# A Few Basics of Probability

Philosophy 57

Spring, 2004

## 1 Introduction

This handout distinguishes between inductive and deductive logic, and then introduces probability, a concept essential to the study of inductive logic.

After reading this handout, you should know what the inductive/deductive distinction is, basically what a probability distribution is, and how to do a few simple probability problems.

## 2 Deductive vs Inductive Reasoning

The reasoning that we've learned so far has been deductive. Deductive reasoning is all or nothing. Consider a valid deductive argument:

| |
|---|
| All men are mortal. |
| Socrates is a man. |
| Socrates is mortal. |

If the premises are true, the conclusion isn't just more likely, it is necessarily true. Valid deductive arguments are great when they exist, but sometimes an argument can have force without being valid. For instance:

| |
|---|
| The sun has risen every day for the past 1000 years. |
| The sun will rise tomorrow. |

This is not a valid argument. It is possible to imagine that the sun rises every day for 1000 years and then doesn't rise the next day. However, it is a pretty good argument, in that the premise does seem to make the conclusion more likely. It seems to be good evidence for the conclusion.

Much of the reasoning you do on a daily basis is inductive. Perhaps you study for a test in order to do better; you may conclude that class will be at noon because it has always been at noon before and that's what it says on the syllabus; when you evaluate a textbook's veracity you take into account the author's credentials. All of these reasoning is inductive—it is not certainly valid and the conclusions you reach may be false.

## 2.1 Monotonicity

One notable difference between deductive and inductive logic is that adding extra premises can make a good inductive argument into a bad one. For instance:

| The sun has risen every day for the past 1000 years. |
|---|
| The sun just blew up an hour ago. |
| The sun will rise tomorrow. |

The addition of the second premise has turned a reasonably good argument into a bad one.

In deductive logic, addition of a premise can never turn a valid argument into an invalid one. For example, this argument:

| All men are mortal. |
|---|
| Socrates is a man. |
| Socrates is widely thought to be immortal. |
| Socrates is mortal. |

is still valid deductively.

## 2.2 Sample problems

There is no homework due on probability, but to help you learn the material there are some sample problems interspersed through this handout.

1. Make up two more good probabilisitic arguments that are not deductive arguments.

2. Take the two good probabilistic arguments from the previous question and add a premise to each, making them into bad arguments.

3. Suppose in the "Socrates is a man; all men are mortal; thus Socrates is mortal" argument we added the premise "Socrates is immortal." Would the argument still be logically valid? Why or why not? If not, doesn't this contradict the statement above that adding a premise to a deductive argument can never make it deductively invalid?

# 3 Probability measures

All common approaches to inductive logic make use of probabilities. In this section we will introduce the notion of a probability measure, then the axioms of probability, and finally will show how a few basic facts about probability can be deduced from them.

## 3.1 Intuitive probability

Intuitively, a probability measure is a function P that assigns to each sentence a number between 0 and 1 inclusive depending on how likely the sentence is to be true. Here are a few examples:

| Statement $S$ | possible value of $P(S)$ |
|---|---|
| Logic is fun and logic is not fun. | 0 |
| Nader will be next president. | $0.00\ldots0001$ |
| My lottery ticket will win. | 0.0000001 |
| I'll get murdered this year. | 0.0001 |
| I'll get killed in car accident this year. | 0.001 |
| On next die toss, will get 6. | 0.17 |
| Kerry will be next president. | 0.4 |
| Next coin flip will land heads. | 0.5 |
| Bush will be next president | 0.6 |
| I'll live to see 70th birthday | 0.75 |
| Next die toss will not land 6. | 0.83 |
| On 7 coin flips, will get at least one heads. | 0.99 |
| Sun will rise tomorrow. | $0.999\ldots99$ |
| Logic is fun or logic is not fun. | 1 |

(Disclaimer: All values are approximate and not necessarily calculated from mortality statistics or valid political data.)

The idea is that sentences which cannot be true (like logical falsehoods of the form $A \wedge \neg A$) get probability 0. Sentences which must be true (like tautologies of the form $A \vee \neg A$) get probability 1. Sentences that may be true or may not with equal probability (like a coin landing heads or tails) get probability 0.5. Sentences of intermediate likelihood get assigned an intermediate number.

Note that there is much philosophical discussion on what exactly probabilities are. We won't be discussing this in this class, but see the Hájek article mentioned in the further readings section for more information.

## 3.2 The axioms of probability

More formally a probability measure P is a function from sentences to $\mathbb{R}$ (the real numbers) such that:

1. $\mathrm{P}(S) \geq 0$ for all sentences $S$.
2. If $S$ is a logical truth, then $\mathrm{P}(S) = 1$.
3. If $A$ and $B$ are mutually exclusive (i.e. $A \wedge B$ is logically impossible), then $\mathrm{P}(A \vee B) = \mathrm{P}(A) + \mathrm{P}(B)$.

Each of these axioms should make some sense based on the intuitive notion of probability: The first axiom says that the probability must be greater than 0. Something that can't happen gets probability 0, so it doesn't make sense to give something probability less than that.

The second axiom says that logical truths should get probability 1. Again this makes sense because logical truths are necessarily true, and are known with certainty. Probability 1 is the highest probability that any statements are supposed to get (see below) so it makes sense to give logical truths probability 1.

The third axiom is a bit more confusing than the others but still makes intuitive sense. If two sentences cannot be true together, then the probability that one or the other happens is the sum of the probabilities of each happening. For instance, a die may land on 1 or 2 with probability $\frac{1}{6}$ each, and a die cannot land on 1 or 2 simultaneously. Thus the probability that the die lands on 1 or 2 is $\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

## 3.3 Simple Derivations from the Axioms

Although the axioms are quite simple, we can still deduce a number of useful facts from them. The following subsections each prove some elementary result.

### 3.3.1 Probability of a negation

It is simple to calculate the probability of a negation if you know the original probability: $\mathrm{P}(\neg A) = 1 - \mathrm{P}(A)$

> **Proof:** $A$ and $\neg A$ are mutually exclusive, so by axiom 3, $\mathrm{P}(A) + \mathrm{P}(\neg A) = \mathrm{P}(A \vee \neg A)$. But since $A \vee \neg A$ is a logical truth (in fact, a tautology), $\mathrm{P}(A \vee \neg A) = 1$ by axiom 2. Then $\mathrm{P}(A) + \mathrm{P}(\neg A) = 1$, so $\mathrm{P}(\neg A) = 1 - \mathrm{P}(A)$.

For example, if the probability that you graduate this year is 95%, then the probability that you don't graduate this year is 5%.

### 3.3.2 Probability of contradiction

If a sentence $A$ is a logical contradiction (this of course includes tautological or first order contradictions) then $\mathrm{P}(A) = 0$. Thus contradictory sentences have no chance of being true.

> **Proof:** If $A$ is a contradiction, then $\neg A$ must be a logical truth. Then $\mathrm{P}(\neg A) = 1$ by axiom 2. This along with the fact that $\mathrm{P}(A) + \mathrm{P}(\neg A) = 1$ (see 3.3.1) implies $\mathrm{P}(A) = 0$ as desired.

### 3.3.3 All probabilities between 0 and 1

For any sentence $A$, $0 \leq \mathrm{P}(A) \leq 1$.

> **Proof:** By axiom 1 we know that $\mathrm{P}(A) \geq 0$. We just need to show that $\mathrm{P}(A) \leq 1$. So note that $\mathrm{P}(A) = 1 - \mathrm{P}(\neg A)$ (see 3.3.1 again). But by axiom 1 we also have $0 \leq \mathrm{P}(\neg A)$. Adding this inequality to the previous equation we get $\mathrm{P}(A) \leq 1$ as desired.

In order for probabilities to make sense this result must hold, but we haven't proven it from the axioms until now.

### 3.3.4 Probability of logically equivalent sentences

Suppose $A$ and $B$ are logically equivalent. Then $\mathrm{P}(A) = \mathrm{P}(B)$.

> **Proof:** By the deductive logic we've learned so far, we know that $A \vee \neg B$ is a logical truth, and that $A$ and $\neg B$ are mutually exclusive. Thus
>
> $$\begin{aligned} \mathrm{P}(A) &= \mathrm{P}(A \vee \neg B) - \mathrm{P}(\neg B) && \text{by axiom 3} \\ &= 1 - \mathrm{P}(\neg B) && \text{since } \mathrm{P}(A \vee \neg B) = 1 \text{ by axiom 2} \\ &= 1 - (1 - \mathrm{P}(B)) && \text{since } \mathrm{P}(\neg B) = 1 - \mathrm{P}(B) \\ &= \mathrm{P(B)} \end{aligned}$$

An intuitive gloss on this result is that if two sentences say the same thing, then they will either both happen or both not happen, so they are equally likely to be true.

### 3.3.5 Probability of conjunction

Unlike the negation case, the probability of a conjunction doesn't depend only on the probability of the subsentences. However, we do have both that $\mathrm{P}(A \wedge B) \leq \mathrm{P}(A)$ and that $\mathrm{P}(A \wedge B) \leq \mathrm{P}(B)$.

> **Proof:** $\mathrm{P}(A) = \mathrm{P}((A \wedge \neg B) \vee (A \wedge B))$ by our equivalence result (3.3.4). Then $\mathrm{P}(A) = \mathrm{P}(A \wedge \neg B) + \mathrm{P}(A \wedge B)$ by axiom 3. Since $\mathrm{P}(A \wedge \neg B) \geq 0$ by axiom 1, $\mathrm{P}(A) \geq \mathrm{P}(A \wedge B)$.
>
> By doing the same starting with $B$ instead of $A$, we get that $\mathrm{P}(B) \geq \mathrm{P}(A \wedge B)$.

Note that this result is a special case of problem 2 below.

## 3.4 Sample problems

These problems involve working with the axioms at a low level. For a few extra problems you can try deriving the results above without looking at the proofs given to you.

1. Prove that $\mathrm{P}(A \vee B) \geq \mathrm{P}(A)$.
2. Prove that $\mathrm{P}(B) \leq \mathrm{P}(A \rightarrow B)$.
3. Show that if $\mathrm{P}(A) = 1$, then $\mathrm{P}(A \wedge B) = \mathrm{P}(B)$

4. *Show that if $\mathrm{P}(A \leftrightarrow B) = 1$, then $\mathrm{P}(A) = \mathrm{P}(B)$.

5. *Prove that if $A$ entails $B$, then $\mathrm{P}(A) \leq \mathrm{P}(B)$.

   **New example solution:** Note that $A$ is logically equivalent to $(A \wedge B) \vee (A \wedge \neg B)$, and that $A \wedge \neg B$ is logically impossible if $A$ entails $B$. Then:

$$
\begin{array}{rcll}
\mathrm{P}(A) & = & \mathrm{P}((A \wedge B) \vee (A \wedge \neg B)) & \text{by (3.3.4)} \\
& = & \mathrm{P}(A \wedge B) + \mathrm{P}(A \wedge \neg B) & \text{by axiom 3} \\
& = & \mathrm{P}(A \wedge B) & \text{by (3.3.2)} \\
& \leq & \mathrm{P}(B) & \text{by (3.3.5)}
\end{array}
$$

# 4 Independence and Conditional Probability

## 4.1 Conditional probability

Here we introduce some new notation. Besides normal probabilities, like $\mathrm{P}(A)$, we can also talk about conditional probabilities, like $\mathrm{P}(A|B)$. The symbols $\mathrm{P}(A|B)$ are read "the probability of $A$ given $B$" and the quantity intuitively refers to the probability that you would give to $A$ if you knew that $B$ were true. Here's the formal definition:

**Definition:**

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \wedge B)}{\mathrm{P}(B)} \text{ where } \mathrm{P}(B) \neq 0$$

## 4.2 Independent sentences

The notion of two sentences being *probabilistically independent* is very important. The idea behind independence is that two independent sentences don't depend on one another, so learning about one doesn't tell you anything about the probability of the other.

**Definition:** Sentences $A$ and $B$ are *independent* iff $\mathrm{P}(A) = \mathrm{P}(A|B)$.

By the definition of conditional probability, two sentences $A$ and $B$ (with $\mathrm{P}(B) > 0$) are independent just in case $\mathrm{P}(A \wedge B) = \mathrm{P}(A)\mathrm{P}(B)$.

The definition of independence for three or more statements won't be stated formally, but similar multiplication holds. For instance, if $A$, $B$, and $C$ are all independent, then $\mathrm{P}(B \wedge C) = \mathrm{P}(B)\mathrm{P}(C)$ and $\mathrm{P}(A \wedge B \wedge C) = \mathrm{P}(A)\mathrm{P}(B)\mathrm{P}(C)$.

### 4.2.1 Independence of negations

Suppose $A$ and $B$ are independent, and $0 < \mathrm{P}(B) < 1$. Then so are $A$ and $\neg B$.

> **Proof:** Since $A$ and $(A \wedge B) \vee (A \wedge \neg B)$ are logically equivalent, we have $\mathrm{P}(A) = \mathrm{P}(A \wedge B) + \mathrm{P}(A \wedge \neg B)$ by axiom 3 and our result from 3.3.4. Now the rest follow smoothly from this idea:
>
> $$\begin{aligned} \mathrm{P}(A \wedge \neg B) &= \mathrm{P}(A) - \mathrm{P}(A \wedge B) && \text{result immediately above} \\ &= \mathrm{P}(A) - \mathrm{P}(A)\mathrm{P}(B) && \text{because } A \text{ and } B \text{ are independent} \\ &= \mathrm{P}(A)(1 - \mathrm{P}(B)) && \text{factoring out } \mathrm{P}(A). \\ &= \mathrm{P}(A)\mathrm{P}(\neg B) && \text{by negation result from 3.3.1} \end{aligned}$$

## 4.3 Example: coin flips

The standard example of independent sentences involves separate coin flips or die rolls. Unless stated otherwise, you can always assume that the results of coin flips and die rolls are independent. You can also usually assume that coins and die are "fair", meaning that a coin has a 50% chance of coming up heads, and that each side of an n-sided die has $\frac{1}{n}$ probability of landing up.

Interestingly, the unbiasedness assumption about coins may not be true. See `http://www.sciencenews.org/articles/20040228/fob2.asp` for a description of work by a Stanford professor that says a coin is biased about 51-49.

### 4.3.1 Probability of repeated flips

Suppose a coin is flipped 3 times. Let $T_1$ and $H_1$ mean that the first flip landed tails and heads repectively, and let $T_2$, $T_3$, $H_2$, $H_3$ mean the same with respect to the second and third flips.

Then $\mathrm{P}(T_1 \wedge T_2 \wedge T_3) = \mathrm{P}(T_1)\mathrm{P}(T_2)\mathrm{P}(T_3) = \frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{8}$. Since we could have done this to any other result statement, like $H_1 \wedge T_2 \wedge H_3$, all these sequences have the same probability. There are eight different ways the coin could land when flipped three times, and all have probability $\frac{1}{8}$.

Similar results hold when we look longer sequences, say of 5 or 100 tosses.

### 4.3.2 Conditional coin probability example

Suppose that a coin is flipped 3 times. The probability that the last two tosses are heads, given that the first toss is a tail, is 25%. In symbols, $\mathrm{P}(H_2 \wedge H_3 | T_1) = \frac{1}{4}$.

**Proof:**

$$P(H_2 \wedge H_3 | T_1) = \frac{P(T_1 \wedge H_2 \wedge H_3)}{P(T_1)} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}$$

## 4.4 Probability by counting

Often when many events are equally probable it is easier to think about counting the possibilities than calculating the probabilities explicitly.

For instance, suppose a die is rolled, and we want to know the probability of the die landing 1, 2, or 3. We could calculate $P(D_1 \vee D_2 \vee D_3) = P(D_1) + P(D_2) + P(D_3) = \frac{1}{3} + \frac{1}{3} + \frac{1}{3} = \frac{1}{2}$. On the other hand, we could get the answer by observing that there are 6 equally probable outcomes, and the statement is true for 3 of them. Then the statement has probability $\frac{3}{6} = \frac{1}{2}$.

### 4.4.1 Adding up die roll totals

Suppose two (six-sided) dice are rolled. Then the probability that they add up to 7 is $\frac{1}{6}$.

> **Proof:** There are 36 possibilities total—the first die could land in any of 6 positions, and so can the second. Under six of these possibilities the dice add up to 7: The first could land 1 and the second 6, or the first 2 and the second 5, or 3 and 4, or 4 and 3, or 5 and 2, or 6 and 1. Thus the probability is $\frac{6}{36} = \frac{1}{6}$.

## 4.5 Reasoning about conditional probability

Often probabilistic reasoning and inference are useful in everyday life. We haven't really learned enough to apply them to many situations, but we know enough to analyze some simple cases.

### 4.5.1 Taxing example

Suppose there is a 55% chance that Bush gets elected, and a 45% chance that Kerry gets elected (and one of the two will definitely get elected). Now suppose that the probability of taxes getting raised given that Bush gets elected is 0.75, and the probability of taxes getting raised if Kerry gets elected is 0.95.

Then the probability that taxes get raised is 0.84.

**Proof:** Let $B$ mean that Bush gets elected, and $\neg B$ that Kerry gets elected. Let $T$ mean that taxes get raised.

Then $\mathrm{P}(T) = \mathrm{P}(T \wedge B) + \mathrm{P}(T \wedge \neg B)$ by axiom 3 and the result from 3.3.4. Since $\mathrm{P}(T \wedge B) = \mathrm{P}(T|B)\mathrm{P}(B)$ and $\mathrm{P}(T \wedge \neg B) = \mathrm{P}(T|\neg B)\mathrm{P}(\neg B)$ (see the first sample problem below), $\mathrm{P}(T) = \mathrm{P}(T|B)\mathrm{P}(B) + \mathrm{P}(T|\neg B)\mathrm{P}(\neg B) = 0.75 \cdot 0.55 + 0.95 \cdot 0.45 = 0.84$

## 4.6 Sample Problems:

Here are some sample problems involving conditional probability and independence.

1. Prove that $\mathrm{P}(A) = \mathrm{P}(A|B)\mathrm{P}(B) + \mathrm{P}(A|\neg B)\mathrm{P}(\neg B)$ for $0 < \mathrm{P}(B) < 1$.

2. A fair coin is flipped 50 times in a row. What is the probability that the last flip lands heads given that the 49 previous flips all landed heads?

3. A coin is flipped 3 times. What is the probability that exactly two of the tosses land heads?

4. A coin is flipped 3 times. What is the probability that they all land heads, given that at least two of the flips land heads?

   **New example solution:** The probability that all land heads, $\mathrm{P}(AllHeads) = \mathrm{P}(H_1 \wedge H_2 \wedge H_3)$ is $\mathrm{P}(H_1)\mathrm{P}(H_2)\mathrm{P}(H_3) = \frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{8}$ because the coin is fair and flips are independent. In fact, any particular sequence has the same probability. The probability that at least two of the coins land heads, $\mathrm{P}(TwoHeads)$ is the the probability of $(H_1 \wedge H_2 \wedge T_3) \vee (H_1 \wedge T_2 \wedge H_3) \vee (T_1 \wedge H_2 \wedge H_3) \vee (H_1 \wedge H_2 \wedge H_3)$, which has probability $\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{1}{2}$.

   Thus the requested probability is $\frac{\mathrm{P}(AllHeads \wedge TwoHeads)}{\mathrm{P}(TwoHeads)} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}$.

5. Two (six-sided) dice are rolled, what is the probability that the results add up to 6?

6. Three dice are rolled, what is the probability that they add up to 12, given that the first roll lands 5?

7. Let $A$ mean that Sally majors in math, let $B$ mean that she majors in philosophy, and $C$ mean that she takes another logic course. Assume that she doesn't double major, so that $A$ and $B$ are mutually exclusive.

Also assume $P(A) = 0.5$, $P(B) = 0.25$. If $P(C|A) = 0.5$, $P(C|B) = 0.9$, and $P(C|\neg A \wedge \neg B) = 0.2$, what is the probability that Sally takes another logic class?

8. *A deck is shuffled randomly (meaning that every ordering is equally likely). What is the probability that the first two cards drawn from the deck will be aces?

9. **A deck is randomized, and a five card hand is drawn from it. What is the probability the the hand will contain two pairs of face cards (and no three-of-a-kind)? (E.g. two jacks, two kings, and an 8.)

# 5 Further Readings

Hurley's textbook on logic has a chapter on induction which covers probability. Possibly useful because it broaches the topic from the same direction.

- Hurley, Patrick J., *A Concise Introduction to Logic*, 2003.

The following is an introductory book in probability, often used in beginning statistics classes. Requires a bit of math to start. Also note that they introduce probabilities by defining them over sets instead of propositions.

- Ross, Sheldon, *A First Course in Probability*, 1994.

This text covers much of what Ross's does and continues into statistics. Requires a bit more math than Ross's textbook. Only read if you get really into this...

- Casella, George, and Berger, Roger, *Statistical Inference*, 2001.

For more on what probabilities really are on a philosophical level (e.g. subjectivists vs frequentists), this online article is a good introduction:

- Hájek, Alan, "Interpretations of Probability", *The Stanford Encyclopedia of Philosophy*, `http://plato.stanford.edu/archives/sum2003/entries/probability-interpret/`

Finally, this book is pretty long but introduces probability from scratch (see chapter 2) and discusses many statistical and philosophical topics related to probability. Probabilities initially defined over propositions, the way we do. Defends reasoning from Bayesian perspective.

- Howson, Colin, and Urbach, Peter, *Scientific Reasoning: The Bayesian Approach*, 1993.