

Probability

OPRE 6301

Random Experiment...

Recall that our eventual goal in this course is to go from the random sample to the population. The theory that allows for this transition is the theory of probability.

A **random experiment** is an action or process that leads to one of many possible **outcomes**. Examples:

Experiment	Outcomes
Flip a coin	Heads, Tails
Roll a die	Numbers: 1, 2, 3, 4, 5, 6
Exam Marks	Numbers: (0, 100)
Course Grades	F, D, C, B, A
Task completion times	Nonnegative values

The list of possible outcomes of a random experiment must be **exhaustive** and **mutually exclusive**.

Sample Space. . .

The set of all possible outcomes of an experiment is called the **sample space**.

We will denote the outcomes by O_1, O_2, \dots , and the sample space by S . Thus, in set-theory notation,

$$S = \{O_1, O_2, \dots\}$$

Events...

An *individual* outcome in the sample space is called a **simple event**, while...

An **event** is a collection or set of one or more simple events in a sample space.

Example: Roll of a Die

$$S = \{1, 2, \dots, 6\}$$

Simple Event: The outcome “3”.

Event: The outcome is an even number (one of 2, 4, 6)

Event: The outcome is a low number (one of 1, 2, 3)

Assigning Probabilities...

Requirements

Given a sample space $S = \{O_1, O_2, \dots\}$, the **probabilities** assigned to events must satisfy these requirements:

1. The probability of any event must be nonnegative, e.g., $P(O_i) \geq 0$ for each i .
2. The probability of the entire sample space must be 1, i.e., $P(S) = 1$.
3. For two *disjoint* events A and B , the probability of the *union* of A and B is equal to the sum of the probabilities of A and B , i.e.,

$$P(A \cup B) = P(A) + P(B).$$

Approaches

There are three ways to assign probabilities to events: **classical approach**, **relative-frequency approach**, **subjective approach**. Details...

Classical Approach...

If an experiment has n simple outcomes, this method would assign a probability of $1/n$ to each outcome. In other words, each outcome is assumed to have an equal probability of occurrence.

This method is also called the **axiomatic** approach.

Example 1: Roll of a Die

$$S = \{1, 2, \dots, 6\}$$

Probabilities: Each simple event has a $1/6$ chance of occurring.

Example 2: Two Rolls of a Die

$$S = \{(1, 1), (1, 2), \dots, (6, 6)\}$$

Assumption: The two rolls are “independent.”

Probabilities: Each simple event has a $(1/6) \cdot (1/6) = 1/36$ chance of occurring.

Relative-Frequency Approach...

Probabilities are assigned on the basis of experimentation or historical data.

Formally, Let A be an event of interest, and assume that you have performed the same experiment n times so that n is the number of times A *could have* occurred. Further, let n_A be the number of times that A did occur. Now, consider the *relative frequency* n_A/n . Then, in this method, we “attempt” to define $P(A)$ as:

$$P(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n}.$$

The above can only be viewed as an attempt because it is not physically feasible to repeat an experiment an infinite number of times. Another important issue with this definition is that *two* sets of n experiments will typically result in two *different* ratios. However, we expect the discrepancy to converge to 0 for large n . Hence, for large n , the ratio n_A/n may be taken as a reasonable approximation for $P(A)$.

Example 1: Roll of a Die

$$S = \{1, 2, \dots, 6\}$$

Probabilities: Roll the given die 100 times (say) and suppose the number of times the outcome 1 is observed is 15. Thus, $A = \{1\}$, $n_A = 15$, and $n = 100$. Therefore, we say that $P(A)$ is approximately equal to $15/100 = 0.15$.

Example 2: Computer Sales

A computer store tracks the daily sales of desktop computers in the past 30 days.

The resulting data is:

Desktops Sold	No. of Days
0	1
1	2
2	10
3	12
4	5
5 or more	0

The approximate probabilities are:

Desktops Sold	No. of Days	Probability
0	1	$1/30 = 0.03$
1	2	$2/30 = 0.07$
2	10	$10/30 = 0.33$
3	12	$12/30 = 0.40$
4	5	$5/30 = 0.17$
5 or more	0	0

Thus, for example, there is a 40% chance that the store will sell 3 desktops on any given day.

Subjective Approach...

In the subjective approach, we define probability as the degree of *belief* that we hold in the occurrence of an event. Thus, judgment is used as the basis for assigning probabilities.

Notice that the classical approach of assigning equal probabilities to simple events is, in fact, also based on judgment.

What is somewhat different here is that the use of the subjective approach is usually limited to experiments that are *unrepeatable*.

Example 1: Horse Race

Consider a horse race with 8 horses running. What is the probability for a particular horse to win? Is it reasonable to assume that the probability is $1/8$? Note that we can't apply the relative-frequency approach.

People regularly place bets on the outcomes of such "one-time" experiments based on their judgment as to how likely it is for a particular horse to win. Indeed, having different judgments is what makes betting possible!

Example 2: Stock Price

What is the probability for a particular stock to go up tomorrow? Again, this "experiment" can't be repeated, and we can't apply the relative-frequency approach.

Sophisticated models (that rely on past data) are often used to make such predictions, as blindly following ill-founded judgments is often dangerous.

Basic Rules of Probability...

All three definitions of probability must follow the same rules. We now describe some basic concepts and rules.

Complement

Let A be an event. The complement of A , denoted by A^c , corresponds to the event that A does *not* occur. By definition, we have $A \cap A^c = \emptyset$ (the empty set) and $A \cup A^c = S$. Here, the intersection operation \cap corresponds to “and”; and the union operation \cup corresponds to “or”.

Since $1 = P(S) = P(A \cup A^c) = P(A) + P(A^c)$, we have

$$P(A^c) = 1 - P(A).$$

Example: Roll of a Die

$$P(\{3\}) = 1/6$$

$$P(\text{the outcome is } \textit{not} \text{ a } 3) = 1 - 1/6 = 5/6$$

Union/Addition

Let A and B be two events. Then,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

The subtraction of $P(A \cap B)$ is necessary because A and B may “overlap.” If A and B are *mutually exclusive*, i.e., $A \cap B = \emptyset$, then

$$P(A \cup B) = P(A) + P(B).$$

Example: Roll of a Die

$$P(\text{even}) = 3/6 \text{ and } P(\text{low}) = 3/6$$

$$P(\text{even and low}) = P(\{2\}) = 1/6$$

$$P(\text{even or low}) = 3/6 + 3/6 - 1/6 = 5/6$$

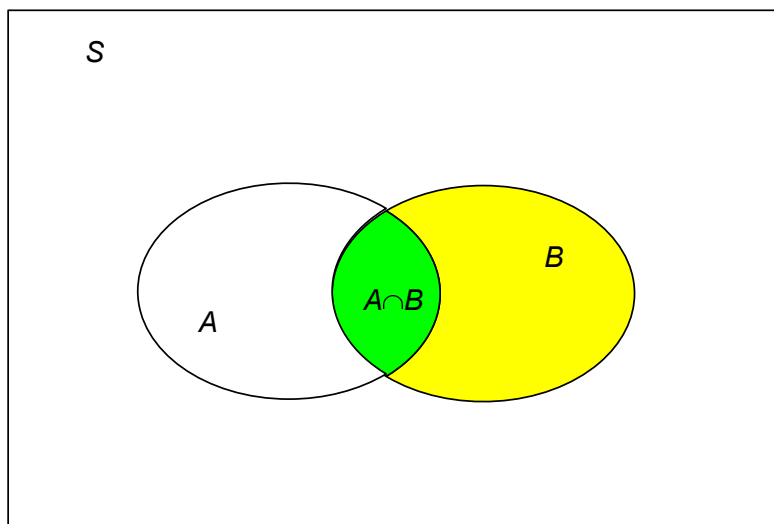
$$P(\{1\} \text{ or } \{6\}) = 1/6 + 1/6 - 0 = 2/6$$

Conditional Probability

Let A and B be two events. Then, the *conditional* probability of A given that B has occurred, $P(A \mid B)$, is defined as:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}. \quad (1)$$

The reasoning behind this definition is that if B has occurred, then only the “portion” of A that is contained in B , i.e., $A \cap B$, could occur; moreover, the original probability of $A \cap B$ must be recalculated to reflect the fact that the “new” sample space is B .



Venn Diagram

Example: Pick a Card from a Deck

Suppose a card is drawn randomly from a deck and found to be an Ace. What is the conditional probability for this card to be Spade Ace?

$A = \text{Spade Ace}$

$B = \text{an Ace}$

$A \cap B = \text{Spade Ace}$

$P(A) = 1/52$; $P(B) = 4/52$; and $P(A \cap B) = 1/52$

Hence,

$$P(A | B) = \frac{1/52}{4/52} = \frac{1}{4}.$$

Multiplication

The *multiplication* rule is used to calculate the *joint* probability of two events. It is simply a rearrangement of the conditional probability formula; see (1). Formally,

$$P(A \cap B) = P(A | B)P(B);$$

or,

$$P(A \cap B) = P(B | A)P(A).$$

Example 1: Drawing a Spade Ace

A = an Ace

B = a Spade

$A \cap B$ = the Spade Ace

$P(B) = 13/52$; $P(A | B) = 1/13$

Hence,

$$\begin{aligned} P(A \cap B) &= P(A | B)P(B) \\ &= \frac{1}{13} \cdot \frac{13}{52} = \frac{1}{52}. \end{aligned}$$

Example 2: Selecting Students

A statistics course has seven male and three female students. The professor wants to select two students at random to help her conduct a research project. What is the probability that the two students chosen are female?

A = the first student selected is female

B = the second student selected is female

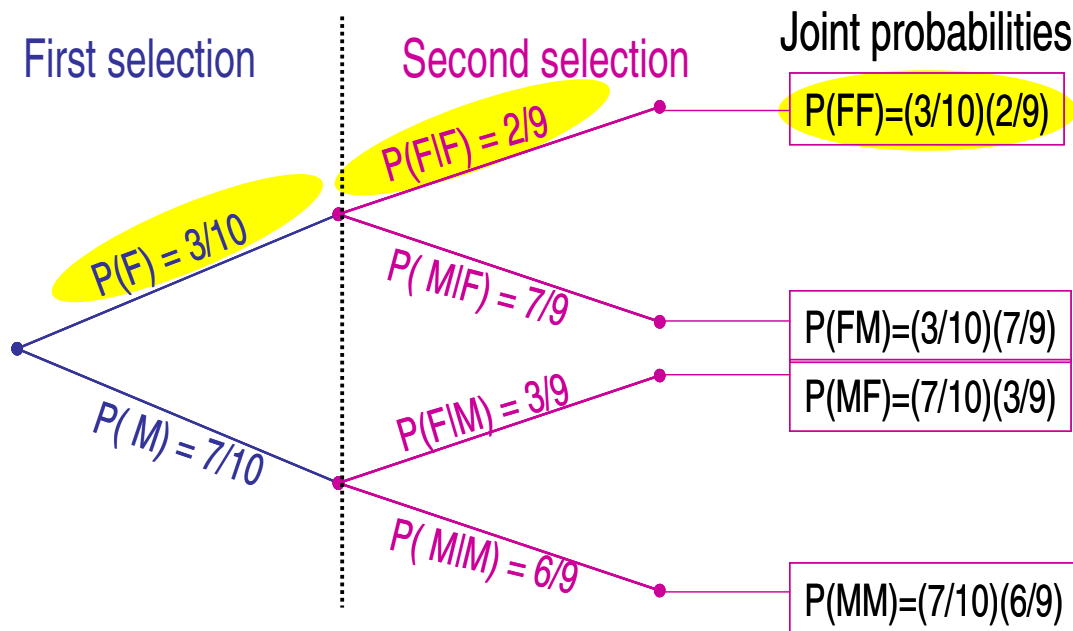
$A \cap B$ = both chosen students are female

$$P(A) = 3/10; P(B | A) = 2/9$$

Hence,

$$\begin{aligned} P(A \cap B) &= P(B | A)P(A) \\ &= \frac{2}{9} \cdot \frac{3}{10} \\ &= \frac{1}{15}. \end{aligned}$$

The calculation above can be visualized as multiplying probabilities along the branches of a **probability tree**:



This is a helpful approach to computing *joint* probabilities.

Independence

Two events are said to be *independent* if the occurrence of either one of the two events does *not* affect the occurrence probability of the other event. This is an important concept, and is formally stated as: Two events A and B are independent if

$$P(A \mid B) = P(A);$$

or,

$$P(B \mid A) = P(B).$$

Note that the above is equivalent to

$$P(A \cap B) = P(A)P(B).$$

Example: Drawings with Replacement

Two balls are successively drawn from an urn that contains seven red balls and three black balls. The first ball drawn is put back into the urn after noting its color. What is the probability that the two balls drawn are both black?

A = the first ball drawn is black

B = the second ball drawn is black

$A \cap B$ = both balls are black

$P(A) = 3/10$; $P(B | A) = P(B) = 3/10$, i.e., B and A
are independent

Hence,

$$\begin{aligned} P(A \cap B) &= P(B | A)P(A) \\ &= \frac{3}{10} \cdot \frac{3}{10} \\ &= \frac{9}{100}. \end{aligned}$$

Bayes' Law...

Suppose we know the conditional probabilities of an event for all possible “causes” of the event. We can use this information to find the probability of the possible cause, given that this event has occurred.

Example: Multiple-Choice Exam

In a multiple-choice exam, each question has m possible answers, but only one of them is correct. Suppose a student adopts the strategy of picking one of the possible answers randomly whenever he does *not* know the correct answer to a question. Assume that the probability for the student to know the answer to a question is p .

Suppose the student answered a question correctly. What is the conditional probability for the student to have truly known the correct answer?

K = the student truly knows the answer

C = the student answered a question correctly

$$P(K | C) = ?$$

Let K^c = the student does *not* know the answer

We know that:

$$P(K) = p$$

$$P(K^c) = 1 - p$$

$$P(C | K) = 1$$

$$P(C | K^c) = 1/m$$

From (1), we have:

$$P(K | C) = \frac{P(K \cap C)}{P(C)}$$

Now,

$$\begin{aligned} P(K \cap C) &= P(C \cap K) \\ &= P(C | K)P(K) \\ &= 1 \cdot p \end{aligned}$$

To compute $P(C)$, observe that

$$P(C) = P(C \cap K) + P(C \cap K^c),$$

and that the second probability above is

$$\begin{aligned} P(C \cap K^c) &= P(C \mid K^c)P(K^c) \\ &= \frac{1}{m} \cdot (1 - p). \end{aligned}$$

Hence,

$$P(C) = 1 \cdot p + \frac{1}{m} \cdot (1 - p).$$

It follows that

$$P(K \mid C) = \frac{p}{p + (1 - p)/m}.$$

Summary:

$$P(K \mid C) = \frac{P(C \mid K)P(K)}{P(C \mid K)P(K) + P(C \mid K^c)P(K^c)}.$$

Note that $P(K \mid C)$ is expressed in terms of both $P(C \mid K)$ and $P(C \mid K^c)$.

Bayes' Law

Let B_1, B_2, \dots, B_k be k mutually exclusive events (or “causes”) such that

$$\sum_{i=1}^k P(B_i) = 1.$$

Then, for any event A , the **Bayes' Law** is:

$$P(B_i | A) = \frac{P(A | B_i)P(B_i)}{\sum_{i=1}^k P(A | B_i)P(B_i)}. \quad (2)$$

The probabilities $P(B_1), P(B_2), \dots, P(B_k)$ are called **prior** probabilities.

The probabilities $P(B_1 | A), P(B_2 | A), \dots, P(B_k | A)$ are called **posterior** probabilities.

The transformation from the prior probabilities to the posterior probabilities is called a *Bayesian update*.

Divide and Conquer...

The method we used to compute the denominator of (2) is extremely powerful. Namely,

$$P(A) = \sum_{i=1}^k P(A \mid B_i)P(B_i). \quad (3)$$

The approach in (3) can be viewed as *divide and conquer*:

Step 1: Divide the sample space S into B_1, B_2, \dots

Step 2: Conquer the probabilities $P(A \cap B_i)$ for $i = 1, 2, \dots$ via

$$P(A \cap B_i) = P(A \mid B_i)P(B_i)$$

Step 3: Compute $P(A)$ by summing the pieces in Step 2

This approach can also be implemented by constructing a *probability tree*.