

2.4: Conditional Probability

The purpose of this section is to study how probabilities are updated in light of new information, clearly an absolutely essential topic. If you are a new student of probability, you may want to skip the technical details.

Definitions and Interpretations

The Basic Definition

As usual, we start with a random experiment modeled by a probability space $(S, \mathcal{S}, \mathbb{P})$. Thus, S is the set of outcomes, \mathcal{S} the collection of events, and \mathbb{P} the probability measure on the sample space (S, \mathcal{S}) . Suppose now that we know that an event B has occurred. In general, this information should clearly alter the probabilities that we assign to other events. In particular, if A is another event then A occurs if and only if A and B occur; effectively, the sample space has been reduced to B . Thus, the probability of A , given that we know B has occurred, should be proportional to $\mathbb{P}(A \cap B)$.



Figure 2.4.1: Events A and B

However, conditional probability, given that B has occurred, should still be a probability measure, that is, it must satisfy the axioms of probability. This forces the proportionality constant to be $1/\mathbb{P}(B)$. Thus, we are led inexorably to the following definition:

Let A and B be events with $\mathbb{P}(B) > 0$. The *conditional probability* of A given B is defined to be

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad (2.4.1)$$

The Law of Large Numbers

The [definition](#) above was based on the axiomatic definition of probability. Let's explore the idea of conditional probability from the less formal and more intuitive notion of relative frequency (the law of large numbers). Thus, suppose that we run the experiment repeatedly. For $n \in \mathbb{N}_+$ and an event E , let $N_n(E)$ denote the number of times E occurs (the frequency of E) in the first n runs. Note that $N_n(E)$ is a random variable in the compound experiment that consists of replicating the original experiment. In particular, its value is unknown until we actually run the experiment n times.

If $N_n(B)$ is large, the conditional probability that A has occurred, given that B has occurred, should be close to the *conditional relative frequency* of A given B , namely the relative frequency of A for the runs on which B occurred: $N_n(A \cap B)/N_n(B)$. But note that

$$\frac{N_n(A \cap B)}{N_n(B)} = \frac{N_n(A \cap B)/n}{N_n(B)/n} \quad (2.4.2)$$

The numerator and denominator of the main fraction on the right are the relative frequencies of $A \cap B$ and B , respectively. So by the law of large numbers again, $N_n(A \cap B)/n \rightarrow \mathbb{P}(A \cap B)$ as $n \rightarrow \infty$ and $N_n(B)/n \rightarrow \mathbb{P}(B)$ as $n \rightarrow \infty$. Hence

$$\frac{N_n(A \cap B)}{N_n(B)} \rightarrow \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \text{ as } n \rightarrow \infty \quad (2.4.3)$$

and we are led again to the [definition](#) above.

In some cases, conditional probabilities can be computed *directly*, by effectively reducing the sample space to the given event. In other cases, the formula in the mathematical definition is better. In some cases, conditional probabilities are known from modeling assumptions, and then are used to compute other probabilities. We will see examples of all of these situations in the [computational exercises](#) below.

It's very important that you not confuse $\mathbb{P}(A | B)$, the probability of A given B , with $\mathbb{P}(B | A)$, the probability of B given A . Making that mistake is known as the *fallacy of the transposed conditional*. (How embarrassing!)

Conditional Distributions

Suppose that X is a random variable for the experiment with values in T . Mathematically, X is a function from S into T , and $\{X \in A\}$ denotes the event $\{s \in S : X(s) \in A\}$ for $A \subseteq T$. Intuitively, X is a variable of interest in the experiment, and every meaningful statement about X defines an event. Recall that the probability distribution of X is the probability measure on T given by

$$A \mapsto \mathbb{P}(X \in A), \quad A \subseteq T \quad (2.4.4)$$

This has a natural extension to a conditional distribution, given an event.

If B is an event with $\mathbb{P}(B) > 0$, then the *conditional distribution of X given B* is the probability measure on T given by

$$A \mapsto \mathbb{P}(X \in A \mid B), \quad A \subseteq T \quad (2.4.5)$$

Details

Recall that T will come with a σ -algebra of admissible subsets so that (T, \mathcal{T}) is a measurable space, just like the sample space (S, \mathcal{S}) . Random variable X is required to be measurable as a function from S into T . This ensures that $\{X \in A\}$ is a valid event for each $A \in \mathcal{T}$, so that the definition makes sense.

Basic Theory

Preliminary Results

Our first result is of fundamental importance, and indeed was a crucial part of the argument for the definition of conditional probability.

Suppose again that B is an event with $\mathbb{P}(B) > 0$. Then $A \mapsto \mathbb{P}(A \mid B)$ is a probability measure on S .

Proof

Clearly $\mathbb{P}(A \mid B) \geq 0$ for every event A , and $\mathbb{P}(S \mid B) = 1$. Thus, suppose that $\{A_i : i \in I\}$ is a countable collection of pairwise disjoint events. Then

$$\mathbb{P}\left(\bigcup_{i \in I} A_i \mid B\right) = \frac{1}{\mathbb{P}(B)} \mathbb{P}\left[\left(\bigcup_{i \in I} A_i\right) \cap B\right] = \frac{1}{\mathbb{P}(B)} \mathbb{P}\left(\bigcup_{i \in I} (A_i \cap B)\right) \quad (2.4.6)$$

But the collection of events $\{A_i \cap B : i \in I\}$ is also pairwise disjoint, so

$$\mathbb{P}\left(\bigcup_{i \in I} A_i \mid B\right) = \frac{1}{\mathbb{P}(B)} \sum_{i \in I} \mathbb{P}(A_i \cap B) = \sum_{i \in I} \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \sum_{i \in I} \mathbb{P}(A_i \mid B) \quad (2.4.7)$$

It's hard to overstate the importance of the [last result](#) because this theorem means that any result that holds for probability measures in general holds for conditional probability, as long as the conditioning event remains fixed. In particular the basic probability rules in the section on Probability Measure have analogs for conditional probability. To give two examples,

$$\mathbb{P}(A^c \mid B) = 1 - \mathbb{P}(A \mid B) \quad (2.4.8)$$

$$\mathbb{P}(A_1 \cup A_2 \mid B) = \mathbb{P}(A_1 \mid B) + \mathbb{P}(A_2 \mid B) - \mathbb{P}(A_1 \cap A_2 \mid B) \quad (2.4.9)$$

By the same token, it follows that the conditional distribution of a random variable with values in T , given in [above](#), really does define a probability distribution on T . No further proof is necessary. Our next results are very simple.

Suppose that A and B are events with $\mathbb{P}(B) > 0$.

1. If $B \subseteq A$ then $\mathbb{P}(A \mid B) = 1$.
2. If $A \subseteq B$ then $\mathbb{P}(A \mid B) = \mathbb{P}(A)/\mathbb{P}(B)$.
3. If A and B are disjoint then $\mathbb{P}(A \mid B) = 0$.

Proof

These results follow directly from the definition of conditional probability. In part (a), note that $A \cap B = B$. In part (b) note that $A \cap B = A$. In part (c) note that $A \cap B = \emptyset$.

Parts (a) and (c) certainly make sense. Suppose that we know that event B has occurred. If $B \subseteq A$ then A becomes a certain event. If $A \cap B = \emptyset$ then A becomes an impossible event. A conditional probability can be computed relative to a probability measure that is itself a conditional probability measure. The following result is a consistency condition.

Suppose that A , B , and C are events with $\mathbb{P}(B \cap C) > 0$. The probability of A given B , relative to $\mathbb{P}(\cdot | C)$, is the same as the probability of A given B and C (relative to \mathbb{P}). That is,

$$\frac{\mathbb{P}(A \cap B | C)}{\mathbb{P}(B | C)} = \mathbb{P}(A | B \cap C) \quad (2.4.10)$$

Proof

From the definition,

$$\frac{\mathbb{P}(A \cap B | C)}{\mathbb{P}(B | C)} = \frac{\mathbb{P}(A \cap B \cap C) / \mathbb{P}(C)}{\mathbb{P}(B \cap C) / \mathbb{P}(C)} = \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(B \cap C)} = \mathbb{P}(A | B \cap C) \quad (2.4.11)$$

Correlation

Our next discussion concerns an important concept that deals with how two events are related, in a probabilistic sense.

Suppose that A and B are events with $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$.

1. $\mathbb{P}(A | B) > \mathbb{P}(A)$ if and only if $\mathbb{P}(B | A) > \mathbb{P}(B)$ if and only if $\mathbb{P}(A \cap B) > \mathbb{P}(A)\mathbb{P}(B)$. In this case, A and B are *positively correlated*.
2. $\mathbb{P}(A | B) < \mathbb{P}(A)$ if and only if $\mathbb{P}(B | A) < \mathbb{P}(B)$ if and only if $\mathbb{P}(A \cap B) < \mathbb{P}(A)\mathbb{P}(B)$. In this case, A and B are *negatively correlated*.
3. $\mathbb{P}(A | B) = \mathbb{P}(A)$ if and only if $\mathbb{P}(B | A) = \mathbb{P}(B)$ if and only if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. In this case, A and B are *uncorrelated or independent*.

Proof

These properties follow directly from the definition of conditional probability and simple algebra. Recall that multiplying or dividing an inequality by a positive number preserves the inequality.

Intuitively, if A and B are positively correlated, then the occurrence of either event means that the other event is more likely. If A and B are negatively correlated, then the occurrence of either event means that the other event is less likely. If A and B are uncorrelated, then the occurrence of either event does not change the probability of the other event. Independence is a fundamental concept that can be extended to more than two events and to random variables; these generalizations are studied in the next section on Independence. A much more general version of correlation, for random variables, is explored in the section on Covariance and Correlation in the chapter on Expected Value.

Suppose that A and B are events. Note from (4) that if $A \subseteq B$ or $B \subseteq A$ then A and B are positively correlated. If A and B are disjoint then A and B are negatively correlated.

Suppose that A and B are events in a random experiment.

1. A and B have the same correlation (positive, negative, or zero) as A^c and B^c .
2. A and B have the opposite correlation as A and B^c (that is, positive-negative, negative-positive, or 0-0).

Proof

1. Using DeMorgan's law and the complement law.

$$\mathbb{P}(A^c \cap B^c) - \mathbb{P}(A^c)\mathbb{P}(B^c) = \mathbb{P}[(A \cup B)^c] - \mathbb{P}(A^c)\mathbb{P}(B^c) = [1 - \mathbb{P}(A \cup B)] - [1 - \mathbb{P}(A)][1 - \mathbb{P}(B)] \quad (2.4.12)$$

Using the inclusion-exclusion law and algebra,

$$\mathbb{P}(A^c \cap B^c) - \mathbb{P}(A^c)\mathbb{P}(B^c) = \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) \quad (2.4.13)$$

2. Using the difference rule and the complement law:

$$\mathbb{P}(A \cap B^c) - \mathbb{P}(A)\mathbb{P}(B^c) = \mathbb{P}(A) - \mathbb{P}(A \cap B) - \mathbb{P}(A)[1 - \mathbb{P}(B)] = -[\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)] \quad (2.4.14)$$

The Multiplication Rule

Sometimes conditional probabilities are known and can be used to find the probabilities of other events. Note first that if A and B are events with positive probability, then by the very definition of conditional probability,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B | A) = \mathbb{P}(B)\mathbb{P}(A | B) \quad (2.4.15)$$

The following generalization is known as the *multiplication rule* of probability. As usual, we assume that any event conditioned on has positive probability.

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

$$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 | A_1)\mathbb{P}(A_3 | A_1 \cap A_2) \cdots \mathbb{P}(A_n | A_1 \cap A_2 \cap \dots \cap A_{n-1}) \quad (2.4.16)$$

Proof

The product on the right is a collapsing product in which only the probability of the intersection of all n events survives. The product of the first two factors is $\mathbb{P}(A_1 \cap A_2)$, and hence the product of the first three factors is $\mathbb{P}(A_1 \cap A_2 \cap A_3)$, and so forth. The proof can be made more rigorous by induction on n .

The multiplication rule is particularly useful for experiments that consist of dependent stages, where A_i is an event in stage i . Compare the multiplication rule of probability with the multiplication rule of combinatorics.

As with any other result, the multiplication rule can be applied to a conditional probability measure. In the context above, if E is another event, then

$$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n | E) = \mathbb{P}(A_1 | E)\mathbb{P}(A_2 | A_1 \cap E)\mathbb{P}(A_3 | A_1 \cap A_2 \cap E) \cdots \mathbb{P}(A_n | A_1 \cap A_2 \cap \dots \cap A_{n-1} \cap E) \quad (2.4.17)$$

Conditioning and Bayes' Theorem

Suppose that $\mathcal{A} = \{A_i : i \in I\}$ is a countable collection of events that partition the sample space S , and that $\mathbb{P}(A_i) > 0$ for each $i \in I$.

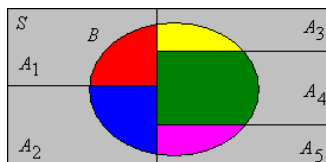


Figure 2.4.2: A partition of S induces a partition of B .

The following theorem is known as the *law of total probability*.

If B is an event then

$$\mathbb{P}(B) = \sum_{i \in I} \mathbb{P}(A_i)\mathbb{P}(B | A_i) \quad (2.4.18)$$

Proof

Recall that $\{A_i \cap B : i \in I\}$ is a partition of B . Hence

$$\mathbb{P}(B) = \sum_{i \in I} \mathbb{P}(A_i \cap B) = \sum_{i \in I} \mathbb{P}(A_i)\mathbb{P}(B | A_i) \quad (2.4.19)$$

The following theorem is known as *Bayes' Theorem*, named after Thomas Bayes:

If B is an event then

$$\mathbb{P}(A_j | B) = \frac{\mathbb{P}(A_j)\mathbb{P}(B | A_j)}{\sum_{i \in I} \mathbb{P}(A_i)\mathbb{P}(B | A_i)}, \quad j \in I \quad (2.4.20)$$

Proof

Again the numerator is $\mathbb{P}(A_j \cap B)$ while the denominator is $\mathbb{P}(B)$ by the [law of total probability](#).

These two theorems are most useful, of course, when we know $\mathbb{P}(A_i)$ and $\mathbb{P}(B | A_i)$ for each $i \in I$. When we compute the probability of $\mathbb{P}(B)$ by the [law of total probability](#), we say that we are *conditioning* on the partition \mathcal{A} . Note that we can think of the sum as a *weighted average* of the conditional probabilities $\mathbb{P}(B | A_i)$ over $i \in I$, where $\mathbb{P}(A_i)$, $i \in I$ are the weight factors. In the context of [Bayes theorem](#), $\mathbb{P}(A_j)$ is the *prior probability* of A_j and $\mathbb{P}(A_j | B)$ is the *posterior probability* of A_j for $j \in I$. We will study more general versions of conditioning and Bayes theorem in the section on Discrete Distributions in the chapter on Distributions, and again in the section on Conditional Expected Value in the chapter on Expected Value.

Once again, the law of total probability and Bayes' theorem can be applied to a conditional probability measure. So, if E is another event with $\mathbb{P}(A_i \cap E) > 0$ for $i \in I$ then

$$\mathbb{P}(B | E) = \sum_{i \in I} \mathbb{P}(A_i | E) \mathbb{P}(B | A_i \cap E) \quad (2.4.21)$$

$$\mathbb{P}(A_j | B \cap E) = \frac{\mathbb{P}(A_j | E) \mathbb{P}(B | A_j \cap E)}{\sum_{i \in I} \mathbb{P}(A_i \cap E) \mathbb{P}(B | A_i \cap E)}, \quad j \in I \quad (2.4.22)$$

Examples and Applications

Basic Rules

Suppose that A and B are events in an experiment with $\mathbb{P}(A) = \frac{1}{3}$, $\mathbb{P}(B) = \frac{1}{4}$, $\mathbb{P}(A \cap B) = \frac{1}{10}$. Find each of the following:

1. $\mathbb{P}(A | B)$
2. $\mathbb{P}(B | A)$
3. $\mathbb{P}(A^c | B)$
4. $\mathbb{P}(B^c | A)$
5. $\mathbb{P}(A^c | B^c)$

Answer

1. $\frac{2}{5}$
2. $\frac{3}{10}$
3. $\frac{3}{5}$
4. $\frac{7}{10}$
5. $\frac{31}{45}$

Suppose that A , B , and C are events in a random experiment with $\mathbb{P}(A | C) = \frac{1}{2}$, $\mathbb{P}(B | C) = \frac{1}{3}$, and $\mathbb{P}(A \cap B | C) = \frac{1}{4}$. Find each of the following:

1. $\mathbb{P}(B \setminus A | C)$
2. $\mathbb{P}(A \cup B | C)$
3. $\mathbb{P}(A^c \cap B^c | C)$
4. $\mathbb{P}(A^c \cup B^c | C)$
5. $\mathbb{P}(A^c \cup B | C)$
6. $\mathbb{P}(A | B \cap C)$

Answer

1. $\frac{1}{12}$
2. $\frac{7}{12}$
3. $\frac{5}{12}$
4. $\frac{3}{4}$
5. $\frac{3}{4}$
6. $\frac{3}{4}$

Suppose that A and B are events in a random experiment with $\mathbb{P}(A) = \frac{1}{2}$, $\mathbb{P}(B) = \frac{1}{3}$, and $\mathbb{P}(A | B) = \frac{3}{4}$.

1. Find $\mathbb{P}(A \cap B)$
2. Find $\mathbb{P}(A \cup B)$

3. Find $\mathbb{P}(B \cup A^c)$
4. Find $\mathbb{P}(B | A)$
5. Are A and B positively correlated, negatively correlated, or independent?

Answer

1. $\frac{1}{4}$
2. $\frac{7}{12}$
3. $\frac{3}{4}$
4. $\frac{1}{2}$
5. positively correlated.

Open the conditional probability experiment.

1. Given $\mathbb{P}(A)$, $\mathbb{P}(B)$, and $\mathbb{P}(A \cap B)$, in the table, verify all of the other probabilities in the table.
2. Run the experiment 1000 times and compare the probabilities with the relative frequencies.

Simple Populations

In a certain population, 30% of the persons smoke cigarettes and 8% have COPD (*Chronic Obstructive Pulmonary Disease*). Moreover, 12% of the persons who smoke have COPD.

1. What percentage of the population smoke and have COPD?
2. What percentage of the population with COPD also smoke?
3. Are smoking and COPD positively correlated, negatively correlated, or independent?

Answer

1. 3.6%
2. 45%
3. positively correlated.

A company has 200 employees: 120 are women and 80 are men. Of the 120 female employees, 30 are classified as managers, while 20 of the 80 male employees are managers. Suppose that an employee is chosen at random.

1. Find the probability that the employee is female.
2. Find the probability that the employee is a manager.
3. Find the conditional probability that the employee is a manager given that the employee is female.
4. Find the conditional probability that the employee is female given that the employee is a manager.
5. Are the events *female* and *manager* positively correlated, negatively correlated, or independent?

Answer

1. $\frac{120}{200}$
2. $\frac{50}{200}$
3. $\frac{30}{120}$
4. $\frac{30}{50}$
5. independent

Dice and Coins

Consider the experiment that consists of rolling 2 standard, fair dice and recording the sequence of scores $\mathbf{X} = (X_1, X_2)$. Let Y denote the sum of the scores. For each of the following pairs of events, find the probability of each event and the conditional probability of each event given the other. Determine whether the events are positively correlated, negatively correlated, or independent.

1. $\{X_1 = 3\}, \{Y = 5\}$
2. $\{X_1 = 3\}, \{Y = 7\}$
3. $\{X_1 = 2\}, \{Y = 5\}$
4. $\{X_1 = 3\}, \{X_1 = 2\}$

Answer

In each case below, the answers are for $\mathbb{P}(A)$, $\mathbb{P}(B)$, $\mathbb{P}(A | B)$, and $\mathbb{P}(B | A)$

1. $\frac{1}{6}, \frac{1}{9}, \frac{1}{4}, \frac{1}{6}$. Positively correlated.
2. $\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}$. Independent.
3. $\frac{1}{6}, \frac{1}{9}, \frac{1}{4}, \frac{1}{6}$. Positively correlated.
4. $\frac{1}{6}, \frac{1}{6}, 0, 0$. Negatively correlated.

Note that positive correlation is not a transitive relation. From the previous exercise, for example, note that $\{X_1 = 3\}$ and $\{Y = 5\}$ are positively correlated, $\{Y = 5\}$ and $\{X_1 = 2\}$ are positively correlated, but $\{X_1 = 3\}$ and $\{X_1 = 2\}$ are negatively correlated (in fact, disjoint).

In dice experiment, set $n = 2$. Run the experiment 1000 times. Compute the empirical conditional probabilities corresponding to the conditional probabilities in the last exercise.

Consider again the experiment that consists of rolling 2 standard, fair dice and recording the sequence of scores $\mathbf{X} = (X_1, X_2)$. Let Y denote the sum of the scores, U the minimum score, and V the maximum score.

1. Find $\mathbb{P}(U = u | V = 4)$ for the appropriate values of u .
2. Find $\mathbb{P}(Y = y | V = 4)$ for the appropriate values of y .
3. Find $\mathbb{P}(V = v | Y = 8)$ for appropriate values of v .
4. Find $\mathbb{P}(U = u | Y = 8)$ for the appropriate values of u .
5. Find $\mathbb{P}[(X_1, X_2) = (x_1, x_2) | Y = 8]$ for the appropriate values of (x_1, x_2) .

Answer

1. $\frac{2}{7}$ for $u \in \{1, 2, 3\}$, $\frac{1}{7}$ for $u = 4$
2. $\frac{2}{7}$ for $y \in \{5, 6, 7\}$, $\frac{1}{7}$ for $y = 8$
3. $\frac{1}{5}$ for $v = 4$, $\frac{2}{5}$ for $v \in \{5, 6\}$
4. $\frac{2}{5}$ for $u \in \{2, 3\}$, $\frac{1}{5}$ for $u = 4$
5. $\frac{1}{5}$ for $(x_1, x_2) \in \{(2, 6), (6, 2), (3, 5), (5, 3), (4, 4)\}$

In the die-coin experiment, a standard, fair die is rolled and then a fair coin is tossed the number of times showing on the die. Let N denote the die score and H the event that all coin tosses result in heads.

1. Find $\mathbb{P}(H)$.
2. Find $\mathbb{P}(N = n | H)$ for $n \in \{1, 2, 3, 4, 5, 6\}$
3. Compare the results in (b) with $\mathbb{P}(N = n)$ for $n \in \{1, 2, 3, 4, 5, 6\}$. In each case, note whether the events H and $\{N = n\}$ are positively correlated, negatively correlated, or independent.

Answer

1. $\frac{21}{128}$
2. $\frac{64}{63} \frac{1}{2^n}$ for $n \in \{1, 2, 3, 4, 5, 6\}$
3. positively correlated for $n \in \{1, 2\}$ and negatively correlated for $n \in \{3, 4, 5, 6\}$

Run the die-coin experiment 1000 times. Let H and N be as defined in the previous exercise.

1. Compute the empirical probability of H . Compare with the true probability in the previous exercise.
2. Compute the empirical probability of $\{N = n\}$ given H , for $n \in \{1, 2, 3, 4, 5, 6\}$. Compare with the true probabilities in the previous exercise.

Suppose that a bag contains 12 coins: 5 are fair, 4 are biased with probability of heads $\frac{1}{3}$; and 3 are two-headed. A coin is chosen at random from the bag and tossed.

1. Find the probability that the coin is heads.
2. Given that the coin is heads, find the conditional probability of each coin type.

Answer

1. $\frac{41}{72}$
2. $\frac{15}{41}$ that the coin is fair, $\frac{8}{41}$ that the coin is biased, $\frac{18}{41}$ that the coin is two-headed

Compare [die-coin experiment](#) and [bag of coins experiment](#). In the die-coin experiment, we toss a coin with a *fixed* probability of heads a *random* number of times. In the bag of coins experiment, we effectively toss a coin with a *random* probability of heads a *fixed* number of times. The random experiment of tossing a coin with a fixed probability of heads p a fixed number of times n is known as the *binomial experiment* with parameters n and p . This is a very basic and important experiment that is studied in more detail in the section on the binomial distribution in the chapter on Bernoulli Trials. Thus, the die-coin and bag of coins experiments can be thought of as modifications of the binomial experiment in which a parameter has been *randomized*. In general, interesting new random experiments can often be constructed by randomizing one or more parameters in another random experiment.

In the coin-die experiment, a fair coin is tossed. If the coin lands tails, a fair die is rolled. If the coin lands heads, an ace-six flat die is tossed (faces 1 and 6 have probability $\frac{1}{4}$ each, while faces 2, 3, 4, and 5 have probability $\frac{1}{8}$ each). Let H denote the event that the coin lands heads, and let Y denote the score when the chosen die is tossed.

1. Find $\mathbb{P}(Y = y)$ for $y \in \{1, 2, 3, 4, 5, 6\}$.
2. Find $\mathbb{P}(H | Y = y)$ for $y \in \{1, 2, 3, 4, 5, 6\}$.
3. Compare each probability in part (b) with $\mathbb{P}(H)$. In each case, note whether the events H and $\{Y = y\}$ are positively correlated, negatively correlated, or independent.

Answer

1. $\frac{5}{24}$ for $y \in \{1, 6\}$, $\frac{7}{48}$ for $y \in \{2, 3, 4, 5\}$
2. $\frac{3}{5}$ for $y \in \{1, 6\}$, $\frac{3}{7}$ for $y \in \{2, 3, 4, 5\}$
3. Positively correlated for $y \in \{1, 6\}$, negatively correlated for $y \in \{2, 3, 4, 5\}$

Run the coin-die experiment 1000 times. Let H and Y be as defined in the previous exercise.

1. Compute the empirical probability of $\{Y = y\}$, for each y , and compare with the true probability in the previous exercise
2. Compute the empirical probability of H given $\{Y = y\}$ for each y , and compare with the true probability in the previous exercise.

Cards

Consider the card experiment that consists of dealing 2 cards 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. For each of the following pairs of events, compute the probability of each event, and the conditional probability of each event given the other. Determine whether the events are positively correlated, negatively correlated, or independent.

1. Q_1, H_1
2. Q_1, Q_2
3. Q_2, H_2
4. Q_1, H_2

Answer

The answers below are for $\mathbb{P}(A)$, $\mathbb{P}(B)$, $\mathbb{P}(A | B)$, and $\mathbb{P}(B | A)$ where A and B are the given events

1. $\frac{1}{13}, \frac{1}{4}, \frac{1}{13}, \frac{1}{4}$, independent.
2. $\frac{1}{13}, \frac{1}{13}, \frac{3}{51}, \frac{3}{51}$, negatively correlated.
3. $\frac{1}{13}, \frac{1}{4}, \frac{1}{13}, \frac{1}{4}$, independent.
4. $\frac{1}{13}, \frac{1}{4}, \frac{1}{13}, \frac{1}{4}$, independent.

In the card experiment, set $n = 2$. Run the experiment 500 times. Compute the conditional relative frequencies corresponding to the conditional probabilities in the last exercise.

Consider the card experiment that consists of dealing 3 cards from a standard deck and recording the sequence of cards dealt. Find the probability of the following events:

1. All three cards are all hearts.
2. The first two cards are hearts and the third is a spade.
3. The first and third cards are hearts and the second is a spade.

Proof

1. $\frac{11}{850}$
2. $\frac{13}{850}$
3. $\frac{13}{850}$

In the card experiment, set $n = 3$ and run the simulation 1000 times. Compute the empirical probability of each event in the previous exercise and compare with the true probability.

Bivariate Uniform Distributions

Recall that Buffon's coin experiment consists of tossing a coin with radius $r \leq \frac{1}{2}$ 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, parallel to the sides. Since the needle is dropped randomly, the basic modeling assumption is that (X, Y) is uniformly distributed on the square $[-1/2, 1/2]^2$.

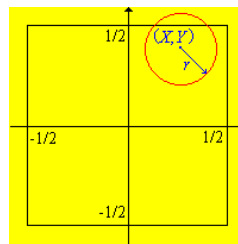


Figure 2.4.3: Buffon's coin experiment

In Buffon's coin experiment,

1. Find $\mathbb{P}(Y > 0 \mid X < Y)$
2. Find the conditional distribution of (X, Y) given that the coin does not touch the sides of the square.

Answer

1. $\frac{3}{4}$
2. Given $(X, Y) \in [r - \frac{1}{2}, \frac{1}{2} - r]^2$, (X, Y) is uniformly distributed on this set.

Run Buffon's coin experiment 500 times. Compute the empirical probability that $Y > 0$ given that $X < Y$ and compare with the probability in the last exercise.

In the conditional probability experiment, the random points are uniformly distributed on the rectangle S . Move and resize events A and B and note how the probabilities change. For each of the following configurations, run the experiment 1000 times and compare the relative frequencies with the true probabilities.

1. A and B in general position
2. A and B disjoint
3. $A \subseteq B$
4. $B \subseteq A$

Reliability

A plant has 3 assembly lines that produces memory chips. Line 1 produces 50% of the chips and has a defective rate of 4%; line 2 has produces 30% of the chips and has a defective rate of 5%; line 3 produces 20% of the chips and has a defective rate of 1%. A chip is chosen at random from the plant.

1. Find the probability that the chip is defective.
2. Given that the chip is defective, find the conditional probability for each line.

Answer

1. 0.037
2. 0.541 for line 1, 0.405 for line 2, 0.054 for line 3

Suppose that a bit (0 or 1) is sent through a noisy communications channel. Because of the noise, the bit sent may be received incorrectly as the complementary bit. Specifically, suppose that if 0 is sent, then the probability that 0 is received is 0.9 and the probability that 1 is received is 0.1. If 1 is sent, then the probability that 1 is received is 0.8 and the probability that 0 is received is 0.2. Finally, suppose that 1 is sent with probability 0.6 and 0 is sent with probability 0.4. Find the probability that

1. 1 was sent given that 1 was received
2. 0 was sent given that 0 was received

Answer

1. 12/13
2. 3/4

Suppose that T denotes the lifetime of a light bulb (in 1000 hour units), and that T has the following exponential distribution, defined for measurable $A \subseteq [0, \infty)$:

$$\mathbb{P}(T \in A) = \int_A e^{-t} dt \quad (2.4.23)$$

1. Find $\mathbb{P}(T > 3)$
2. Find $\mathbb{P}(T > 5 \mid T > 2)$

Answer

1. e^{-3}
2. e^{-3}

Suppose again that T denotes the lifetime of a light bulb (in 1000 hour units), but that T is uniformly distributed on the interval $[0, 10]$.

1. Find $\mathbb{P}(T > 3)$
2. Find $\mathbb{P}(T > 5 \mid T > 2)$

Answer

1. $\frac{7}{10}$
2. $\frac{5}{8}$

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

Genotype	aa	ab	ao	bb	bo	oo
Probability	0.050	0.038	0.310	0.007	0.116	0.479

Suppose that a person is chosen at random from the population. Let A , B , AB , and O be the events that the person is type A , type B , type AB , and type O respectively. Let H be the event that the person is homozygous, and let D denote the event that the person has an o allele. Find each of the following:

1. $\mathbb{P}(A)$, $\mathbb{P}(B)$, $\mathbb{P}(AB)$, $\mathbb{P}(O)$, $\mathbb{P}(H)$, $\mathbb{P}(D)$

1. $P(A \cap H)$, $P(A | H)$, $P(H | A)$. Are the events A and H positively correlated, negatively correlated, or independent?
3. $P(B \cap H)$, $P(B | H)$, $P(H | B)$. Are the events B and H positively correlated, negatively correlated, or independent?
4. $P(A \cap D)$, $P(A | D)$, $P(D | A)$. Are the events A and D positively correlated, negatively correlated, or independent?
5. $P(B \cap D)$, $P(B | D)$, $P(D | B)$. Are the events B and D positively correlated, negatively correlated, or independent?
6. $P(H \cap D)$, $P(H | D)$, $P(D | H)$. Are the events H and D positively correlated, negatively correlated, or independent?

Answer

1. 0.360, 0.123, 0.038, 0.479, 0.536, 0.905
2. 0.050, 0.093, 0.139. A and H are negatively correlated.
3. 0.007, 0.013, 0.057. B and H are negatively correlated.
4. 0.310, 0.343, 0.861. A and D are negatively correlated.
5. 0.116, 0.128, 0.943. B and D are positively correlated.
6. 0.479, 0.529, 0.894. H and D are negatively correlated.

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 y recessive.

Suppose that a green-pod plant and a yellow-pod plant are bred together. Suppose further that the green-pod plant has a $\frac{1}{4}$ chance of carrying the recessive yellow-pod allele.

1. Find the probability that a child plant will have green pods.
2. Given that a child plant has green pods, find the updated probability that the green-pod parent has the recessive allele.

Answer

1. $\frac{7}{8}$
2. $\frac{1}{7}$

Suppose that two green-pod plants are bred together. Suppose further that with probability $\frac{1}{3}$ neither plant has the recessive allele, with probability $\frac{1}{2}$ one plant has the recessive allele, and with probability $\frac{1}{6}$ both plants have the recessive allele.

1. Find the probability that a child plant has green pods.
2. Given that a child plant has green pods, find the updated probability that both parents have the recessive gene.

Answer

1. $\frac{23}{24}$
2. $\frac{3}{23}$

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 in a certain population, 50% are male and 50% are female. Moreover, suppose that 10% of males are color blind but only 1% of females are color blind.

1. Find the percentage of color blind persons in the population.
2. Find the percentage of color blind persons that are male.

Answer

1. 5.5%
2. 90.9%

Since color blindness is a sex-linked hereditary disorder, note that it's reasonable in the previous exercise that the probability that a female is color blind is the square of the probability that a male is color blind. If p is the probability of the defective allele on the X chromosome, then p is also the probability that a male will be color blind. But since the defective allele is recessive, a woman would need two copies of the defective allele to be color blind, and assuming independence, the probability of this event is p^2 .

A man and a woman do not have a certain sex-linked hereditary disorder, but the woman has a $\frac{1}{3}$ chance of being a carrier.

1. Find the probability that a son born to the couple will be normal.

2. Find the probability that a daughter born to the couple will be a carrier.
3. Given that a son born to the couple is normal, find the updated probability that the mother is a carrier.

Answer

1. $\frac{5}{6}$
2. $\frac{1}{6}$
3. $\frac{1}{5}$

Urn Models

Urn 1 contains 4 red and 6 green balls while urn 2 contains 7 red and 3 green balls. An urn is chosen at random and then a ball is chosen at random from the selected urn.

1. Find the probability that the ball is green.
2. Given that the ball is green, find the conditional probability that urn 1 was selected.

Answer

1. $\frac{9}{20}$
2. $\frac{2}{3}$

Urn 1 contains 4 red and 6 green balls while urn 2 contains 6 red and 3 green balls. A ball is selected at random from urn 1 and transferred to urn 2. Then a ball is selected at random from urn 2.

1. Find the probability that the ball from urn 2 is green.
2. Given that the ball from urn 2 is green, find the conditional probability that the ball from urn 1 was green.

Answer

1. $\frac{9}{25}$
2. $\frac{2}{3}$

An urn initially contains 6 red and 4 green balls. A ball is chosen at random from the urn and its color is recorded. It is then replaced in the urn and 2 new balls of the same color are added to the urn. The process is repeated. Find the probability of each of the following events:

1. Balls 1 and 2 are red and ball 3 is green.
2. Balls 1 and 3 are red and ball 2 is green.
3. Ball 1 is green and balls 2 and 3 are red.
4. Ball 2 is red.
5. Ball 1 is red given that ball 2 is red.

Answer

1. $\frac{4}{35}$
2. $\frac{4}{35}$
3. $\frac{4}{35}$
4. $\frac{3}{5}$
5. $\frac{2}{3}$

Think about the results in the previous exercise. Note in particular that the answers to parts (a), (b), and (c) are the same, and that the probability that the second ball is red in part (d) is the same as the probability that the first ball is red. More generally, the probabilities of events do not depend on the order of the draws. For example, the probability of an event involving the first, second, and third draws is the same as the probability of the corresponding event involving the seventh, tenth and fifth draws. Technically, the sequence of events (R_1, R_2, \dots) is exchangeable. The random process described in this exercise is a special case of *Pólya's urn scheme*, named after George Pólya. We will study Pólya's urn in more detail in the chapter on Finite Sampling Models

An urn initially contains 6 red and 4 green balls. A ball is chosen at random from the urn and its color is recorded. It is then replaced in the urn and two new balls of the *other* color are added to the urn. The process is repeated. Find the probability of

each of the following events:

1. Balls 1 and 2 are red and ball 3 is green.
2. Balls 1 and 3 are red and ball 2 is green.
3. Ball 1 is green and balls 2 and 3 are red.
4. Ball 2 is red.
5. Ball 1 is red given that ball 2 is red.

Answer

1. $\frac{6}{35}$
2. $\frac{6}{35}$
3. $\frac{16}{105}$
4. $\frac{17}{30}$
5. $\frac{9}{17}$

Think about the results in the previous exercise, and compare with [Pólya's urn](#). Note that the answers to parts (a), (b), and (c) are not all the same, and that the probability that the second ball is red in part (d) is not the same as the probability that the first ball is red. In short, the sequence of events (R_1, R_2, \dots) is *not* exchangeable.

Diagnostic Testing

Suppose that we have a random experiment with an event A of interest. When we run the experiment, of course, event A will either occur or not occur. However, suppose that we are not able to observe the occurrence or non-occurrence of A directly. Instead we have a *diagnostic test* designed to indicate the occurrence of event A ; thus the test that can be either *positive* for A or *negative* for A . The test also has an element of randomness, and in particular can be in error. Here are some typical examples of the type of situation we have in mind:

- The event is that a person has a certain disease and the test is a blood test for the disease.
- The event is that a woman is pregnant and the test is a home pregnancy test.
- The event is that a person is lying and the test is a lie-detector test.
- The event is that a device is defective and the test consists of a sensor reading.
- The event is that a missile is in a certain region of airspace and the test consists of radar signals.
- The event is that a person has committed a crime, and the test is a jury trial with evidence presented for and against the event.

Let T be the event that the test is positive for the occurrence of A . The conditional probability $\mathbb{P}(T | A)$ is called the *sensitivity* of the test. The complementary probability

$$\mathbb{P}(T^c | A) = 1 - \mathbb{P}(T | A) \quad (2.4.24)$$

is the *false negative* probability. The conditional probability $\mathbb{P}(T^c | A^c)$ is called the *specificity* of the test. The complementary probability

$$\mathbb{P}(T | A^c) = 1 - \mathbb{P}(T^c | A^c) \quad (2.4.25)$$

is the *false positive* probability. In many cases, the sensitivity and specificity of the test are known, as a result of the development of the test. However, the *user* of the test is interested in the opposite conditional probabilities, namely $\mathbb{P}(A | T)$, the probability of the event of interest, given a positive test, and $\mathbb{P}(A^c | T^c)$, the probability of the complementary event, given a negative test. Of course, if we know $\mathbb{P}(A | T)$ then we also have $\mathbb{P}(A^c | T) = 1 - \mathbb{P}(A | T)$, the probability of the complementary event given a positive test. Similarly, if we know $\mathbb{P}(A^c | T^c)$ then we also have $\mathbb{P}(A | T^c)$, the probability of the event given a negative test. Computing the probabilities of interest is simply a special case of Bayes' theorem.

The probability that the event occurs, given a positive test is

$$\mathbb{P}(A | T) = \frac{\mathbb{P}(A)\mathbb{P}(T | A)}{\mathbb{P}(A)\mathbb{P}(T | A) + \mathbb{P}(A^c)\mathbb{P}(T | A^c)} \quad (2.4.26)$$

The probability that the event does not occur, given a negative test is

$$\mathbb{P}(A^c | T^c) = \frac{\mathbb{P}(A^c)\mathbb{P}(T^c | A^c)}{\mathbb{P}(A)\mathbb{P}(T^c | A) + \mathbb{P}(A^c)\mathbb{P}(T^c | A^c)} \quad (2.4.27)$$

There is often a trade-off between sensitivity and specificity. An attempt to make a test more sensitive may result in the test being less specific, and an attempt to make a test more specific may result in the test being less sensitive. As an extreme example, consider the worthless test that always returns positive, no matter what the evidence. Then $T = S$ so the test has sensitivity 1, but specificity 0. At the opposite extreme is the worthless test that always returns negative, no matter what the evidence. Then $T = \emptyset$ so the test has specificity 1 but sensitivity 0. In between these extremes are helpful tests that are actually based on evidence of some sort.

Suppose that the sensitivity $a = \mathbb{P}(T | A) \in (0, 1)$ and the specificity $b = \mathbb{P}(T^c | A^c) \in (0, 1)$ are fixed. Let $p = \mathbb{P}(A)$ denote the prior probability of the event A and $P = \mathbb{P}(A | T)$ the posterior probability of A given a positive test.

P as a function of p is given by

$$P = \frac{ap}{(a+b-1)p + (1-b)}, \quad p \in [0, 1] \quad (2.4.28)$$

1. P increases continuously from 0 to 1 as p increases from 0 to 1.
2. P is concave downward if $a + b > 1$. In this case A and T are positively correlated.
3. P is concave upward if $a + b < 1$. In this case A and T are negatively correlated.
4. $P = p$ if $a + b = 1$. In this case, A and T are uncorrelated (independent).

Proof

The formula for P in terms of p follows from (42) and algebra. For part (a), note that

$$\frac{dP}{dp} = \frac{a(1-b)}{[(a+b-1)p + (1-b)]^2} > 0 \quad (2.4.29)$$

For parts (b)-(d), note that

$$\frac{d^2P}{dp^2} = \frac{-2a(1-b)(a+b-1)}{[(1+b-1)p + (1-b)]^3} \quad (2.4.30)$$

If $a + b > 1$, $d^2P/dp^2 < 0$ so P is concave downward on $[0, 1]$ and hence $P > p$ for $0 < p < 1$. If $a + b < 1$, $d^2P/dp^2 > 0$ so P is concave upward on $[0, 1]$ and hence $P < p$ for $0 < p < 1$. Trivially if $a + b = 1$, $P = p$ for $0 \leq p \leq 1$.

Of course, part (b) is the typical case, where the test is useful. In fact, we would hope that the sensitivity and specificity are close to 1. In case (c), the test is worse than useless since it gives the wrong information about A . But this case could be turned into a useful test by simply reversing the roles of *positive* and *negative*. In case (d), the test is worthless and gives no information about A . It's interesting that the broad classification above depends only on the *sum* of the sensitivity and specificity.

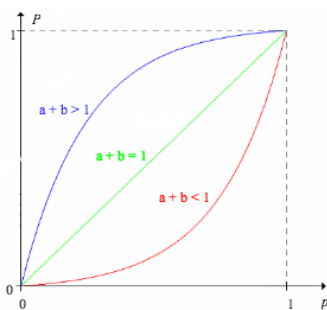


Figure 2.4.4: $P = \mathbb{P}(A | T)$ as a function of $p = \mathbb{P}(A)$ in the three cases

Suppose that a diagnostic test has sensitivity 0.99 and specificity 0.95. Find $\mathbb{P}(A | T)$ for each of the following values of $\mathbb{P}(A)$:

1. 0.001
2. 0.01
3. 0.2
4. 0.5
5. 0.7
6. 0.9

Answer

1. 0.0194
2. 0.1667
3. 0.8319
4. 0.9519
5. 0.9788
6. 0.9944

With sensitivity 0.99 and specificity 0.95, the test in the last exercise superficially looks good. However the small value of $\mathbb{P}(A | T)$ for small values of $\mathbb{P}(A)$ is striking (but inevitable given the [properties above](#)). The moral, of course, is that $\mathbb{P}(A | T)$ depends critically on $\mathbb{P}(A)$ not just on the sensitivity and specificity of the test. Moreover, the correct comparison is $\mathbb{P}(A | T)$ with $\mathbb{P}(A)$, as in the exercise, not $\mathbb{P}(A | T)$ with $\mathbb{P}(T | A)$ —Beware of the fallacy of the transposed conditional! In terms of the correct comparison, the test does indeed work well; $\mathbb{P}(A | T)$ is significantly larger than $\mathbb{P}(A)$ in all cases.

A woman initially believes that there is an even chance that she is or is not pregnant. She takes a home pregnancy test with sensitivity 0.95 and specificity 0.90 (which are reasonable values for a home pregnancy test). Find the updated probability that the woman is pregnant in each of the following cases.

1. The test is positive.
2. The test is negative.

Answer

1. 0.905
2. 0.053

Suppose that 70% of defendants brought to trial for a certain type of crime are guilty. Moreover, historical data show that juries convict guilty persons 80% of the time and convict innocent persons 10% of the time. Suppose that a person is tried for a crime of this type. Find the updated probability that the person is guilty in each of the following cases:

1. The person is convicted.
2. The person is acquitted.

Answer

1. 0.949
2. 0.341

The “Check Engine” light on your car has turned on. Without the information from the light, you believe that there is a 10% chance that your car has a serious engine problem. You learn that if the car has such a problem, the light will come on with probability 0.99, but if the car does not have a serious problem, the light will still come on, under circumstances similar to yours, with probability 0.3. Find the updated probability that you have an engine problem.

Answer

0.268

The standard test for HIV is the *ELISA* (*Enzyme-Linked Immunosorbent Assay*) test. It has sensitivity and specificity of 0.999. Suppose that a person is selected at random from a population in which 1% are infected with HIV, and given the ELISA test. Find the probability that the person has HIV in each of the following cases:

1. The test is positive.
2. The test is negative.

Answer

1. 0.9098
2. 0.00001

The ELISA test for HIV is a very good one. Let's look another test, this one for prostate cancer, that's rather bad.

The PSA test for prostate cancer is based on a blood marker known as the *Prostate Specific Antigen*. An elevated level of PSA is evidence for prostate cancer. To have a diagnostic test, in the sense that we are discussing here, we must decide on a definite level of PSA, above which we declare the test to be positive. A positive test would typically lead to other more invasive tests (such as biopsy) which, of course, carry risks and cost. The PSA test with cutoff 2.6 ng/ml has sensitivity 0.40 and specificity 0.81. The overall incidence of prostate cancer among males is 156 per 100000. Suppose that a man, with no particular risk factors, has the PSA test. Find the probability that the man has prostate cancer in each of the following cases:

1. The test is positive.
2. The test is negative.

Answer

1. 0.00328
2. 0.00116

Diagnostic testing is closely related to a general statistical procedure known as *hypothesis testing*. A separate chapter on hypothesis testing explores this procedure in detail.

Data Analysis Exercises

For the M&M data set, find the empirical probability that a bag has at least 10 reds, given that the weight of the bag is at least 48 grams.

Answer

$$\frac{10}{23}.$$

Consider the Cicada data.

1. Find the empirical probability that a cicada weighs at least 0.25 grams given that the cicada is male.
2. Find the empirical probability that a cicada weighs at least 0.25 grams given that the cicada is the tredecula species.

Answer

1. $\frac{2}{45}$
2. $\frac{7}{44}$

This page titled [2.4: Conditional Probability](#) 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.