mathcal{A} = \{A_i : i \in I\}$ is a collection of events, where I is a nonempty index set. Then \mathcal{A} is *exchangeable* if the probability of the intersection of a finite number of the events depends only on the number of events. That is, if J and K are finite subsets of I and $\#(J) = \#(K)$ then

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \mathbb{P}\left(\bigcap_{k \in K} A_k\right) \quad (2.9.12)$$

Exchangeability has the same basic inheritance property that we have seen before.

Suppose that \mathcal{A} is a collection of events.

1. If \mathcal{A} is exchangeable then \mathcal{B} is exchangeable for every $\mathcal{B} \subseteq \mathcal{A}$.
2. Conversely, if \mathcal{B} is exchangeable for every finite $\mathcal{B} \subseteq \mathcal{A}$ then \mathcal{A} is exchangeable.

For a collection of exchangeable events, the inclusion exclusion law for the probability of a union is much simpler than the general version.

Suppose that $\{A_1, A_2, \dots, A_n\}$ is an exchangeable collection of events. For $J \subseteq \{1, 2, \dots, n\}$ with $\#(J) = k$, let $p_k = \mathbb{P}\left(\bigcap_{j \in J} A_j\right)$. Then

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n (-1)^{k-1} \binom{n}{k} p_k \quad (2.9.13)$$

Proof

The inclusion-exclusion rule gives

$$\mathbb{P}\left(\bigcup_{i \in I} A_i\right) = \sum_{k=1}^n (-1)^{k-1} \sum_{J \subseteq I, \#(J)=k} \mathbb{P}\left(\bigcap_{j \in J} A_j\right) \quad (2.9.14)$$

But $p_k = \mathbb{P}\left(\bigcap_{j \in J} A_j\right)$ for every $J \subseteq \{1, 2, \dots, n\}$ with $\#(J) = k$, and there are $\binom{n}{k}$ such subsets.

The concept of exchangeability can be extended to random variables in the natural way. Suppose that (T, \mathcal{T}) is a measurable space.

Suppose that \mathcal{A} is a collection of random variables, each taking values in T . The collection \mathcal{A} is *exchangeable* if for any $\{X_1, X_2, \dots, X_n\} \subseteq \mathcal{A}$, the distribution of the random vector (X_1, X_2, \dots, X_n) depends only on n .

Thus, the distribution of the random vector is unchanged if the coordinates are permuted. Once again, exchangeability has the same basic inheritance property as a collection of independent variables.

Suppose that \mathcal{A} is a collection of random variables, each taking values in T .

1. If \mathcal{A} is exchangeable then \mathcal{B} is exchangeable for every $\mathcal{B} \subseteq \mathcal{A}$.
2. Conversely, if \mathcal{B} is exchangeable for every finite $\mathcal{B} \subseteq \mathcal{A}$ then \mathcal{A} is exchangeable.

Suppose that \mathcal{A} is a collection of random variables, each taking values in T , and that \mathcal{A} is exchangeable. Then trivially the variables are identically distributed: if $X, Y \in \mathcal{A}$ and $A \in \mathcal{T}$, then $\mathbb{P}(X \in A) = \mathbb{P}(Y \in A)$. Also, the definition of exchangeable variables subsumes the definition for events:

Suppose that \mathcal{A} is a collection of events, and let $\mathcal{B} = \{\mathbf{1}_A : A \in \mathcal{A}\}$ denote the corresponding collection of indicator random variables. Then \mathcal{A} is exchangeable if and only if \mathcal{B} is exchangeable.

Tail Events and Variables

Suppose again that we have a random experiment modeled by a probability space $(S, \mathcal{S}, \mathbb{P})$.

Suppose that (X_1, X_2, \dots) be a sequence of random variables. The *tail sigma algebra* of the sequence is

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma\{X_n, X_{n+1}, \dots\} \quad (2.9.15)$$

1. An event $B \in \mathcal{T}$ is a *tail event* for the sequence.
2. A random variable Y that is measurable with respect to \mathcal{T} is a *tail random variable* for the sequence.

Informally, a tail event (random variable) is an event (random variable) that can be defined in terms of $\{X_n, X_{n+1}, \dots\}$ for each $n \in \mathbb{N}_+$. The tail sigma algebra for a sequence of events (A_1, A_2, \dots) is defined analogously (or simply let $X_k = \mathbf{1}(A_k)$, the indicator variable of A , for each k). For the following results, you may need to review some of the definitions in the section on Convergence.

Suppose that (A_1, A_2, \dots) is a sequence of events.

1. If the sequence is increasing then $\lim_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} A_n$ is a tail event of the sequence.
2. If the sequence is decreasing then $\lim_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} A_n$ is a tail event of the sequence.

Proof

1. If the sequence is increasing then $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=k}^{\infty} A_n \in \sigma\{A_k, A_{k+1}, \dots\}$ for every $k \in \mathbb{N}_+$.
2. If the sequence is decreasing then $\bigcap_{n=1}^{\infty} A_n = \bigcap_{n=k}^{\infty} A_n \in \sigma\{A_k, A_{k+1}, \dots\}$ for every $k \in \mathbb{N}_+$.

Suppose again that (A_1, A_2, \dots) is a sequence of events. Each of the following is a tail event of the sequence:

1. $\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i$
2. $\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} A_i$

Proof

1. The events $\bigcup_{i=n}^{\infty} A_i$ are decreasing in n and hence $\limsup_{n \rightarrow \infty} A_n = \lim_{n \rightarrow \infty} \bigcup_{i=n}^{\infty} A_i \in \mathcal{T}$ by the previous result.
2. The events $\bigcap_{i=n}^{\infty} A_i$ are increasing in n and hence $\liminf_{n \rightarrow \infty} A_n = \lim_{n \rightarrow \infty} \bigcap_{i=n}^{\infty} A_i \in \mathcal{T}$ by the previous result.

Suppose that $\mathbf{X} = (X_1, X_2, \dots)$ is a sequence of real-valued random variables.

1. $\{X_n \text{ converges as } n \rightarrow \infty\}$ is a tail event for \mathbf{X} .
2. $\liminf_{n \rightarrow \infty} X_n$ is a tail random variable for \mathbf{X} .
3. $\limsup_{n \rightarrow \infty} X_n$ is a tail random variable for \mathbf{X} .

Proof

1. The Cauchy criterion for convergence (named for Augustin Cauchy of course) states that X_n converges as $n \rightarrow \infty$ if and only if for every $\epsilon > 0$ there exists $N \in \mathbb{N}_+$ (depending on ϵ) such that if $m, n \geq N$ then $|X_n - X_m| < \epsilon$. In this criterion, we can without loss of generality take ϵ to be rational, and for a given $k \in \mathbb{N}_+$ we can insist that $m, n \geq k$. With these restrictions, the Cauchy criterion is a countable intersection of events, each of which is in $\sigma\{X_k, X_{k+1}, \dots\}$.
2. Recall that $\liminf_{n \rightarrow \infty} X_n = \lim_{n \rightarrow \infty} \inf\{X_k : k \geq n\}$.
3. Similarly, recall that $\limsup_{n \rightarrow \infty} X_n = \lim_{n \rightarrow \infty} \sup\{X_k : k \geq n\}$.

The random variable in part (b) may take the value $-\infty$, and the random variable in (c) may take the value ∞ . From parts (b) and (c) together, note that if $X_n \rightarrow X_\infty$ as $n \rightarrow \infty$ on the sample space \mathcal{S} , then X_∞ is a tail random variable for \mathbf{X} .

There are a number of *zero-one laws* in probability. These are theorems that give conditions under which an event will be essentially deterministic; that is, have probability 0 or probability 1. Interestingly, it can sometimes be difficult to determine which of these extremes is actually the case. The following result is the *Kolmogorov zero-one law*, named for Andrey Kolmogorov. It states that an event in the tail σ -algebra of an *independent* sequence will have probability 0 or 1.

Suppose that $\mathbf{X} = (X_1, X_2, \dots)$ is an independent sequence of random variables

1. If B is a tail event for \mathbf{X} then $\mathbb{P}(B) = 0$ or $\mathbb{P}(B) = 1$.
2. If Y is a real-valued tail random variable for \mathbf{X} then Y is constant with probability 1.

Proof

1. By definition $B \in \sigma\{X_{n+1}, X_{n+2}, \dots\}$ for each $n \in \mathbb{N}_+$, and hence $\{X_1, X_2, \dots, X_n, \mathbf{1}_B\}$ is an independent set of random variables. Thus $\{X_1, X_2, \dots, \mathbf{1}_B\}$ is an independent set of random variables. But $B \in \sigma\{X_1, X_2, \dots\}$, so it follows that the event B is independent of itself. Therefore $\mathbb{P}(B) = 0$ or $\mathbb{P}(B) = 1$.

2. The function $y \mapsto \mathbb{P}(Y \leq y)$ on \mathbb{R} is the (cumulative) distribution function of Y . This function is clearly increasing. Moreover, simple applications of the continuity theorems show that it is right continuous and that $\mathbb{P}(Y \leq y) \rightarrow 0$ as $y \rightarrow -\infty$ and $\mathbb{P}(Y \leq y) \rightarrow 1$ as $y \rightarrow \infty$. (Explicit proofs are given in the section on distribution functions in the chapter on Distributions.) But since Y is a tail random variable, $\{Y \leq y\}$ is a tail event and hence $\mathbb{P}(Y \leq y) \in \{0, 1\}$ for each $y \in \mathbb{R}$. It follows that there exists $c \in \mathbb{R}$ such that $\mathbb{P}(Y \leq y) = 0$ for $y < c$ and $\mathbb{P}(Y \leq y) = 1$ for $y \geq c$. Hence $\mathbb{P}(Y = c) = 1$.

From the Komogorov zero-one law and the [result above](#), note that if (A_1, A_2, \dots) is a sequence of independent events, then $\limsup_{n \rightarrow \infty} A_n$ must have probability 0 or 1. The Borel-Cantelli lemmas give conditions for which of these is correct:

Suppose that (A_1, A_2, \dots) is a sequence of independent events.

1. If $\sum_{i=1}^{\infty} \mathbb{P}(A_i) < \infty$ then $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 0$.
2. If $\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty$ then $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 1$.

Another proof of the Kolmogorov zero-one law will be given using the martingale convergence theorem.

Examples and Exercises

As always, be sure to try the computational exercises and proofs yourself before reading the answers and proofs in the text.

Counterexamples

Equal probability certainly does not imply equivalent events.

Consider the simple experiment of tossing a fair coin. The event that the coin lands heads and the event that the coin lands tails have the same probability, but are not equivalent.

Proof

Let S denote the sample space, and H the event of heads, so that H^c is the event of tails. Since the coin is fair, $\mathbb{P}(H) = \mathbb{P}(H^c) = \frac{1}{2}$. But $H \triangle H^c = S$, so $\mathbb{P}(H \triangle H^c) = 1$, so H and H^c are as far from equivalent as possible.

Similarly, equivalent distributions does not imply equivalent random variables.

Consider the experiment of rolling a standard, fair die. Let X denote the score and $Y = 7 - X$. Then X and Y have the same distribution but are not equivalent.

Proof

Since the die is fair, X is uniformly distributed on $S = \{1, 2, 3, 4, 5, 6\}$. Also $\mathbb{P}(Y = k) = \mathbb{P}(X = 7 - k) = \frac{1}{6}$ for $k \in S$, so Y also has the uniform distribution on S . But $\mathbb{P}(X = Y) = \mathbb{P}\left(X = \frac{7}{2}\right) = 0$, so X and Y are as far from equivalent as possible.

Consider the experiment of rolling two standard, fair dice and recording the sequence of scores (X, Y) . Then X and Y are independent and have the same distribution, but are not equivalent.

Proof

Since the dice are fair, (X, Y) has the uniform distribution on $\{1, 2, 3, 4, 5, 6\}^2$. Equivalently, X and Y are independent, and each has the uniform distribution on $\{1, 2, 3, 4, 5, 6\}$. But $\mathbb{P}(X = Y) = \frac{1}{6}$, so X and Y are not equivalent.

This page titled [2.9: Probability Spaces Revisited](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist](#) ([Random Services](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.