Handout 1: Basic Probability Theory

Suggested readings: Casella and Berger (2001), Sections 1.1 – 1.3.

1.1 Introduction

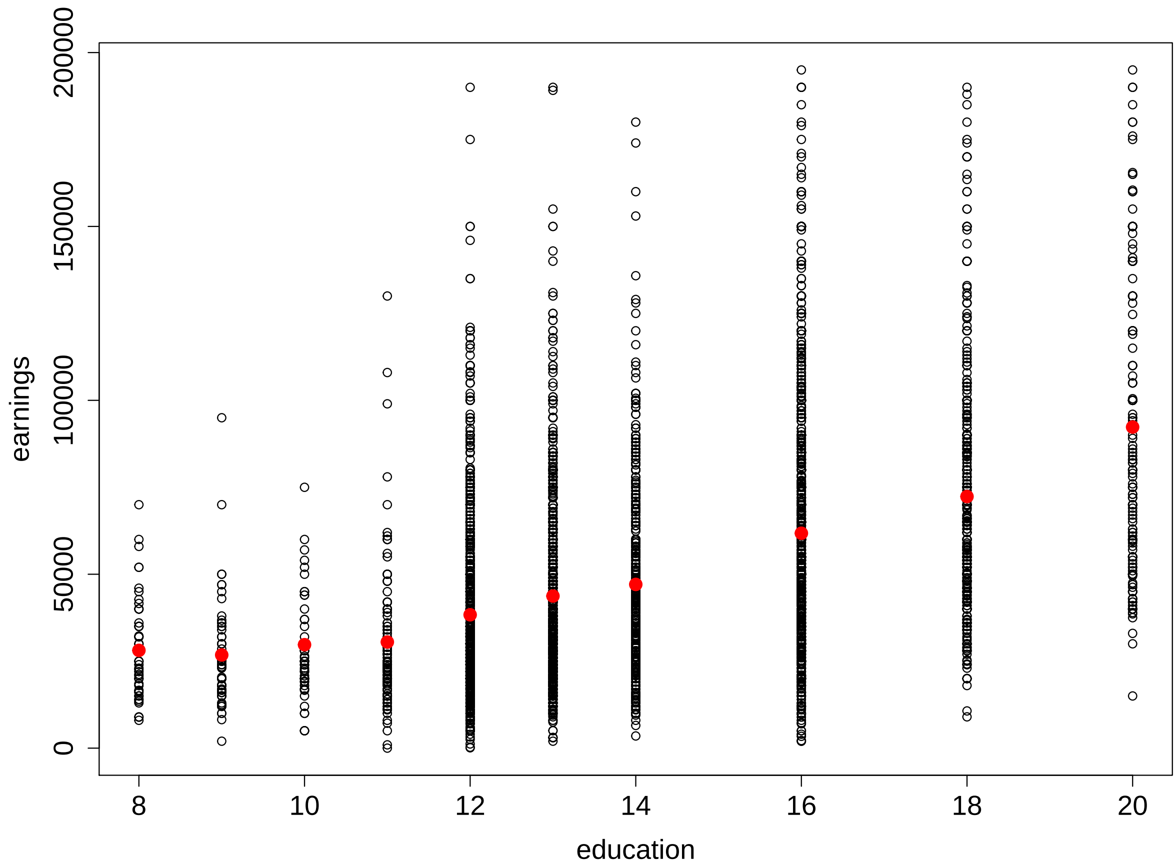
Economic models often involve uncertainty. For example, a firm might want to choose the quantities it produces and the prices of its products to maximize profit. The firm might be fully aware of the costs structure, but it might be uncertain about the demand of consumers. Hence, the firm does not know the exact profit when producing given quantities and asking for certain prices. Agents might also be uncertain about the strategic behavior of other agents. For example, Aldi's decision to open a supermarket in a new geographic market depends on the demand of consumers, but also on the behavior of the competitors. If in response to Aldi's entry, Lidl will also open a new supermarket in the same market, Aldi's profit will suffer. Since Aldi might be uncertain about Lidl's strategic behavior, Aldi does not know the exact profit a new supermarket will produce and whether opening a new store is actually a good idea. Similarly, in macroeconomic models, where agents maximize future discounted utility subject to aggregate resource constraints, agents are typically uncertain about the exact future states of the economy.

In all these situations, we can make progress by modeling the uncertainty. For example, we could assume that Aldi knows how likely it is that Lidl enters the market as well. That is, Aldi knows the *probability* that Lidl enters. Then, instead of maximizing the exact profit, Aldi can base its decision on the *expected* profit. Similarly, we could assume that an agent knows how likely the different potential future states are. The agent can then maximize the expected future discounted utility.

Uncertainty is modeled using probability theory. Specifically, uncertainly is typically introduced as *random variables*, which are, loosely speaking, numerical outcomes with an element of chance to them. For example, if demand of consumers is uncertain, it is typically assumed to be a random variable. How likely the different possible outcomes are is described by the *distribution function* of a random variable, which can then be used to calculate expected values. Models often contain more than one source of uncertainty. For example, there might be many different states in the future that an agent is uncertain about. Sometimes an agent already has information about some of these states and then makes a decision given the available information. In these cases, we need to think about so called *conditional probabilities* and *conditional distributions*.

More specifically, in econometrics we will learn about applying statistical methods to analyze relationships between variables, estimate economic model, test economic theories, and use the estimated relationships for policy evaluations, forecasting, and decision making. As an example, suppose we have a very simple model, which posits a positive relationship between individual earnings and education. Of course, there is no exact relationships between education and earnings and some highly educated people might earn less than people with a high school degree only. Again, there is some uncertainty in our model. To give empirical content to this relationship, suppose we have a data set of $2,500$ individuals, which contains, among others, information about their yearly earnings and their years of education. Figure 1.1 shows a scatter plot of the data, which is a subset of the US Consumer Population Survey (CPS). Here, earnings is measured in USD and education is measured in years.

Figure 1.1: Earnings and education



We can see from the scatter plot that people with more education tend to earn more than people with less education. However, as already discussed, there is no exact relationship, and some people with only 12 years of education earn more than $\$150,000$ a year while some people with 18 years of education earn significantly less. To summarize the data, we could look at the average earnings for the different years of education, which are the solid red circles in the figure. We can see that these sample averages are not monotone in education. For example, in this sample, people with 9 years of education earn on average less than people with 8 years of education. However, this difference is very small and it seems plausible that a different data set yields different results. That is, if we took another sample of $2,500$ individuals from the CPS, maybe people with 9 years of education would on average have higher earnings than people with 8 years of education.

Therefore, similar to economic models more generally, we want to start out by thinking about the *population* distribution of the data, rather than a specific *sample*. As you might know, a feature of population distributions is for example the expected value, while a sample average is a feature of a specific sample. Intuitively, for now, you could think about the population as an arbitrary large or infinite sample. Ideally, we then would like to make claims about the population, using the data from a specific sample.

To make all the above statements more precise and to be able to write down simple economic and econometric models, we have to start with some important concepts in probability theory and statistics, including random variables and their properties, modes of convergence, point estimation, hypothesis testing, and confidence intervals. This is exactly what we will do in the first half of the semester. While this material is quite theoretical and involves a lot of definitions, formal results, and proofs, it is essential to understand econometric methods and to develop a deeper understanding of economic models. The main reference for the first part of the class is Casella and Berger (2001) and we will cover parts of chapters 1 – 9. Casella and Berger write in the preface of their book that "the book can be used for a two-semester (...) course in statistics", but we clearly do not want to spend that much time on the material. We will therefore focus on the parts of the material that are most important for economics in general and econometrics in particular, which implies that we will skip many sections. These lecture notes are meant to summarize all important results, but I still recommend that you read the corresponding sections in Casella and Berger (2001), especially for proofs of many of the results and additional examples. In this handout, we will start with basic concepts in probability theory. For more details see Casella and Berger (2001), Sections 1.1 - 1.3.

You are probably aware that students taking this class have heterogeneous backgrounds and many of you might have seen at least parts of the material before. The main goals of this class is to bring all students to a similar level, to provide you with the necessary tools to carry out empirical research in economics, and to get closer to the forefront of research in econometrics.

1.2 Set theory

To define the most important concepts in probability theory, we will start by thinking about an experiment, which can result in several possible outcomes. As a simple example, suppose we roll a die with six faces. Then the possible outcomes are 1, 2, 3, 4, 5, and 6. This simple concept gives us our first definition.

Definition 1.1. The set of all possible outcomes of an experiment is called the *sample space* and is denoted by S.

When we roll a die we have $S = \{1, 2, 3, 4, 5, 6\}$. The sample space does not have to consist of numerical values. For example, when we flip a coin, the outcomes could be *heads* and *tails* in which case $S = \{heads, tails\}$. The sample space could also be infinite. For example, suppose we throw a dart at a dartboard and measure the distance to the center. Since there is no natural upper bound, we could have $S = [0, \infty)$. In general, we will distinguish between sample spaces that are *countable* or *uncountable*. Countable simply means that we can attach an integer to every possible outcome in the sample space. Therefore, every sample space that consists of finitely many elements is countable, but intervals such as $[0, 1]$ are not countable.

Once we know the sample space, we can also define subsets of it.

Definition 1.2. A subset of the sample space, denoted by $A \subset S$, is called an *event*.

An event is by definition simply a collection of possible outcomes of the experiment, such as $A = \{1, 3, 5\}$ in the die rolling example.

We will continue talking about sets and subsets for a while. It is therefore important to quickly discuss (hopefully review) some basic set operations. Here are some important definitions and operations we will use.

- Containment: $A \subset B$ means if $x \in A$, then $x \in B$. That is, A is a subset of B.
- Equality: $A = B$ means $x \in A$ if and only if $x \in B$. Equivalently, $A = B$ if $A \subset B$ and $B \subset A$.
- Union: the union of A and B is $A \cup B = \{x \in A \text{ or } x \in B\}$. The union of two sets are simply all elements that are in A <u>or</u> B (or in both sets).
- Intersection: the intersection of A and B is $A \cap B = \{x \in A \text{ and } x \in B\}$. The intersection of two sets are simply all elements that are in A <u>and</u> B.
- Complement: the complement of A in S is $A^c = \{x \in S : x \notin A\}$. The complement consists of all elements in S that are not in A.

Here are some useful properties we will use:

- Commutativity: $A \cup B = B \cup A$ and $A \cap B = B \cap A$.
- Associativity: $A \cup (B \cup C) = (A \cup B) \cup C$ and $A \cap (B \cap C) = (A \cap B) \cap C$.
- Distributive laws: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ and $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
- DeMorgan's Laws: $(A \cup B)^c = A^c \cap B^c$ and $(A \cap B)^c = A^c \cup B^c$.

Most of these properties are easy to show and easy to illustrate in figures (in so called Venn Diagrams). For example, we have

$$\begin{aligned} A \cap (B \cup C) &= \{x : x \in A \text{ and } (x \in B \text{ or } x \in C)\} \\ &= \{x : \text{either } (x \in A \text{ and } x \in B) \text{ or } (x \in A \text{ and } x \in C)\} \\ &= \{x : x \in A \cap B \text{ or } x \in A \cap C\} \\ &= (A \cap B) \cup (A \cap C). \end{aligned}$$

Notice that $A \cup (B \cup C)$ is simply the union of three sets, which we usually denote by

$$A \cup B \cup C = \{x \in A \text{ or } x \in B \text{ or } x \in C\}.$$

Similarly, we write the intersection of three sets as

$$A \cap B \cap C = \{x \in A \text{ and } x \in B \text{ and } x \in C\}.$$

Sometimes it will also be useful to take unions or intersections of countably many sets. We denote the union and intersection of $A_1, A_2, \ldots$ by

$$\cup_{i=1}^{\infty} A_i = \{x \in S : x \in A_i \text{ for some } i\}$$

and

$$\cap_{i=1}^{\infty} A_i = \{x \in S : x \in A_i \text{ for all } i\},$$

respectively. For example, let $A_i = (1/i, 2)$ for $i = 1, 2, \ldots$. Then

$$\cap_{i=1}^{\infty} A_i = (1, 2)$$

and

$$\cup_{i=1}^{\infty} A_i = (0, 2).$$

It is of course possible that intersections of two or more sets are empty. We denote the empty set, which does not contain any element, by $\emptyset$. For example, suppose $A = \{1, 2, 3\}$ and $B = \{4, 5, 6\}$. Then $A \cap B = \emptyset$. We end this section with two more definitions.

Definition 1.3. Two events A and B are *disjoint* (or *mutually exclusive*) if $A \cap B = \emptyset$. The events $A_1, A_2, \ldots$ are *pairwise disjoint* (or *pairwise mutually exclusive*) if $A_i \cap A_j = \emptyset$ for all $i \neq j$.

Definition 1.4. If $A_1, A_2, \ldots$ are pairwise disjoint and $\cup_{i=1}^{\infty} A_i = S$, then $A_1, A_2, \ldots$ form a *partition* of S.

For instance, in the die rolling example, $A_1 = \{1, 2\}$, $A_2 = \{3, 4\}$, and $A_3 = \{5, 6\}$ form a partition of $S = \{1, 2, 3, 4, 5, 6\}$.

1.3 Basics of probability theory

Now that we learned about sample spaces and events, we want to assign probabilities to outcomes of experiments. In particular, we want to assign probabilities to events and we will denote the probability of an event A by $P(A)$. We want to define probabilities in a reasonable way. For example, all probabilities will be between 0 and 1. Moreover, since an experiment always results in some outcome, it seems reasonable that $P(S) = 1$. Similarly, it also seems desirable that $P(\emptyset) = 0$. Finally, we might want the probability of a union of disjoint events to be equal to the sum of the individual probabilities. For example, the probability that we roll a 2 or a 6 should be equal to the probability that we roll a 2 plus the probability that we roll a 6. We will formally define a probability function below, but these will be essentially all the properties that we require.

Now more formally, the probability function $P(\cdot)$ assigns numbers to events. It would then seem natural to try to define the function in a way that it assigns a probability to each possible subset of S. However, if S is uncountable, such as $S = [0, \infty)$, it turns out that it is not possible to define a probability function which assigns a value to every subset of S and which satisfies the definition of a probability function below. We therefore have to be a bit careful when we define the domains of probability functions, in particular, the collection of events we assign probabilities to.

To define the domains of probability functions, we need the following definition.

Definition 1.5. A collection of subsets of S is called a *sigma algebra*, denoted by $\mathcal{B}$, if it satisfies the following three properties:

1. $\emptyset \in \mathcal{B}$.

2. If $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$.

3. If $A_1, A_2, \ldots \in \mathcal{B}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$.

As defined below, the domain of a probability function will always be a sigma algebra. If S is countable, then $\mathcal{B}$ could be all subsets of S, including S itself and the empty set. For any given experiment there are typically many different sigma algebras. For example, if $S = \{1, 2, 3, 4, 5, 6\}$, then one possible sigma algebra is $\{\emptyset, S\}$ and another possibility is $\{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, S\}$. Notice that a sigma algebra is a set of sets.

We are now ready to define a probability function.

Definition 1.6. Given a sample space S and an associated sigma algebra $\mathcal{B}$, a *probability function* is a function P with domain $\mathcal{B}$ that satisfies:

1. $P(A) \geq 0$ for all $A \in \mathcal{B}$.

2. $P(S) = 1$.

3. If $A_1, A_2, \ldots \in \mathcal{B}$ are pairwise disjoint, then $P\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.

Example 1.1. Suppose $S = \{1, 2, 3, 4, 5, 6\}$ and let $\mathcal{B}$ be the collection of all possible subsets of S, including S and the empty set. Let $|A|$ be the cardinality of A, which is the number of elements in A. It is then not hard to show that $P(A) = |A|/6$ satisfies the definition of a probability function. The probabilities this function assigns are simply proportional to the number of elements in the set. For these specific sets S and $\mathcal{B}$, we can define many other probability functions satisfying the definition above. As another concrete example suppose $P(A) = 1$ if $6 \in A$ and $P(A) = 0$ if $6 \notin A$.

We said that we might not be able to define a probability function when the domain consists of all subsets of S. For example, when $S = (-\infty, \infty)$, there exists no function defined on the collection of all subsets of S that satisfies the definition of probability function. In this particular case, a possible (and commonly used) sigma algebra is the smallest sigma algebra which contains all open subsets of $\mathbb{R}$. By the definition of a sigma algebra, this collection of sets, contains all open intervals (a, b), closed intervals $[a, b]$, half-open and half-closed intervals $(a, b]$ and $[a, b)$, as well as their finite or countably infinite unions and intersections. While this sigma algebra does not contain all possible subsets of S, it is more than sufficient for all of our econometrics applications and there is no reason for us to worry about these types of technicalities any further. This sigma algebra is often refereed to as the *Borel sigma-algebra*.

There are many other properties that follow immediately from the definition of a probability function, some of which are stated in the following two theorems.

Theorem 1.1. *If P is a probability function and A is any set in $\mathcal{B}$, then*

1. $P(\emptyset) = 0$;

2. $P(A) \leq 1$;

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

Theorem 1.2. *If P is a probability function and A and B are any sets in $\mathcal{B}$, then*

1. $P(B \cap A^c) = P(B) - P(A \cap B)$;

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

3. *If $A \subset B$, then $P(A) \leq P(B)$.*

Part (b) of Theorem 1.2 implies a useful inequality, which is referred to as *Bonferroni's inequality.* Specifically, since $P(A \cup B) \leq 1$, we get

$$P(A \cap B) \geq P(A) + P(B) - 1.$$

Part (b) also implies that the probability of a union is smaller than the sum of the probabilities, namely

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

This result extends to unions of countably many events; see Theorem 1.2.11 of Casella and Berger (2001).

1.4 Conditional probabilities and independence

In this section, we introduce the concept of conditional probabilities. It allows us to compute probabilities under the knowledge that an event occurred. As soon as we know that some event occurred, the relevant sample space changes to the subset of the original sample space that is compatible with the additional information provided by the known event. We then also have to update all probabilities, which are the so-called conditional probability.

Before, we look at the formal definition, let's return to our die rolling example with $S = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{B}$ being the collection of all possible subsets of S, and $P(A) = |A|/6$. Now suppose that we already know that we roll an even number, but we have no additional information. Since we ruled out half of the possible outcomes it seems that, given the additional information, the probability that we roll a 1 (or a 3 or a 5) should be 0, while the probability that we roll a 2 (or a 4 or a 6) should be $\frac{1}{3}$.

We now formalize conditional probabilities.

Definition 1.7. If A and B are events in S, and $P(B) > 0$, then the *conditional probability of A given B*, written $P(A \mid B)$, is

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

Note that the conditional probability is properly defined only when $P(B) > 0$. Moreover, now essentially B becomes the new sample space because $P(B \mid B) = 1$ and $P(A \mid B) = 0$ whenever $A \cap B = \emptyset$.

In the previous die rolling example with $P(A) = |A|/6$, this definition yields conditional probabilities that equal the previous suggestion. In particular, let $B = \{2, 4, 6\}$. Then

$$P(\{1\} \mid \{2,4,6\}) = \frac{P(\{1\} \cap \{2,4,6\})}{P(\{2,4,6\})} = \frac{P(\emptyset)}{P(\{2,4,6\})} = \frac{0}{\frac{1}{2}} = 0$$

and

$$P(\{2\} \mid \{2,4,6\}) = \frac{P(\{2\} \cap \{2,4,6\})}{P(\{2,4,6\})} = \frac{P(\{2\})}{P(\{2,4,6\})} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}.$$

Similarly

$$P(\{3\} \mid \{2,4,6\}) = P(\{5\} \mid \{2,4,6\}) = 0$$

and

$$P(\{4\} \mid \{2,4,6\}) = P(\{6\} \mid \{2,4,6\}) = \frac{1}{3}.$$

Conditional probabilities can be tricky to work with and the results might not be intuitive at first, which is illustrated in the following example.

Example 1.2. Suppose you are applying for a US visa and new regulations require you to get tested for a deadly disease (take your pick). Suppose the disease is very rare and only 1 in $100,000$ people are infected. So coming in, you believe that the probability that you are infected is $1/100,000$. The test you take is quite accurate, but it is not perfect. In particular, let's assume that if you are infected, the test will provide the correct result. However, if you are healthy, the test will give the correct answer only 99.9% of the time and it incorrectly classifies you as infected 0.1% of the time (a so called false positive). False positives are very rare in this example and only 1 in $1,000$ uninfected people get this test result. Now suppose you take the test and the result shows that you are infected. Is it time to panic? In particular, what is the probability that you are actually infected, given that the test results show that you are?

It seems like we need to calculate a conditional probability. Let the sample space be

$$S = \{\{inf, pos\}, \{inf, neg\}, \{non, pos\}, \{non, neg\}\}$$

That is, a person can be either infected (inf) or non-infected (non) and the test result can be either positive (pos), which means that the results show that you are infected, or negative

(neg). Let $\mathcal{B}$ be the collection of all possible subsets of S and let

$$A = \{\{inf, pos\}, \{inf, neg\}\}.$$

Then $P(A)$ is the unconditional probability of being infected and we said that

$$P(A) = \frac{1}{100,000}.$$

Let

$$B = \{\{inf, pos\}, \{non, pos\}\},$$

which is the collection of all outcomes with positive test results. We are now interested in $P(A \mid B)$. In other words, given that you obtained a positive result, what is the probability that you are actually infected. Our definition says that

$$P\left(A \mid B\right) = \frac{P\left(A \cap B\right)}{P\left(B\right)} = \frac{P(\{inf, pos\})}{P(\{inf, pos\}, \{non, pos\})}.$$

To make further progress, notice that by the definition of conditional probabilities, we can write

$$\begin{aligned} P(\{inf, pos\}) &= P(\{inf, pos\} \cap \{\{inf, pos\}, \{inf, neg\}\}) \\ &= P(\{inf, pos\} \mid \{inf, pos\}, \{inf, neg\}) P(\{inf, pos\}, \{inf, neg\}). \end{aligned}$$

Why is this useful? We now have two probabilities on the right-hand side. First, we have

$$P(\{inf, pos\} \mid \{inf, pos\}, \{inf, neg\}),$$

which is the probability of being infected and getting a positive result, given that you are infected. But we said that for infected people, the test is 100% accurate and therefore

$$P(\{inf, pos\} \mid \{inf, pos\}, \{inf, neg\}) = 1.$$

Second, we already said that

$$P(\{inf, pos\}, \{inf, neg\}) = P(A) = \frac{1}{100,000}.$$

Therefore,

$$P(\{inf, pos\}) = \frac{1}{100,000}.$$

We still have to work out what $P(B) = P(\{inf, pos\}, \{non, pos\})$ equals to, which is the probability of getting a positive test result. To do so, notice that

$$\begin{aligned} P(B) &= P((B \cap A) \cup (B \cap A^c)) \\ &= P(B \cap A) + P(B \cap A^c) \\ &= P(B \mid A) P(A) + P(B \mid A^c) P(A^c) \end{aligned}$$

The result

$$P(B) = P(B \mid A)P(A) + P(B \mid A^c)P(A^c)$$

is called the *law of total probability* (see below). Again, why is this useful? It seems we made things more complicated. But we already worked out that $P(A) = 1/100,000$ and therefore $P(A^c) = 1 - 1/100,000$. Moreover, $P(B \mid A)$ is the probability of getting a positive result conditional on being infected, which we said is 1. Finally, $P(B \mid A^c)$ is the probability of getting a positive result conditional on not being infected, which is $1/1,000$. Putting everything together yields

$$P(B) = 1 \cdot \frac{1}{100,000} + \frac{1}{1,000} \cdot \frac{99,999}{100,000} \approx 0.001001.$$

It follows that

$$P(A \mid B) \approx \frac{\frac{1}{100,000}}{0.001001} = 0.0099$$

which means that even though the test result was positive, the conditional probability that you are infected is less than 1%!

To derive this result, we used two important results. First, as already mentioned, we used the *law of total probability* by which we mean the following result.

Theorem 1.3. *Let $A_1, A_2, \ldots$ be a partition of the sample space, and let B be any set. Then*

$$P(B) = \sum_{i=1}^{\infty} P(B \mid A_i) P(A_i).$$

The law of total probability not only applies to countably infinite, but also to finite partitions of the sample space. In the previous example, we used the partition A, A^c. From the law of total probability and the fact that $P(A_i \cap B) = P(B \mid A_i)P(A_i)$, we immediately get *Bayes' rule.*

Theorem 1.4. *Let $A_1, A_2, \ldots$ be a partition of the sample space, and let B be any set. Then for each $i = 1, 2, \ldots$*

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{\sum_{i=1}^{\infty} P(B \mid A_i) P(A_i)}.$$

Another interesting example of (maybe not intuitive) conditional probabilities is the well known Monty Hall Problem (or Ziegenproblem in German). See problems below for details.

When we calculate conditional probabilities, we update the probabilities given the new information. Suppose the new information that B occurred is useless. That is, it does not affect the probability that A occurs and therefore $P(A) = P(A \mid B)$. In this case, we say that A and B are *independent.* But if $P(A) = P(A \mid B)$, then by the definition of conditional probabilities,

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

which implies that $P(A \cap B) = P(A)P(B)$. We can therefore define independence of two events as follows.

Definition 1.8. Two events, A and B, are *independent* if

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

Sometimes people also use the term *statistically independent* in this case. Here are some other useful properties which are easy to show.

Theorem 1.5. *If A and B are independent then the following pairs are also independent:*

1. *A and B^c*
2. *A^c and B*
3. *A^c and B^c*

Next, we generalize the definition to any arbitrary finite collection of events.

Definition 1.9. The events $A_1, A_2, ..., A_n$ are *mutually independent* if for any subcollection $A_{i_1}, \ldots, A_{i_k}$

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P\left(A_{i_j}\right)$$

It is important to note that the definition of mutual independence requires checking the independence condition for all possible (non-singleton) subcollections. For example, for the collection A, B, C, we need to check:

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

Therefore, independence of each pair (called *pairwise independence*) is a necessary, but not sufficient condition for mutual independence.

Problems

Problem 1.1. Illustrate graphically that $B = (B \cap A) \cup (B \cap A^c)$. Also prove this equality formally.

Problem 1.2. Suppose $S = \{a, b, c, d, e, f, g, h\}$. Which of the following collections of sets is a sigma-algebra:

1. $\{\emptyset, S\}$
2. $\{\emptyset, \{a, b, c, d\}, \{e, f, g, h\}\}$.
3. $\{\emptyset, \{a, c\}, \{b, d\}, \{a, b, c, d\}, \{e, f, g, h\}, S\}$.

Problem 1.3. Prove Theorem 1.1.

Problem 1.4. Prove Theorem 1.2.

Problem 1.5. During the lecture we briefly discussed the Monte Hall Problem, which we now analyze formally. The setup is as follows:

Suppose you are on a game show, and you are given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?"

To analyze this question notice that the host always opens one door with a goat behind it regardless of your initial choice. You may assume that if you picked the door with the car initially, the host picks one of the two doors at random. Now

(a) calculate the probability that the car is behind the door you initially picked, given what's behind the door the host opened and

(b) calculate the probability that the car is behind the other door, given what's behind the door the host opened.

Define all objects, such as the sample space, etc. carefully. Should you switch?

Problem 1.6. Let $S = \{1, 2, 3, 4\}$ and let $\mathcal{B}$ consist of all subsets of S. Let $P(\{i\}) = 1/4$ for all $i = 1, 2, 3, 4$. Now define $A = \{1, 2\}$, $B = \{1, 3\}$, $C = \{2, 3\}$. Are A,B,C pairwise independent? Are they mutually independent?

Problem 1.7. Suppose that a pair of events A and B are mutually exclusive (i.e. $A \cap B = \emptyset$) and that $P(A) > 0$ and $P(B) > 0$. Prove that A and B are not independent.

Problem 1.8. Another important concept is *conditional independence* defined as follows.

Definition 1.10. Let A_1, A_2, and B be three events such that $P(B) > 0$. The events A_1 and A_2, are *(statistically) independent conditional on* B if

$$P\left(A_1 \cap A_2 \mid B\right) = P\left(A_1 \mid B\right) P\left(A_2 \mid B\right).$$

The question we want to analyze is whether independence implies conditional independence and vice versa.

(a) Consider the experiment of tossing two coins independently. Each coin has the possible outcomes H or T (heads or tails) with equal probability. Then we can write

$$S = \{\{H,H\},\{H,T\},\{T,H\},\{T,T\}\}.$$

Let

$$A_1 = \{\text{first toss is heads}\} = \{\{H,H\} \cup \{H,T\}\}$$

and

$$A_2 = \{\text{second toss is heads}\} = \{\{H,H\} \cup \{T,H\}\}.$$

Also let

$$B = \{\text{the two tosses are different}\} = \{\{H,T\} \cup \{T,H\}\}.$$

Show that A_1 and A_2 are statistically independent, but A_1 and A_2 are not statistically independent conditional on B.

(b) The next experiment will be a bit more complicated. Again we have two coins, now denoted by C_1 and C_2 and the possible outcomes of a coin flip are H or T (heads or tails). For coin 1 (C_1), the probability of heads is 0.9 and the probability of tails is 0.1. On the other hand, for coin 2 (C_2), the probability of heads is 0.1 and the probability of tails is 0.9. We now first select a coin at random. Each coin has a probability of 0.5 of being selected. We then flip the selected coin twice and those two coin flips are independent of each other. Therefore, the sample space is

$$\begin{aligned} S = \{&\{C_1,H,H\},\{C_1,H,T\},\{C_1,T,H\},\{C_1,T,T\},\\ &\{C_2,H,H\},\{C_2,H,T\},\{C_2,T,H\},\{C_2,T,T\}\} \end{aligned}$$

That is, we can pick either coin and then get either H or T on each of the two flips. Now let

$$\begin{aligned} A_1 &= \{\text{first toss is } H\} \\ &= \{\{C_1, H, H\}, \{C_1, H, T\}, \{C_2, H, H\}, \{C_2, H, T\}\} \end{aligned}$$

and

$$\begin{aligned} A_2 &= \{\text{second toss is } H\} \\ &= \{\{C_1, H, H\}, \{C_1, T, H\}, \{C_2, H, H\}, \{C_2, T, H\}\}. \end{aligned}$$

Also let

$$\begin{aligned} B &= \{\text{select coint 1}\} \\ &= \{\{C_1, H, H\}, \{C_1, H, T\}, \{C_1, T, H\}, \{C_1, T, T\}\}. \end{aligned}$$

Show that A_1 and A_2 are not statistically independent, but A_1 and A_2 are statistically independent conditional on B.

Problem 1.9 (Computational problem)**.** Use R and the functions *union* and *intersect* to find $A_1 \cup A_2 \cup A_3$ and $A_1 \cap A_2 \cap A_3$ where

$$\begin{aligned} A_1 &= \{13, 5, 9, 17, 268, 183\} \\ A_2 &= \{36, 5, 9, 7, 268\} \\ A_3 &= \{5, 268, 81\} \end{aligned}$$

Also use these functions to illustrate the distributive laws, i.e. that

$$A_1 \cap (A_2 \cup A_3) = (A_1 \cap A_2) \cup (A_1 \cap A_3)$$

and

$$A_1 \cup (A_2 \cap A_3) = (A_1 \cup A_2) \cap (A_1 \cup A_3).$$

Problem 1.10 (Computational problem)**.** This problem asks you to numerically verify the results of Problem 1.5. To do so, perform the following steps:

1. Use the command *sample(1:3,1)* to pick one of three doors that contain the car.
2. Use the command *sample(1:3,1)* to pick one of three doors that is initially picked by the contestant.

3. Depending on the outcomes of step 1 and 2, construct the set of doors that have not been picked by the contestant and that do not contain the car. Select a door from this set at random, which is the door that will be opened.

4. Implement the strategy that the contestant does not switch.

5. Implement the strategy that the contestant switches.

6. Repeat steps 1-4 at least 1000 times and calculate the fractions of times that the contestant wins with either strategy.

References

Casella, G. and R. Berger (2001). *Statistical Inference*. Duxbury Resource Center.