2.5: Independence

In this section, we will discuss independence, one of the fundamental concepts in probability theory. Independence is frequently invoked as a modeling assumption, and moreover, (classical) probability itself is based on the idea of independent replications of the experiment. As usual, if you are a new student of probability, you may want to skip the technical details.

Basic Theory

As usual, our starting point is a random experiment modeled by a probability space $(S, \mathscr{S}, \mathbb{P})$ so that S is the set of outcomes, $\mathscr{S}$ the collection of events, and $\mathbb{P}$ the probability measure on the sample space $(S, \mathscr{S})$. We will define independence for two events, then for collections of events, and then for collections of random variables. In each case, the basic idea is the same.

Independence of Two Events

Two events A and B are *independent* if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \tag{2.5.1}$$

If both of the events have positive probability, then independence is equivalent to the statement that the conditional probability of one event given the other is the same as the unconditional probability of the event:

$$\mathbb{P}(A \mid B) = \mathbb{P}(A) \iff \mathbb{P}(B \mid A) = \mathbb{P}(B) \iff \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \tag{2.5.2}$$

This is how you should think of independence: knowledge that one event has occurred does not change the probability assigned to the other event. Independence of two events was discussed in the last section in the context of correlation. In particular, for two events, *independent* and *uncorrelated* mean the same thing.

The terms *independent* and *disjoint* sound vaguely similar but they are actually very different. First, note that disjointness is purely a set-theory concept while independence is a probability (measure-theoretic) concept. Indeed, two events can be independent relative to one probability measure and dependent relative to another. But most importantly, two disjoint events can *never* be independent, except in the trivial case that one of the events is null.

Suppose that A and B are disjoint events, each with positive probability. Then A and B are dependent, and in fact are negatively correlated.

Proof

Note that $\mathbb{P}(A \cap B) = \mathbb{P}(\emptyset) = 0$ but $\mathbb{P}(A)\mathbb{P}(B) > 0$.

If A and B are independent events then intuitively it seems clear that any event that can be constructed from A should be independent of any event that can be constructed from B. This is the case, as the next result shows. Moreover, this basic idea is essential for the generalization of independence that we will consider shortly.

If A and B are independent events, then each of the following pairs of events is independent:

1. A^c, B
2. B, A^c
3. A^c, B^c

Proof

Suppose that A and B are independent. Then by the difference rule and the complement rule,

$$\mathbb{P}(A^c \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(B) - \mathbb{P}(A)\,\mathbb{P}(B) = \mathbb{P}(B)\,[1 - \mathbb{P}(A)] = \mathbb{P}(B)\mathbb{P}(A^c) \tag{2.5.3}$$

Hence A^c and B are equivalent. Parts (b) and (c) follow from (a).

An event that is “essentially deterministic”, that is, has probability 0 or 1, is independent of any other event, even itself.

Suppose that A and B are events.

1. If $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$, then A and B are independent.
2. A is independent of itself if and only if $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

Proof

1. Recall that if $\mathbb{P}(A)=0$ then $\mathbb{P}(A\cap B)=0$, and if $\mathbb{P}(A)=1$ then $\mathbb{P}(A\cap B)=\mathbb{P}(B)$. In either case we have $\mathbb{P}(A\cap B)=\mathbb{P}(A)\mathbb{P}(B)$.
2. The independence of A with itself gives $\mathbb{P}(A)=[\mathbb{P}(A)]^2$ and hence either $\mathbb{P}(A)=0$ or $\mathbb{P}(A)=1$.

General Independence of Events

To extend the definition of independence to more than two events, we might think that we could just require *pairwise independence*, the independence of each pair of events. However, this is not sufficient for the *strong* type of independence that we have in mind. For example, suppose that we have three events A, B, and C. *Mutual* independence of these events should not only mean that each pair is independent, but also that an event that can be constructed from A and B (for example $A\cup B^c$) should be independent of C. Pairwise independence does not achieve this; an exercise below gives three events that are pairwise independent, but the intersection of two of the events is related to the third event in the strongest possible sense.

Another possible generalization would be to simply require the probability of the intersection of the events to be the product of the probabilities of the events. However, this condition does not even guarantee pairwise independence. An exercise below gives an example. However, the definition of independence for two events does generalize in a natural way to an arbitrary collection of events.

Suppose that A_i is an event for each i in an index set I. Then the collection $\mathscr{A}=\{A_i: i\in I\}$ is *independent* if for every finite $J\subseteq I$,

$$\mathbb{P}\left(\bigcap_{j\in J}A_j\right)=\prod_{j\in J}\mathbb{P}(A_j) \tag{2.5.4}$$

Independence of a collection of events is much stronger than mere pairwise independence of the events in the collection. The basic *inheritance property* in the following result follows immediately from the definition.

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

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

For a finite collection of events, the number of conditions required for mutual independence grows exponentially with the number of events.

There are 2^n-n-1 non-trivial conditions in the definition of the independence of n events.

1. Explicitly give the 4 conditions that must be satisfied for events A, B, and C to be independent.
2. Explicitly give the 11 conditions that must be satisfied for events A, B, C, and D to be independent.

Answer

There are 2^n subcollections of the n events. One is empty and n involve a single event. The remaining 2^n-n-1 subcollections involve two or more events and correspond to non-trivial conditions.

1. A, B, C are independent if and only if

$$\begin{aligned}\mathbb{P}(A\cap B)&=\mathbb{P}(A)\mathbb{P}(B)\\ \mathbb{P}(A\cap C)&=\mathbb{P}(A)\mathbb{P}(C)\\ \mathbb{P}(B\cap C)&=\mathbb{P}(B)\mathbb{P}(C)\\ \mathbb{P}(A\cap B\cap C)&=\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)\end{aligned}$$

2. A, B, C, D are independent if and only if

$$\mathbb{P}(A\cap B)=\mathbb{P}(A)\mathbb{P}(B)$$
$$\mathbb{P}(A\cap C)=\mathbb{P}(A)\mathbb{P}(C)$$
$$\mathbb{P}(A\cap D)=\mathbb{P}(A)\mathbb{P}(D)$$
$$\mathbb{P}(B\cap C)=\mathbb{P}(B)\mathbb{P}(C)$$
$$\mathbb{P}(B\cap D)=\mathbb{P}(B)\mathbb{P}(D)$$
$$\mathbb{P}(C\cap D)=\mathbb{P}(C)\mathbb{P}(D)$$
$$\mathbb{P}(A\cap B\cap C)=\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$$
$$\mathbb{P}(A\cap B\cap D)=\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(D)$$
$$\mathbb{P}(A\cap C\cap D)=\mathbb{P}(A)\mathbb{P}(C)\mathbb{P}(D)$$
$$\mathbb{P}(B\cap C\cap D)=\mathbb{P}(B)\mathbb{P}(C)\mathbb{P}(D)$$
$$\mathbb{P}(A\cap B\cap C\cap D)=\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)\mathbb{P}(D)$$

If the events $A_1, A_2, \ldots, A_n$ are independent, then it follows immediately from the definition that

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i\right)=\prod_{i=1}^n \mathbb{P}(A_i) \tag{2.5.5}$$

This is known as the *multiplication rule* for independent events. Compare this with the general multiplication rule for conditional probability.

The collection of *essentially deterministic* events $\mathscr{D}=\{A\in\mathscr{S}:\mathbb{P}(A)=0 \text{ or } \mathbb{P}(A)=1\}$ is independent.

Proof

Suppose that $\{A_1, A_2, \ldots, A_n\}\subseteq\mathscr{D}$. If $\mathbb{P}(A_i)=0$ for some $i\in\{1,2,\ldots,n\}$then $\mathbb{P}(A_1\cap A_2\cap\cdots\cap A_n)=0$. If $\mathbb{P}(A_i)=1$ for every $i\in\{1,2,\ldots,n\}$then $\mathbb{P}(A_1\cap A_2\cap\cdots\cap A_n)=1$. In either case, $\mathbb{P}(A_1\cap A_2\cdots\cap A_n)=\mathbb{P}(A_1)\mathbb{P}(A_2)\cdots\mathbb{P}(A_n)$.

The next result generalizes the theorem above on the complements of two independent events.

Suppose that $\mathscr{A}=\{A_i: i\in I\}$ and $\mathscr{B}=\{B_i: i\in I\}$ are two collections of events with the property that for each $i\in I$, either $B_i=A_i$ or $B_i=A_i^c$. Then $\mathscr{A}$ is independent if and only if $\mathscr{B}$ is an independent.

Proof

The proof is actually very similar to the proof for two events, except for more complicated notation. First, by the symmetry of the relation between $\mathscr{A}$ and $\mathscr{B}$, it suffices to show $\mathscr{A}$ indpendent implies $\mathscr{B}$ independent. Next, by the inheritance property, it suffices to consider the case where the index set I is finite.

1. Fix $k\in I$ and define $B_k=A_k^c$ and $B_i=A_i$ for $i\in I\setminus\{k\}$. Suppose now that $J\subseteq I$. If $k\notin J$ then trivially, $\mathbb{P}\left(\bigcap_{j\in J}B_j\right)=\prod_{j\in J}\mathbb{P}(B_j)$. If $k\in J$, then using the difference rule,
$$\begin{aligned}\mathbb{P}\left(\bigcap_{j\in J}B_j\right)&=\mathbb{P}\left(\bigcap_{j\in J\setminus\{k\}}A_j\right)-\mathbb{P}\left(\bigcap_{j\in J}A_j\right)\\&=\prod_{j\in J\setminus\{k\}}\mathbb{P}(A_j)-\prod_{j\in J}\mathbb{P}(A_j)=\left[\prod_{j\in J\setminus\{k\}}\mathbb{P}(A_j)\right][1-\mathbb{P}(A_k)]=\prod_{j\in J}\mathbb{P}(B_j)\end{aligned}$$
Hence $\{B_i: i\in I\}$ is a collection of independent events.
2. Suppose now that $\mathscr{B}=\{B_i: i\in I\}$ is a general collection of events where $B_i=A_i$ or $B_i=A_i^c$ for each $i\in I$. Then $\mathscr{B}$ can be obtained from $\mathscr{A}$ by a finite sequence of complement changes of the type in (a), each of which preserves independence.

The last theorem in turn leads to the type of strong independence that we want. The following exercise gives examples.

If A, B, C, and D are independent events, then

1. $A\cup B$, C^c, D are independent.
2. $A\cup B^c$, $C^c\cup D^c$ are independent.

Proof

We will give proofs that use the complement theorem, but to do so, some additional notation is helpful. If E is an event, let $E^1 = E$ and $E^0 = E^c$.

1. Note that $A \cup B = \bigcup_{(i,j)\in I} A^i \cap B^j$ where $I = \{(1,0),(0,1),(1,1)\}$ and note that the events in the union are disjoint. By the distributive property, $(A\cup B)\cap C^c = \bigcup_{(i,j)\in I} A^i \cap B^j \cap C^0$ and again the events in the union are disjoint. By additivity and complement theorem,
$$\mathbb{P}[(A\cup B)\cap C^c] = \sum_{(i,j)\in I} \mathbb{P}(A^i)\mathbb{P}(B^j)\mathbb{P}(C^0) = \left(\sum_{(i,j)\in I}\mathbb{P}(A^i)\mathbb{P}(B^j)\right)\mathbb{P}(C^0) = \mathbb{P}(A\cup B)\mathbb{P}(C^c) \tag{2.5.6}$$
By exactly the same type of argument, $\mathbb{P}[(A\cup B)\cap D] = \mathbb{P}(A\cup B)\mathbb{P}(D)$ and $\mathbb{P}[(A\cup B)\cap C^c \cap D] = \mathbb{P}(A\cup B)\mathbb{P}(C^c)\mathbb{P}(D)$. Directly from the result above on complements, $\mathbb{P}(C^c\cap D) = \mathbb{P}(C^c)\mathbb{P}(D)$.
2. Note that $A\cup B^c = \bigcup_{(i,j)\in I} A^i\cap B^j$ where $I = \{(0,0),(1,0),(1,1)\}$ and note that the events in the union are disjoint. Similarly $C^c \cup D^c = \bigcup_{(k,l)\in J} C^i \cap D^j$ where $J = \{(0,0),(1,0),(0,1)\}$, and again the events in the union are disjoint. By the distributive rule for set operations,
$$(A\cup B^c)\cap(C^c\cup D^c) = \bigcup_{(i,j,k,l)\in I\times J} A^i\cap B^j\cap C^k\cap D^l \tag{2.5.7}$$
and once again, the events in the union are disjoint. By additivity and the complement theorem,
$$\mathbb{P}[(A\cup B^c)\cap(C^c\cup D^c)] = \sum_{(i,j,k,l)\in I\times J}\mathbb{P}(A^i)\mathbb{P}(B^j)\mathbb{P}(C^k)\mathbb{P}(D^l) \tag{2.5.8}$$
But also by additivity, the complement theorem, and the distributive property of arithmetic,
$$\mathbb{P}(A\cup B^c)\mathbb{P}(C^c\cup D^c) = \left(\sum_{(i,j)\in I}\mathbb{P}(A^i)\mathbb{P}(B^j)\right)\left(\sum_{(k,l)\in J}\mathbb{P}(C^k)\mathbb{P}(D^l)\right) = \sum_{(i,j,k,l)\in I\times J}\mathbb{P}(A^i)\mathbb{P}(B^j)\mathbb{P}(C^k)\mathbb{P}(D^l) \tag{2.5.9}$$

The complete generalization of these results is a bit complicated, but roughly means that if we start with a collection of indpendent events, and form new events from disjoint subcollections (using the set operations of union, intersection, and complment), then the new events are independent. For a precise statement, see the section on measure spaces. The importance of the complement theorem lies in the fact that any event that can be defined in terms of a finite collection of events $\{A_i : i \in I\}$ can be written as a disjoint union of events of the form $\bigcap_{i\in I} B_i$ where $B_i = A_i$ or $B_i = A_i^c$ for each $i \in I$.

Another consequence of the general complement theorem is a formula for the probability of the union of a collection of independent events that is much nicer than the inclusion-exclusion formula.

If $A_1, A_2, \ldots, A_n$ are independent events, then
$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = 1 - \prod_{i=1}^n [1-\mathbb{P}(A_i)] \tag{2.5.10}$$

Proof

From DeMorgan's law and the independence of $A_1^c, A_2^c, \ldots, A_n^c$ we have
$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = 1 - \mathbb{P}\left(\bigcap_{i=1}^n A_i^c\right) = 1 - \prod_{i=1}^n \mathbb{P}(A_i^c) = 1 - \prod_{i=1}^n[1-\mathbb{P}(A_i)] \tag{2.5.11}$$

Independence of Random Variables

Suppose now that X_i is a random variable for the experiment with values in a set T_i for each i in a nonempty index set I. Mathematically, X_i is a function from S into T_i, and recall that $\{X_i \in B\}$ denotes the event $\{s\in S : X_i(s)\in B\}$ for $B\subseteq T_i$. Intuitively, X_i is a variable of interest in the experiment, and every meaningful statement about X_i defines an event. Intuitively, the random variables are independent if information about some of the variables tells us nothing about the other variables. Mathematically, independence of a collection of random variables can be reduced to the independence of collections of events.

The collection of random variables $\mathscr{X} = \{X_i : i \in I\}$ is *independent* if the collection of events $\{\{X_i \in B_i\} : i \in I\}$ is independent for every choice of $B_i \subseteq T_i$ for $i\in I$. Equivalently then, $\mathscr{X}$ is independent if for every finite $J\subseteq I$, and for every choice of $B_j \subseteq T_j$

for $j \in J$ we have

$$\mathbb{P}\left(\bigcap_{j\in J}\{X_j \in B_j\}\right) = \prod_{j\in J}\mathbb{P}(X_j \in B_j) \tag{2.5.12}$$

Details

Recall that T_i will have a σ-algebra $\mathscr{T}_i$ of admissible subsets so that $(T_i, \mathscr{T}_i)$ is a measurable space just like the sample space $(S, \mathscr{S})$ for each $i \in I$. Also X_i is measurable as a function from S into T_i for each $i \in I$. These technical assumptions ensure that the definition makes sense.

Suppose that $\mathscr{X}$ is a collection of random variables.

1. If $\mathscr{X}$ is independent, then $\mathscr{Y}$ is independent for every $\mathscr{Y} \subseteq \mathscr{X}$
2. If $\mathscr{Y}$ is independent for every finite $\mathscr{Y} \subseteq \mathscr{X}$ then $\mathscr{X}$ is independent.

It would seem almost obvious that if a collection of random variables is independent, and we transform each variable in deterministic way, then the new collection of random variables should still be independent.

Suppose now that g_i is a function from T_i into a set U_i for each $i \in I$. If $\{X_i : i \in I\}$ is independent, then $\{g_i(X_i) : i \in I\}$ is also independent.

Proof

Except for the abstract setting, the proof of independence is easy. Suppose that $C_i \subseteq U_i$ for each $i \in I$. Then $\{g_i(X_i) \in C_i\} = \left\{X_i \in g_i^{-1}(C_i)\right\}$ for $i \in I$. By the independence of $\{X_i : i \in I\}$, the collection of events $\left\{\left\{X_i \in g_i^{-1}(C_i)\right\} : i \in I\right\}$ is independent.

Technically, the set U_i will have a σ-algebra $\mathscr{U}_i$ of admissible subsets so that $(U_i, \mathscr{U}_i)$ is a measurable space just like $(T_i, \mathscr{T}_i)$ and just like the sample space $(S, \mathscr{S})$. The function g_i is required to be measurable as a function from T_i into U_i just as X_i is measurable as a function form S into T_i. In the proof above, $C_i \in \mathscr{U}_i$ so that $g^{-1}(C_i) \in \mathscr{T}_i$ and hence $\{X_i \in g^{-1}(C_i)\} \in \mathscr{S}$.

As with events, the (mutual) independence of random variables is a very strong property. If a collection of random variables is independent, then any subcollection is also independent. New random variables formed from disjoint subcollections are independent. For a simple example, suppose that X, Y, and Z are independent real-valued random variables. Then

1. $\sin(X)$, $\cos(Y)$, and e^Z are independent.
2. (X, Y) and Z are independent.
3. $X^2 + Y^2$ and $\arctan(Z)$ are independent.
4. X and Z are independent.
5. Y and Z are independent.

In particular, note that statement 2 in the list above is much stronger than the conjunction of statements 4 and 5. Contrapositively, if X and Z are dependent, then (X, Y) and Z are also dependent. Independence of random variables subsumes independence of events.

A collection of events $\mathscr{A}$ is independent if and only if the corresponding collection of indicator variables $\{\mathbf{1}_A : A \in \mathscr{A}\}$ is independent.

Proof

Let $\mathscr{A} = \{A_i : i \in I\}$ where I is a nonempty index set. For $i \in I$, the only non-trivial events that can be defined in terms of $\mathbf{1}_{A_i}$ are $\{\mathbf{1}_{A_i} = 1\} = A_i$ and $\{\mathbf{1}_{A_i} = 0\} = A_i^c$. So $\{\mathbf{1}_{A_i} : i \in I\}$ is independent if and only if every collection of the form $\{B_i : i \in I\}$ is independent, where for each $i \in I$, either $B_i = A_i$ or $B_i = A_i^c$. But by the complement theorem, this is equivalent to the independence of $\{A_i : i \in I\}$.

Many of the concepts that we have been using informally can now be made precise. A compound experiment that consists of "independent stages" is essentially just an experiment whose outcome is a sequence of independent random variables $\boldsymbol{X} = (X_1, X_2, \ldots)$ where X_i is the outcome of the ith stage.

In particular, suppose that we have a basic experiment with outcome variable X. By definition, the outcome of the experiment that consists of "independent replications" of the basic experiment is a sequence of independent random variables $\boldsymbol{X} = (X_1, X_2, \ldots)$ each with the same probability distribution as X. This is fundamental to the very concept of probability, as expressed in the law of large numbers. From a statistical point of view, suppose that we have a population of objects and a vector of measurements X of interest for the objects in the sample. The sequence $\boldsymbol{X}$ above corresponds to *sampling from the distribution of* X; that is, X_i is the vector of measurements for the ith object drawn from the sample. When we sample from a finite population, sampling *with replacement* generates independent random variables while sampling *without replacement* generates dependent random variables.

Conditional Independence and Conditional Probability

As noted at the beginning of our discussion, independence of events or random variables depends on the underlying probability measure. Thus, suppose that B is an event with positive probability. A collection of events or a collection of random variables is *conditionally independent given* B if the collection is independent relative to the conditional probability measure $A \mapsto \mathbb{P}(A \mid B)$. For example, a collection of events $\{A_i : i \in I\}$ is conditionally independent given B if for every finite $J \subseteq I$,

$$\mathbb{P}\left(\bigcap_{j \in J} A_j \,\middle|\, B\right) = \prod_{j \in J} \mathbb{P}(A_j \mid B) \tag{2.5.13}$$

Note that the definitions and theorems of this section would still be true, but with all probabilities conditioned on B.

Conversely, conditional probability has a nice interpretation in terms of independent replications of the experiment. Thus, suppose that we start with a basic experiment with S as the set of outcomes. We let X denote the outcome random variable, so that mathematically X is simply the identity function on S. In particular, if A is an event then trivially, $\mathbb{P}(X \in A) = \mathbb{P}(A)$. Suppose now that we replicate the experiment independently. This results in a new, compound experiment with a sequence of independent random variables $(X_1, X_2, \ldots)$, each with the same distribution as X. That is, X_i is the outcome of the ith repetition of the experiment.

Suppose now that A and B are events in the basic experiment with $\mathbb{P}(B) > 0$. In the compound experiment, the event that "when B occurs for the first time, A also occurs" has probability

$$\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A \mid B) \tag{2.5.14}$$

Proof

In the compound experiment, if we record $(X_1, X_2, \ldots)$ then the new set of outcomes is $S^\infty = S \times S \times \cdots$. The event that "when B occurs for the first time, A also occurs" is

$$\bigcup_{n=1}^\infty \{X_1 \notin B, X_2 \notin B, \ldots, X_{n-1} \notin B, X_n \in A \cap B\} \tag{2.5.15}$$

The events in the union are disjoint. Also, since $(X_1, X_2, \ldots)$ is a sequence of independent variables, each with the distribution of X we have

$$\mathbb{P}\left(X_1 \notin B, X_2 \notin B, \ldots, X_{n-1} \notin B, X_n \in A \cap B\right) = \left[\mathbb{P}\left(B^c\right)\right]^{n-1} \mathbb{P}(A \cap B) = [1 - \mathbb{P}(B)]^{n-1} \mathbb{P}(A \cap B) \tag{2.5.16}$$

Hence, using geometric series, the probability of the union is

$$\sum_{n=1}^\infty [1 - \mathbb{P}(B)]^{n-1} \mathbb{P}(A \cap B) = \frac{\mathbb{P}(A \cap B)}{1 - [1 - \mathbb{P}(B)]} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \tag{2.5.17}$$

Heuristic Argument

Suppose that we create a new experiment by repeating the basic experiment until B occurs for the first time, and then record the outcome of just the last repetition of the basic experiment. Now the set of outcomes is simply B and the appropriate probability measure on the new experiment is $A \mapsto \mathbb{P}(A \mid B)$.

Suppose that A and B are disjoint events in a basic experiment with $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$. In the compound experiment obtained by replicating the basic experiment, the event that "A occurs before B" has probability

$$\frac{\mathbb{P}(A)}{\mathbb{P}(A) + \mathbb{P}(B)} \tag{2.5.18}$$

Proof

Note that the event "A occurs before B" is the same as the event "when $A \cup B$ occurs for the first time, A occurs".

Examples and Applications

Basic Rules

Suppose that A, B, and C are independent events in an experiment with $\mathbb{P}(A) = 0.3$, $\mathbb{P}(B) = 0.4$, and $\mathbb{P}(C) = 0.8$. Express each of the following events in set notation and find its probability:

1. All three events occur.
2. None of the three events occurs.
3. At least one of the three events occurs.
4. At least one of the three events does not occur.
5. Exactly one of the three events occurs.
6. Exactly two of the three events occurs.

Answer

1. $\mathbb{P}(A \cap B \cap C) = 0.096$
2. $\mathbb{P}(A^c \cap B^c \cap C^c) = 0.084$
3. $\mathbb{P}(A \cup B \cup C) = 0.916$
4. $\mathbb{P}(A^c \cup B^c \cup C^c) = 0.904$
5. $\mathbb{P}[(A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c) \cup (A^c \cap B^c \cap C)] = 0.428$
6. $\mathbb{P}[(A \cap B \cap C^c) \cup (A \cap B^c \cap C) \cup (A^c \cap B \cap C)] = 0.392$

Suppose that A, B, and C are independent events for an experiment with $\mathbb{P}(A) = \frac{1}{3}$, $\mathbb{P}(B) = \frac{1}{4}$, and $\mathbb{P}(C) = \frac{1}{5}$. Find the probability of each of the following events:

1. $(A \cap B) \cup C$
2. $A \cup B^c \cup C$
3. $(A^c \cap B^c) \cup C^c$

Answer

1. $\frac{4}{15}$
2. $\frac{13}{15}$
3. $\frac{9}{10}$

Simple Populations

A small company has 100 employees; 40 are men and 60 are women. There are 6 male executives. How many female executives should there be if gender and rank are independent? The underlying experiment is to choose an employee at random.

Answer

9

Suppose that a farm has four orchards that produce peaches, and that peaches are classified by size as small, medium, and large. The table below gives total number of peaches in a recent harvest by orchard and by size. Fill in the body of the table with counts for the various intersections, so that orchard and size are independent variables. The underlying experiment is to select a peach at random from the farm.

Frequency	Size Small	Medium	Large	Total
Orchard 1			**Total2000**	400
2			**Total2000**	600
3			**Total2000**	300
4			**Total2000**	700
Total	400	1000	600	2000

Answer

Frequency	Size Small	Medium	Large	Total

Orchard 1	80	200	120	400
2	120	300	180	600
3	60	150	90	300
4	140	350	210	700
total	400	1000	600	2000

Note from the last two exercises that you cannot "see" independence in a Venn diagram. Again, independence is a measure-theoretic concept, not a set-theoretic concept.

Bernoulli Trials

A *Bernoulli trials sequence* is a sequence $\boldsymbol{X} = (X_1, X_2, \ldots)$ of independent, identically distributed indicator variables. Random variable X_i is the outcome of trial i, where in the usual terminology of reliability theory, 1 denotes *success* and 0 denotes *failure*. The canonical example is the sequence of scores when a coin (not necessarily fair) is tossed repeatedly. Another basic example arises whenever we start with an basic experiment and an event A of interest, and then repeat the experiment. In this setting, X_i is the indicator variable for event A on the ith run of the experiment. The Bernoulli trials process is named for Jacob Bernoulli, and has a single basic *parameter* $p = \mathbb{P}(X_i = 1)$. This random process is studied in detail in the chapter on Bernoulli trials.

For $(x_1, x_2, \ldots, x_n) \in \{0,1\}^n$,

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = p^{x_1 + x_2 + \cdots + x_n}(1-p)^{n-(x_1+x_2+\cdots+x_n)} \tag{2.5.19}$$

Proof

If X is a generic Bernoulli trial, then by definition, $\mathbb{P}(X=1) = p$ and $\mathbb{P}(X=0) = 1-p$. Equivalently, $\mathbb{P}(X=x) = p^x(1-p)^{1-x}$ for $x \in \{0,1\}$. Thus the result follows by independence.

Note that the sequence of indicator random variables $\boldsymbol{X}$ is exchangeable. That is, if the sequence $(x_1, x_2, \ldots, x_n)$ in the previous result is permuted, the probability does not change. On the other hand, there are exchangeable sequences of indicator random variables that are dependent, as Pólya's urn model so dramatically illustrates.

Let Y denote the number of successes in the first n trials. Then

$$\mathbb{P}(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}, \quad y \in \{0, 1, \ldots, n\} \tag{2.5.20}$$

Proof

Note that $Y = \sum_{i=1}^n X_i$, where X_i is the outcome of trial i, as in the previous result. For $y \in \{0, 1, \ldots, n\}$, the event $\{Y = y\}$ occurs if and only if exactly y of the n trials result in success (1). The number of ways to choose the y trials that result in success is $\binom{n}{y}$, and by the previous result, the probability of any particular sequence of y successes and $n-y$ failures is $p^y(1-p)^{n-y}$. Thus the result follows by the additivity of probability.

The distribution of Y is called the *binomial distribution* with parameters n and p. The binomial distribution is studied in more detail in the chapter on Bernoulli Trials.

More generally, a *multinomial trials sequence* is a sequence $\boldsymbol{X} = (X_1, X_2, \ldots)$ of independent, identically distributed random variables, each taking values in a finite set S. The canonical example is the sequence of scores when a k-sided die (not necessarily fair) is thrown repeatedly. Multinomial trials are also studied in detail in the chapter on Bernoulli trials.

Cards

Consider the experiment that consists of dealing 2 cards at random from a standard deck and recording the sequence of cards dealt. For $i \in \{1, 2\}$, let Q_i be the event that card i is a queen and H_i the event that card i is a heart. Compute the appropriate probabilities to verify the following results. Reflect on these results.

1. Q_1 and H_1 are independent.
2. Q_2 and H_2 are independent.
3. Q_1 and Q_2 are negatively correlated.

4. H_1 and H_2 are negatively correlated.
5. Q_1 and H_2 are independent.
6. H_1 and Q_2 are independent.

Answer

1. $\mathbb{P}(Q_1) = \mathbb{P}(Q_1 \mid H_1) = \frac{1}{13}$
2. $\mathbb{P}(Q_2) = \mathbb{P}(Q_2 \mid H_2) = \frac{1}{13}$
3. $\mathbb{P}(Q_1) = \frac{1}{13}$, $\mathbb{P}(Q_1 \mid Q_2) = \frac{1}{17}$
4. $\mathbb{P}(H_1) = \frac{1}{4}$, $\mathbb{P}(H_1 \mid H_2) = \frac{4}{17}$
5. $\mathbb{P}(Q_1) = \mathbb{P}(Q_1 \mid H_2) = \frac{1}{13}$
6. $\mathbb{P}(Q_2) = \mathbb{P}(Q_2 \mid H_1) = \frac{1}{13}$

In the card experiment, set $n = 2$. Run the simulation 500 times. For each pair of events in the previous exercise, compute the product of the empirical probabilities and the empirical probability of the intersection. Compare the results.

Dice

The following exercise gives three events that are *pairwise independent*, but not *(mutually) independent*.

Consider the dice experiment that consists of rolling 2 standard, fair dice and recording the sequence of scores. Let A denote the event that first score is 3, B the event that the second score is 4, and C the event that the sum of the scores is 7. Then

1. A, B, C are pairwise independent.
2. $A \cap B$ *implies* (is a subset of) C and hence these events are dependent in the strongest possible sense.

Answer

Note that $A \cap B = A \cap C = B \cap C = \{(3, 4)\}$, and the probability of the common intersection is $\frac{1}{36}$. On the other hand, $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{6}{36} = \frac{1}{6}$.

In the dice experiment, set $n = 2$. Run the experiment 500 times. For each pair of events in the previous exercise, compute the product of the empirical probabilities and the empirical probability of the intersection. Compare the results.

The following exercise gives an example of three events with the property that the probability of the intersection is the product of the probabilities, but the events are not pairwise independent.

Suppose that we throw a standard, fair die one time. Let $A = \{1, 2, 3, 4\}$, $B = C = \{4, 5, 6\}$. Then

1. $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$.
2. B and C are the same event, and hence are dependent in the strongest possbile sense.

Answer

Note that $A \cap B \cap C = \{4\}$, so $\mathbb{P}(A \cap B \cap C) = \frac{1}{6}$. On the other hand, $\mathbb{P}(A) = \frac{4}{6}$ and $\mathbb{P}(B) = \mathbb{P}(C) = \frac{3}{6}$.

Suppose that a standard, fair die is thrown 4 times. Find the probability of the following events.

1. Six does not occur.
2. Six occurs at least once.
3. The sum of the first two scores is 5 and the sum of the last two scores is 7.

Answer

1. $\left(\frac{5}{6}\right)^4 \approx 0.4823$
2. $1 - \left(\frac{5}{6}\right)^4 \approx 0.5177$
3. $\frac{1}{54}$

Suppose that a pair of standard, fair dice are thrown 8 times. Find the probability of each of the following events.

1. Double six does not occur.
2. Double six occurs at least once.

3. Double six does not occur on the first 4 throws but occurs at least once in the last 4 throws.

Answer

1. $\left(\frac{35}{36}\right)^8 \approx 0.7982$
2. $1-\left(\frac{35}{36}\right)^8 \approx 0.2018$
3. $\left(\frac{35}{36}\right)^4\left[1-\left(\frac{35}{36}\right)^4\right] \approx 0.0952$

Consider the dice experiment that consists of rolling n, k-sided dice and recording the sequence of scores $\boldsymbol{X}=(X_1, X_2, \ldots, X_n)$.The following conditions are equivalent (and correspond to the assumption that the dice are fair):

1. $\boldsymbol{X}$ is uniformly distributed on $\{1,2,\ldots,k\}^n$.
2. $\boldsymbol{X}$ is a sequence of independent variables, and X_i is uniformly distributed on $\{1,2,\ldots,k\}$ for each i.

Proof

Let $S=\{1,2,\ldots,k\}$ and note that S^n has k^n points. Suppose that $\boldsymbol{X}$ is uniformly distributed on S^n. Then $\mathbb{P}(\boldsymbol{X}=\boldsymbol{x})=1/k^n$ for each $\boldsymbol{x}\in S^n$ so $\mathbb{P}(X_i=x)=k^{n-1}/k^n=1/k$ for each $x\in S$. Hence X_i is uniformly distributed on S. Moreover,

$$\mathbb{P}(\boldsymbol{X}=\boldsymbol{x})=\mathbb{P}(X_1=x_1)\mathbb{P}(X_2=x_2)\cdots\mathbb{P}(X_n=x_n), \quad \boldsymbol{x}=(x_1,x_2,\ldots,x_n)\in S^n \tag{2.5.21}$$

so $\boldsymbol{X}$ is an independent sequence. Conversely, if $\boldsymbol{X}$ is an independent sequence and X_i is uniformly distributed on S for each i then $\mathbb{P}(X_i=x)=1/k$ for each $x\in S$ and hence $\mathbb{P}(\boldsymbol{X}=\boldsymbol{x})=1/k^n$ for each $\boldsymbol{x}\in S^n$. Thus $\boldsymbol{X}$ is uniformly distributed on S^n.

A pair of standard, fair dice are thrown repeatedly. Find the probability of each of the following events.

1. A sum of 4 occurs before a sum of 7.
2. A sum of 5 occurs before a sum of 7.
3. A sum of 6 occurs before a sum of 7.
4. When a sum of 8 occurs the first time, it occurs “the hard way” as $(4,4)$.

Answer

1. $\frac{3}{9}$
2. $\frac{4}{10}$
3. $\frac{5}{11}$
4. $\frac{1}{5}$

Problems of the type in the last exercise are important in the game of *craps*. Craps is studied in more detail in the chapter on Games of Chance.

Coins

A biased coin with probability of heads $\frac{1}{3}$ is tossed 5 times. Let $\boldsymbol{X}$ denote the outcome of the tosses (encoded as a bit string) and let Y denote the number of heads. Find each of the following:

1. $\mathbb{P}(\boldsymbol{X}=\boldsymbol{x})$ for each $\boldsymbol{x}\in\{0,1\}^5$.
2. $\mathbb{P}(Y=y)$ for each $y\in\{0,1,2,3,4,5\}$.
3. $\mathbb{P}(1\le Y\le 3)$

Answer

1. $\frac{32}{243}$ if $\boldsymbol{x}=00000$, $\frac{16}{243}$ if $\boldsymbol{x}$ has exactly one 1 (there are 5 of these), $\frac{8}{243}$ if $\boldsymbol{x}$ has exactly two 1s (there are 10 of these), $\frac{4}{243}$ if $\boldsymbol{x}$ has exactly three 1s (there are 10 of these), $\frac{2}{243}$ if $\boldsymbol{x}$ has exactly four 1s (there are 5 of these), $\frac{1}{243}$ if $\boldsymbol{x}=11111$
2. $\frac{32}{243}$ if $y=0$, $\frac{80}{243}$ if $y=1$, $\frac{80}{243}$ if $y=2$, $\frac{40}{243}$ if $y=3$, $\frac{10}{243}$ if $y=4$, $\frac{1}{243}$ if $y=5$
3. $\frac{200}{243}$

A box contains a fair coin and a two-headed coin. A coin is chosen at random from the box and tossed repeatedly. Let F denote the event that the fair coin is chosen, and let H_i denote the event that the ith toss results in heads. Then

1. $(H_1,H_2,\ldots)$ are conditionally independent given F, with $\mathbb{P}(H_i\mid F)=\frac{1}{2}$ for each i.
2. $(H_1,H_2,\ldots)$ are conditionally independent given F^c, with $\mathbb{P}(H_i\mid F^c)=1$ for each i.

3. $\mathbb{P}(H_i) = \frac{3}{4}$ for each i.
4. $\mathbb{P}(H_1 \cap H_2 \cap \cdots \cap H_n) = \frac{1}{2^{n+1}} + \frac{1}{2}$.
5. $(H_1, H_2, \ldots)$ are dependent.
6. $\mathbb{P}(F \mid H_1 \cap H_2 \cap \cdots \cap H_n) = \frac{1}{2^n+1}$.
7. $\mathbb{P}(F \mid H_1 \cap H_2 \cap \cdots \cap H_n) \to 0$ as $n \to \infty$.

Proof

Parts (a) and (b) are essentially modeling assumptions, based on the design of the experiment. If we know what kind of coin we have, then the tosses are independent. Parts (c) and (d) follow by conditioning on the type of coin and using parts (a) and (b). Part (e) follows from (c) and (d). Note that the expression in (d) is not $(3/4)^n$. Part (f) follows from part (d) and Bayes' theorem. Finally part (g) follows from part (f).

Consider again the box in the previous exercise, but we change the experiment as follows: a coin is chosen at random from the box and tossed and the result recorded. The coin is returned to the box and the process is repeated. As before, let H_i denote the event that toss i results in heads. Then

1. $(H_1, H_2, \ldots)$ are independent.
2. $\mathbb{P}(H_i) = \frac{3}{4}$ for each i.
3. $\mathbb{P}(H_1 \cap H_2 \cap \cdots H_n) = \left(\frac{3}{4}\right)^n$.

Proof

Again, part (a) is essentially a modeling assumption. Since we return the coin and draw a new coin at random each time, the results of the tosses should be independent. Part (b) follows by conditioning on the type of the ith coin. Part (c) follows from parts (a) and (b).

Think carefully about the results in the previous two exercises, and the differences between the two models. Tossing a coin produces independent random variables *if* the probability of heads is fixed (that is, non-random even if unknown). Tossing a coin with a random probability of heads generally does not produce independent random variables; the result of a toss gives information about the probability of heads which in turn gives information about subsequent tosses.

Uniform Distributions

Recall that Buffon's coin experiment consists of tossing a coin with radius $r \le \frac{1}{2}$ randomly on a floor covered with square tiles of side length 1. The coordinates (X, Y) of the center of the coin are recorded relative to axes through the center of the square in which the coin lands. The following conditions are equivalent:

1. (X, Y) is uniformly distributed on $\left[-\frac{1}{2}, \frac{1}{2}\right]^2$.
2. X and Y are independent and each is uniformly distributed on $\left[-\frac{1}{2}, \frac{1}{2}\right]$.

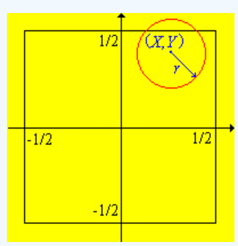

Figure 2.5.1: Buffon's coin experiment

Proof

Let $S = \left[-\frac{1}{2}, \frac{1}{2}\right]$, and let λ_1 denote length measure on S and λ_2 area measure on S^2. Note that $\lambda_1(S) = \lambda_2(S^2) = 1$. Suppose that (X, Y) is uniformly distributed on S^2, so that $\mathbb{P}[(X, Y) \in C] = \lambda_2(C)$ for $C \subseteq S^2$. For $A \subseteq S$,

$$\mathbb{P}(X \in A) = \mathbb{P}[(X, Y) \in A \times S] = \lambda_2(A \times S) = \lambda_1(A) \tag{2.5.22}$$

Hence X is uniformly distributed on S. By a similar argument, Y is also uniformly distributed on S. Moreover, for $A \subseteq S$ and $B \subseteq S$,

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}[(X, Y) \in A \times B] = \lambda_2(A \times B) = \lambda_1(A)\lambda_1(B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B) \tag{2.5.23}$$

so X and Y are independent. Conversely, if X and Y are independent and each is uniformly distributed on S, then for $A \subseteq S$ and $B \subseteq S$,

$$\mathbb{P}[(X, Y) \in A \times B] = \mathbb{P}(X \in A)\mathbb{P}(Y \in B) = \lambda_1(A)\lambda_1(B) = \lambda_2(A \times B) \tag{2.5.24}$$

It then follows that $\mathbb{P}[(X, Y) \in C] = \lambda_2(C)$ for every $C \subseteq S^2$. For more details about this last step, see the advanced section on existence and uniqueness of measures.

Compare this result with the result above for fair dice.

In Buffon's coin experiment, set $r = 0.3$. Run the simulation 500 times. For the events $\{X > 0\}$ and $\{Y < 0\}$, compute the product of the empirical probabilities and the empirical probability of the intersection. Compare the results.

The arrival time X of the A train is uniformly distributed on the interval $(0, 30)$, while the arrival time Y of the B train is uniformly distributed on the interval $(15, 30)$. (The arrival times are in minutes, after 8:00 AM). Moreover, the arrival times are independent. Find the probability of each of the following events:

1. The A train arrives first.
2. Both trains arrive sometime after 20 minutes.

Answer

1. $\frac{3}{4}$
2. $\frac{2}{9}$

Reliability

Recall the simple model of structural reliability in which a system is composed of n components. Suppose in addition that the components operate independently of each other. As before, let X_i denote the state of component i, where 1 means working and 0 means failure. Thus, our basic assumption is that the *state vector* $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is a sequence of independent indicator random variables. We assume that the state of the system (either working or failed) depends only on the states of the components. Thus, the state of the system is an indicator random variable

$$Y = y(X_1, X_2, \ldots, X_n) \tag{2.5.25}$$

where $y : \{0, 1\}^n \to \{0, 1\}$ is the *structure function*. Generally, the probability that a device is working is the *reliability* of the device. Thus, we will denote the reliability of component i by $p_i = \mathbb{P}(X_i = 1)$ so that the vector of component reliabilities is $\boldsymbol{p} = (p_1, p_2, \ldots, p_n)$. By independence, the system reliability r is a function of the component reliabilities:

$$r(p_1, p_2, \ldots, p_n) = \mathbb{P}(Y = 1) \tag{2.5.26}$$

Appropriately enough, this function is known as the *reliability function*. Our challenge is usually to find the reliability function, given the structure function. When the components all have the same probability p then of course the system reliability r is just a function of p. In this case, the state vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ forms a sequence of Bernoulli trials.

Comment on the independence assumption for real systems, such as your car or your computer.

Recall that a *series system* is working if and only if each component is working.

1. The state of the system is $U = X_1 X_2 \cdots X_n = \min\{X_1, X_2, \ldots, X_n\}$.
2. The reliability is $\mathbb{P}(U = 1) = p_1 p_2 \cdots p_n$.

Recall that a *parallel system* is working if and only if at least one component is working.

1. The state of the system is $V = 1 - (1 - X_1)(1 - X_2) \cdots (1 - X_n) = \max\{X_1, X_2, \ldots, X_n\}$.
2. The reliability is $\mathbb{P}(V = 1) = 1 - (1 - p_1)(1 - p_2) \cdots (1 - p_n)$.

Recall that a k *out of* n *system* is working if and only if at least k of the n components are working. Thus, a parallel system is a 1 out of n system and a series system is an n out of n system. A k out of $2k - 1$ system is a *majority rules system*. The reliability function of a general k out of n system is a mess. However, if the component reliabilities are the same, the function has a reasonably simple form.

For a k out of n system with common component reliability p, the system reliability is

$$r(p) = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i} \tag{2.5.27}$$

Consider a system of 3 independent components with common reliability $p = 0.8$. Find the reliability of each of the following:

1. The parallel system.
2. The 2 out of 3 system.
3. The series system.

Answer

1. 0.992
2. 0.896
3. 0.512

Consider a system of 3 independent components with reliabilities $p_1 = 0.8$, $p_2 = 0.8$, $p_3 = 0.7$. Find the reliability of each of the following:

1. The parallel system.
2. The 2 out of 3 system.
3. The series system.

Answer

1. 0.994
2. 0.902
3. 0.504

Consider an airplane with an odd number of engines, each with reliability p. Suppose that the airplane is a majority rules system, so that the airplane needs a majority of working engines in order to fly.

1. Find the reliability of a 3 engine plane as a function of p.
2. Find the reliability of a 5 engine plane as a function of p.
3. For what values of p is a 5 engine plane preferable to a 3 engine plane?

Answer

1. $r_3(p) = 3\,p^2 - 2\,p^3$
2. $r_5(p) = 6\,p^5 - 15\,p^4 + 10p^3$
3. The 5-engine plane would be preferable if $p > \frac{1}{2}$ (which one would hope would be the case). The 3-engine plane would be preferable if $p < \frac{1}{2}$. If $p = \frac{1}{2}$, the 3-engine and 5-engine planes are equally reliable.

The graph below is known as the *Wheatstone bridge network* and is named for Charles Wheatstone. The edges represent components, and the system works if and only if there is a working path from vertex a to vertex b.

1. Find the structure function.
2. Find the reliability function.

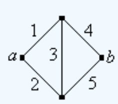

Figure 2.5.2: The Wheatstone bridge netwok

Answer

1. $Y = X_3(X_1 + X_2 - X_1 X_2)(X_4 + X_5 - X_4, X_5) + (1 - X_3)(X_1 X_4 + X_2 X_5 - X_1 X_2 X_4 X_5)$
2. $r(p_1, p_2, p_3, p_4, p_5) = p_3(p_1 + p_2 - p_1 p_2)(p_4 + p_5 - p_4, p_5) + (1 - p_3)(p_1 p_4 + p_2 p_5 - p_1 p_2 p_4 p_5)$

A system consists of 3 components, connected in parallel. Because of environmental factors, the components do not operate independently, so our usual assumption does not hold. However, we will assume that under low stress conditions, the components are independent, each with reliability 0.9; under medium stress conditions, the components are independent with reliability 0.8; and under high stress conditions, the components are independent, each with reliability 0.7. The probability of low stress is 0.5, of medium stress is 0.3, and of high stress is 0.2.

1. Find the reliability of the system.
2. Given that the system works, find the conditional probability of each stress level.

Answer

1. 0.9917. Condition on the stress level.
2. 0.5037 for low, 0.3001 for medium, 0.1962 for high. Use Bayes' theorem and part (a).

Suppose that bits are transmitted across a noisy communications channel. Each bit that is sent, independently of the others, is received correctly with probability 0.9 and changed to the complementary bit with probability 0.1. Using redundancy to improve reliability, suppose that a given bit will be sent 3 times. We naturally want to compute the probability that we correctly identify the bit that was sent. Assume we have no prior knowledge of the bit, so we assign probability $\frac{1}{2}$ each to the event that 000 was sent and the event that 111 was sent. Now find the conditional probability that 111 was sent given each of the 8 possible bit strings received.

Answer

Let $\boldsymbol{X}$ denote the string sent and $\boldsymbol{Y}$ the string received.

$\boldsymbol{y}$	$\mathbb{P}(\boldsymbol{X}=111 \mid \boldsymbol{Y}=\boldsymbol{y})$
111	$729/730$
110	$9/10$
101	$9/10$
011	$9/10$
100	$1/10$
010	$1/10$
001	$1/10$
000	$1/730$

Diagnostic Testing

Recall the discussion of diagnostic testing in the section on Conditional Probability. Thus, we have an event A for a random experiment whose occurrence or non-occurrence we cannot observe directly. Suppose now that we have n tests for the occurrence of A, labeled from 1 to n. We will let T_i denote the event that test i is positive for A. The tests are independent in the following sense:

- If A occurs, then $(T_1, T_2, \ldots, T_n)$ are (conditionally) independent and test i has sensitivity $a_i = \mathbb{P}(T_i \mid A)$.
- If A does not occur, then $(T_1, T_2, \ldots, T_n)$ are (conditionally) independent and test i has specificity $b_i = \mathbb{P}(T_i^c \mid A^c)$.

Note that *unconditionally*, it is not reasonable to assume that the tests are independent. For example, a positive result for a given test presumably is evidence that the condition A has occurred, which in turn is evidence that a subsequent test will be positive. In short, we expect that T_i and T_j should be positively correlated.

We can form a new, compound test by giving a *decision rule* in terms of the individual test results. In other words, the event T that the compound test is positive for A is a function of $(T_1, T_2, \ldots, T_n)$. The typical decision rules are very similar to the reliability structures discussed above. A special case of interest is when the n tests are independent applications of a given basic test. In this case, $a_i = a$ and $b_i = b$ for each i.

Consider the compound test that is positive for A if and only if each of the n tests is positive for A.

1. $T = T_1 \cap T_2 \cap \cdots \cap T_n$
2. The sensitivity is $\mathbb{P}(T \mid A) = a_1 a_2 \cdots a_n$.
3. The specificity is $\mathbb{P}(T^c \mid A^c) = 1 - (1 - b_1)(1 - b_2) \cdots (1 - b_n)$

Consider the compound test that is positive for A if and only if each at least one of the n tests is positive for A.

1. $T = T_1 \cup T_2 \cup \cdots \cup T_n$
2. The sensitivity is $\mathbb{P}(T \mid A) = 1 - (1 - a_1)(1 - a_2) \cdots (1 - a_n)$.
3. The specificity is $\mathbb{P}(T^c \mid A^c) = b_1 b_2 \cdots b_n$.

More generally, we could define the compound k *out of* n *test* that is positive for A if and only if at least k of the individual tests are positive for A. The series test is the n out of n test, while the parallel test is the 1 out of n test. The k out of $2k-1$ test is the *majority rules test*.

Suppose that a woman initially believes that there is an even chance that she is or is not pregnant. She buys three identical pregnancy tests with sensitivity 0.95 and specificity 0.90. Tests 1 and 3 are positive and test 2 is negative.

1. Find the updated probability that the woman is pregnant.
2. Can we just say that tests 2 and 3 cancel each other out? Find the probability that the woman is pregnant given just one positive test, and compare the answer with the answer to part (a).

Answer

1. 0.834
2. No: 0.905.

Suppose that 3 independent, identical tests for an event A are applied, each with sensitivity a and specificity b. Find the sensitivity and specificity of the following tests:

1. 1 out of 3 test
2. 2 out of 3 test
3. 3 out of 3 test

Answer

1. sensitivity $1-(1-a)^3$, specificity b^3
2. sensitivity $3\,a^2$, specificity $b^3+3\,b^2(1-b)$
3. sensitivity a^3, specificity $1-(1-b)^3$

In a criminal trial, the defendant is convicted if and only if all 6 jurors vote guilty. Assume that if the defendant really is guilty, the jurors vote guilty, independently, with probability 0.95, while if the defendant is really innocent, the jurors vote not guilty, independently with probability 0.8. Suppose that 70% of defendants brought to trial are guilty.

1. Find the probability that the defendant is convicted.
2. Given that the defendant is convicted, find the probability that the defendant is guilty.
3. Comment on the assumption that the jurors act independently.

Answer

1. 0.5148
2. 0.99996
3. The independence assumption is not reasonable since jurors collaborate.

Genetics

Please refer to the discussion of genetics in the section on random experiments if you need to review some of the definitions in this section.

Recall first that the *ABO blood type* in humans is determined by three alleles: a, b, and o. Furthermore, a and b are co-dominant and o is recessive. Suppose that in a certain population, the proportion of a, b, and o alleles are p, q, and r respectively. Of course we must have $p>0$, $q>0$, $r>0$ and $p+q+r=1$.

Suppose that the blood genotype in a person is the result of independent alleles, chosen with probabilities p, q, and r as above.

1. The probability distribubtion of the geneotypes is given in the following table:

Genotype	aa	ab	ao	bb	bo	oo
Probability	p^2	$2pq$	$2pr$	q^2	$2qr$	r^2

2. The probability distribution of the blood types is given in the following table:

Blood type	A	B	AB	O
Probability	p^2+2pr	q^2+2qr	$2pq$	r^2

Proof

Part (a) follows from the independence assumption and basic rules of probability. Even though genotypes are listed as unordered pairs, note that there are two ways that a heterozygous genotype can occur, since either parent could contribute either of the two distinct alleles. Part (b) follows from part (a) and basic rules of probability.

The discussion above is related to the *Hardy-Weinberg model* of genetics. The model is named for the English mathematician Godfrey Hardy and the German physician Wilhelm Weiberg

Suppose that the probability distribution for the set of blood types in a certain population is given in the following table:

Blood type	A	B	AB	O
Probability	0.360	0.123	0.038	0.479

Find p, q, and r.

Answer

$p = 0.224$, $q = 0.084$, $r = 0.692$

Suppose next that pod color in certain type of pea plant is determined by a gene with two alleles: g for green and y for yellow, and that g is dominant and o recessive.

Suppose that 2 green-pod plants are bred together. Suppose further that each plant, independently, has the recessive yellow-pod allele with probability $\frac{1}{4}$.

1. Find the probability that 3 offspring plants will have green pods.
2. Given that the 3 offspring plants have green pods, find the updated probability that both parents have the recessive allele.

Answer

1. $\frac{987}{1024}$
2. $\frac{27}{987}$

Next consider a sex-linked hereditary disorder in humans (such as colorblindness or hemophilia). Let h denote the healthy allele and d the defective allele for the gene linked to the disorder. Recall that h is dominant and d recessive for women.

Suppose that a healthy woman initially has a $\frac{1}{2}$ chance of being a carrier. (This would be the case, for example, if her mother and father are healthy but she has a brother with the disorder, so that her mother must be a carrier).

1. Find the probability that the first two sons of the women will be healthy.
2. Given that the first two sons are healthy, compute the updated probability that she is a carrier.
3. Given that the first two sons are healthy, compute the conditional probability that the third son will be healthy.

Answer

1. $\frac{5}{8}$
2. $\frac{1}{5}$
3. $\frac{9}{10}$

Laplace's Rule of Succession

Suppose that we have $m+1$ coins, labeled $0, 1, \ldots, m$. Coin i lands heads with probability $\frac{i}{m}$ for each i. The experiment is to choose a coin at random (so that each coin is equally likely to be chosen) and then toss the chosen coin repeatedly.

1. The probability that the first n tosses are all heads is $p_{m,n} = \frac{1}{m+1} \sum_{i=0}^m \left(\frac{i}{m}\right)^n$
2. $p_{m,n} \to \frac{1}{n+1}$ as $m \to \infty$
3. The conditional probability that toss $n+1$ is heads given that the previous n tosses were all heads is $\frac{p_{m,n+1}}{p_{m,n}}$
4. $\frac{p_{m,n+1}}{p_{m,n}} \to \frac{n+1}{n+2}$ as $m \to \infty$

Proof

Part (a) follows by conditioning on the chosen coin. For part (b), note that $p_{m,n}$ is an approximating sum for $\int_0^1 x^n \, dx = \frac{1}{n+1}$. Part (c) follows from the definition of conditional probability, and part (d) is a trivial consequence of (b), (c).

Note that coin 0 is two-tailed, the probability of heads increases with i, and coin m is two-headed. The limiting conditional probability in part (d) is called *Laplace's Rule of Succession*, named after Simon Laplace. This rule was used by Laplace and others as a general principle for estimating the conditional probability that an event will occur on time $n+1$, given that the event has occurred n times in succession.

Suppose that a missile has had 10 successful tests in a row. Compute Laplace's estimate that the 11th test will be successful. Does this make sense?

Answer

$\frac{11}{12}$. No, not really.