10.4: Probabilistic Reasoning: Rules of Probability

- Inductive logic involves the notion of strength in its definition:

  **Inductive logic** is the part of logic that is concerned with the study of methods of evaluating arguments for strength or weakness.

- And strength in turn was characterized in the last lecture in terms of probability:

  A strong argument is one in which it is probable (but not necessary) that if the premises are true, then the conclusion is true.

- A sound foundation for inductive logic therefore requires a rigorous theoretical understanding of the notion of probability.

- And, in fact, probability theory has become an extremely advanced branch of mathematics.

- In these final lectures we will study the basic laws, or rules, of the probability calculus, which form the basis of probability theory.

**Background to Probability**

- There is serious philosophical disagreement about the precise nature of probability
  - Is it something “objective”, something to be discovered out in the world?
  - Is it just a measure of one’s own subjective feelings, a measure of the strength of one’s belief that something will occur?
• But there is widespread agreement about (a) the probabilities of certain logically distinctive propositions and (b) how the probability of a compound statement is determined by the probabilities of its component statements.

• The probability calculus consists of the basic rules concerning (a) and (b).

• The rules concerning (b) are analogous to the rules of the truth table method of Ch. 7.
  • A truth table does not tell us the truth value of simple statements like F and G.
  • But it does tell us how the truth value of a compound statement like \((F \lor G)\) is determined, given the truth values of F and G.
  • Likewise, the probability calculus does not tell us the probability of simple statements like F and G.
  • But it does tell us how the probability of a compound statement like \((F \lor G)\) — written \(P(F \lor G)\) — is determined, given the probabilities of F and G — written \(P(F)\) and \(P(G)\).

• We will be using the language of statement logic except that the letter “P” will be reserved for the probability operator.
  • Statement letters: A, B, C, …, O, Q, R, …, Z (though I’ll use some alternatives below)
  • Logical operators: \(~, \cdot, \lor, \rightarrow, \leftrightarrow, \) and P
  • And, as before, we will use lowercase italic letters \(p, q, r, \ldots\) as metavariables that stand for arbitrary statements.

The Rules of Probability

• Probability values are expressed as numbers from 0 to 1.
  • 0 is the lowest degree of probability, 1 the highest.
• It is customary to assign a probability of 1 to the tautologies of statement logic, i.e., those that are true in every row of a truth table.

• This is reasonable because tautologies must be true; there is not the smallest probability that a tautology could be false.

• This is in fact the first rule of the probability calculus:

  **Rule 1:** If a statement $p$ is a tautology, then $P(p) = 1$.

• Likewise, a probability of 0 is assigned to contradictions, i.e., those that are false in every row of a truth table.

  **Rule 2:** If a statement $p$ is a contradiction, then $P(p) = 0$.

**Examples**

• By Rule 1, $P(A \lor \neg A) = P(B \rightarrow (A \rightarrow B)) = 1$.

• By Rule 2, $P(A \land \neg A) = P(\neg(B \leftrightarrow B)) = 0$.

**MUTUAL EXCLUSIVITY**

• Consider the statements:
  
  (a) Hillary Clinton will win the US presidency in 2016.

  (b) Jeb Bush will win the US presidency in 2016.

• These statements both have a probability between 0 and 1.

• However they cannot both be true; they are mutually exclusive.

Two statements are **mutually exclusive** if they cannot both be true.
EXHAUSTIVENESS

• Consider the statements:
  (a) W. V. Quine was born before 1900.
  (b) W. V. Quine was born after 1900.
  (c) W. V. Quine was not born before or after 1900.
• Not only are they mutually exclusive, one of them must be true; together they exhaust the possibilities. Hence:

  Statements $p, q, r, \ldots$ are jointly exhaustive if at least one of them must be true.

• Now suppose $p$ and $q$ are mutually exclusive.
  • Let $T =$ The die will turn up 3
  • Let $S =$ The die will turn up 6
  • There is a $1$ in $6$ ($1/6$) chance that the die will land on any given side.
  • So $P(T) = P(S) = 1/6$
  • Hence, since $T$ and $S$ are mutually exclusive, there is a $2$ in $6$ chance that either $T$ or $S$, that is:
    • $P(T \lor S) = P(T) + P(S) = 2/6 = 1/3$.
  • This illustrates the restricted disjunction rule:

    **Rule 3:** If $p$ and $q$ are mutually exclusive, then
    $$P(p \lor q) = P(p) + P(q).$$
Examples

• Suppose we want to draw one card from a well-shuffled deck of 52.

• Since drawing an Ace of Clubs (A♣) and drawing an Ace of Diamonds (A♦) are mutually exclusive, we have:

\[ P(A♣ \lor A♦) = P(A♣) + P(A♦) = \frac{1}{52} + \frac{1}{52} = \frac{2}{52} = \frac{1}{26} \]

• What is the probability of drawing a Queen (of any suit)?

\[ P(Q♣ \lor Q♦ \lor Q♥ \lor Q♠) = P(Q♣) + P(Q♦) + P(Q♥) + P(Q♠) = \frac{1}{52} + \frac{1}{52} + \frac{1}{52} + \frac{1}{52} = \frac{4}{52} = \frac{1}{13} \]

THE PROBABILITY OF NEGATIONS

• The restricted disjunction (RD) rule enables us to calculate the probability of a negation, \(P(\neg p)\), from the probability of the statement negated, \(P(p)\).

• Consider any statement \(p\).

• \(p\) and its negation \(\neg p\) are mutually exclusive.

• Hence, by the RD rule

\[ P(p \lor \neg p) = P(p) + P(\neg p) \]

• But by Rule 1, the rule for tautologies, we also know that

\[ P(p \lor \neg p) = 1 \]

• Putting these two together, we have

\[ P(p) + P(\neg p) = 1 \]

• And, subtracting \(P(p)\) from both sides, we have our fourth rule, the negation rule:

\[ \text{Rule 4: } P(\neg p) = 1 - P(p) \]
**Example 1**

- Suppose we know that the probability, $P(F)$, of throwing a 4 on the next throw of a die is 1 in 6, so $P(F) = 1/6$.

- Then the negation rule enables us to calculate the probability $P(\neg F)$ that a 4 will *not* turn up on the next throw:
  
  $$P(\neg F) = 1 - P(F) = 1 - 1/6 = 6/6 - 1/6 = 5/6.$$  

**Example 2**

- Since there are 13 cards in each suit, the probability, $P(S)$, that we will draw a spade from a well-shuffled deck is 13/52.

- Hence, the probability $P(\neg S)$ that we will *not* draw a spade is:
  

**THE GENERAL DISJUNCTION RULE**

- Obviously, not every pair of statements is mutually exclusive.

  - In many cases $p$ and $q$ can both be true.

  - E.g., Let $K = \text{You draw a King}$ and $C = \text{You draw a Club}$. $K$ and $C$ are not mutually exclusive because of the King of Clubs.

- So we need a more *general* disjunction rule for calculating probabilities $P(p \lor q)$ when $p$ and $q$ are not mutually exclusive.

  - Consider the probability $P(K \lor C)$ of drawing a King or a Club.

    - The sum $P(K) + P(C) = 4/52 + 13/52 = 17/52$ is too high, since we are in effect counting $K\spadesuit$ twice — once as a King and once as a Club.

    - So we need *subtract* the probability of drawing $K\spadesuit$, i.e., the probability $P(K\spadesuit C)$ of drawing *both* a King and a club:

      $$P(K \lor C) = P(K) + P(C) - P(K \spadesuit C) = 4/52 + 13/52 - 1/52 = 16/52 = 4/13$$
• This illustrates the general disjunction rule:

**Rule 5:** \[ P(p \lor q) = P(p) + P(q) - P(p \land q) \]

• Note that, when \( p \) and \( q \) are mutually exclusive, \( P(p \land q) = 0 \).

• Hence, we can derive Rule 3 from Rule 5.

• E.g., since it is impossible to draw both a Club (♣) and a Diamond (♦) on a single draw, the probability of doing so, \( P(♣ \land ♦) \), is 0. Hence:

\[
P(♣ \lor ♦) = P(♣) + P(♦) - P(♣ \land ♦) = \frac{1}{4} + \frac{1}{4} - 0 = \frac{2}{4} = \frac{1}{2}.
\]

**Example**

• What is the probability \( P(R \lor E) \) of drawing a red card (R) or an 8 (E)?

• \( R = \) You draw either a Heart or a Diamond, (♥ v ♦).

• So \( P(R) = P(♥ \lor ♦) = P(♥) + P(♦) \) (since ♥ and ♦ are mutually exclusive) = \( \frac{13}{52} + \frac{13}{52} = \frac{26}{52} (= \frac{1}{2}) \).

• \( E = \) You draw either 8♣, 8♦, 8♥, or 8♠

• So \( P(E) = P(8♣ \lor 8♦ \lor 8♥ \lor 8♠) = P(8♣) + P(8♦) + P(8♥) + P(8♠) = \frac{4}{52} (= \frac{1}{13}) \).

• Since there are two red eights, 8♦ and 8♥, \( P(R \land E) = \frac{2}{52} \).

• \( P(R \lor E) = P(R) + P(E) - P(R \land E) = \frac{26}{52} + \frac{4}{52} - \frac{2}{52} = \frac{28}{52} = \frac{7}{13} \).

**CONDITIONAL PROBABILITY**

• Because \( p \rightarrow q \) is logically equivalent to \( \neg p \lor q \) (recall the MI rule), it follows that \( P(p \rightarrow q) = P(\neg p \lor q) \).

• But, as I’ve noted before, the meaning we’ve assigned to \( \rightarrow \) (via its truth table) does not adequately capture the meaning of “if … then” in every context — notably, those involving judgments of probability.
• Consequently, a rule of probability has been designed to capture the meaning of conditionals in such contexts.

• Specifically, this rule is designed to enable us to calculate the probability that \( q \) is true \textit{conditional on} \( p \)'s being true.

  • We will write “The probability of \( q \) \textit{conditional on} \( p \)” as \( P(q/p) \).
  
  • This notation can also read as:
    • The probability of \( q \) \textit{on the condition that} \( p \).
    • The probability of \( q \) \textit{on} \( p \)
    • The probability of \( q \) \textit{given} \( p \).”
  
  • In statements of the form \( P(q/p) \), \( p \) is the \textit{antecedent} and \( q \) the \textit{consequent}.

• The \textit{conditional rule} is as follows:

\[
\text{Rule 6:} \\
P(q/p) = \frac{P(p \cdot q)}{P(p)}
\]

• Why was it decided that \( P(q/p) \) is the \( P(p \cdot q) \) \textit{divided by} \( P(p) \)?

\textit{Example 1}

• Suppose we are about to draw one card from a well-shuffled deck.

• What's \( P(\clubsuit/A\spadesuit) \), i.e., the probability of our drawing a club \textit{given that} we will draw \( A\spadesuit \)?

  • Intuitively, it is certain, i.e., it should turn out that \( P(\clubsuit/A\spadesuit) = 1 \).

  • \( P(\clubsuit/A\spadesuit) = P(\clubsuit \cdot A\spadesuit)/P(A\spadesuit) = P(A\spadesuit)/P(A\spadesuit) = 1 \).
Example 2

• What’s \( P(\spadesuit/\heartsuit) \), i.e., the probability of our drawing a Spade given that we will draw a Heart?

  • Intuitively, it is nil, i.e., it should turn out that \( P(\spadesuit/\heartsuit) = 0 \). For, given we will draw a Heart, we can’t possibly draw another suit.

  • \( P(\spadesuit/\heartsuit) = \frac{P(\heartsuit \cdot \spadesuit)}{P(\heartsuit)} = \frac{0}{P(\heartsuit)} = 0/\frac{1}{4} = 0 \).

Example 3

• What’s \( P(K\heartsuit/K) \), i.e., the probability of our drawing a King of Hearts given that we will draw a King (of any suit).

  • Intuitively, it should be \( \frac{1}{4} \). For, given that we will draw a King, there is a 1 in 4 chance that it will be the King of Heart instead of one of the other three.

  • \( P(K\heartsuit/K) = \frac{P(K \cdot K\heartsuit)}{P(K)} = \frac{P(K\heartsuit)}{P(K)} = \frac{\frac{1}{52}}{\frac{4}{52}} = \frac{1}{52} \times \frac{52}{4} = \frac{1}{4} \).

Example 4

• What’s \( P(\clubsuit/\clubsuit \lor \spadesuit) \), i.e., the probability of our drawing a Club given that we will draw black card, i.e., either a Club or a Spade?

  • Intuitively, it should be \( \frac{1}{2} \). For, given that we will draw a black card, it must be either a Club or a Spade. Since the number of Clubs = the number of Spades, there is a 1 in 2 chance our card will be a Club.

  • \( P(\clubsuit/\clubsuit \lor \spadesuit) = \frac{P((\clubsuit \lor \spadesuit) \cdot \clubsuit)}{P(\clubsuit \lor \spadesuit)} = \frac{P(\clubsuit)}{P(\clubsuit \lor \spadesuit)} = \frac{\frac{13}{52}}{\frac{26}{52}} = \frac{1/4}{1/2} = 1/4 \times 2 = \frac{1}{2} \).

CONJUNCTION

• The conditional rule is important, not only for what it tells us about conditional probability but also because from it we can immediately deduce the general conjunction rule:

  \[ \text{Rule 7: } P(p \cdot q) = P(p) \times P(q/p) \]
• To prove this, note that by the conditional rule (Rule 6) we have:

\[ P(q/p) = \frac{P(p \cdot q)}{P(p)} \]

• Next, we multiply both sides of the equation by \( P(p) \):

\[ P(p) \times P(q/p) = P(p) \times \frac{P(p \cdot q)}{P(p)} \]

• Since \( \frac{a \times b}{a} = \frac{a}{1} \times \frac{b}{1} = b \) we have:

\[ P(p) \times P(q/p) = P(p \cdot q) \]

• And that is exactly the conjunction rule (with the two sides switched).

**Example 1**

• Consider the situation where you draw a card and then, without replacing the first card, draw a second card.

• Let \( A\spadesuit_1 \) be drawing \( A\spadesuit \) on the first draw and \( A\spadesuit_2 \) be drawing \( A\spadesuit \) on the second draw. What is \( P(A\spadesuit_1 \cdot A\spadesuit_2) \)

\[ P(A\spadesuit_1 \cdot A\spadesuit_2) = P(A\spadesuit_1) \times P(A\spadesuit_2/A\spadesuit_1) = 1/52 \times 0 = 0 \]

**Example 2**

• What is the probability \( P(\text{Red}_1 \cdot \text{Red}_2) \) of choosing a Red card (i.e., a Heart or a Diamond) and then, without putting it back, choosing another?

\[ P(\text{Red}_1 \cdot \text{Red}_2) = P((\text{♥} \spadesuit)_{1} \cdot (\text{♥} \spadesuit)_{2}) = P((\text{♥} \spadesuit)_{1}) \times P((\text{♥} \spadesuit)_{2}/(\text{♥} \spadesuit)_{1}) = 1/2 \times 25/51 = 25/102. \]
Example 3

• What is the probability \( P(A_1 \cdot A_2) \) of drawing an ace on the first draw and (without replacing the first card drawn) another ace on the second draw?

\[
P(A_1 \cdot A_2) = P(A_1) \times P(A_2/A_1) = \frac{4}{52} \times \frac{3}{51} = \frac{1}{13} \times \frac{1}{17} = \frac{1}{221}.
\]

**INDEPENDENCE**

• Our final rule requires us to introduce the important notion of independence.

Two statements \( p \) and \( q \) are **independent** if neither affects the probability of the other, that is, if \( P(q/p) = P(q) \) and \( P(p/q) = P(p) \).

**Example**

• **Hillary Clinton will be the next US President** \((H)\) is independent of **The first card I choose (from a full deck) will be an Ace** \((A)\).

• So \( P(A/H) = P(A) \)

• **The second card I choose will be a Queen** \((Q)\) is **not** independent of **The first card I choose will be Jack** \((J)\).

• In this case, \( P(Q/J) = \frac{4}{51} \).

• When we’re dealing with independent propositions, we can derive a simpler rule for conjunctions, the **restricted conjunction rule**:

\[
\text{Rule 8: } P(p \cdot q) = P(p) \times P(q)
\]

• By Rule 7, \( P(p \cdot q) = P(p) \times P(q/p) \).

• But since \( p \) and \( q \) are independent, \( P(q/p) = P(q) \).
Example

- Consider the probability of selecting an ace twice by drawing from a well-shuffled deck, replacing the card, reshuffling, and drawing a second time.

\[ P(A_1 \cdot A_2) = P(A_1) \times P(A_2) = 1/13 \times 1/13 = 1/169 \]

An Important Observation

- The restricted conjunction rule highlights an important fact about probability.

- Suppose we have a conjunction of independent statements, each of which has a probability of less than 1 but greater than \( \frac{1}{2} \).

- For example, suppose \( P(A) = P(B) = P(C) = 7/10 \).

- What is the probability of the whole conjunction? Because A, B, and C are independent we have:

\[ P(A \cdot B \cdot C) = P(A) \times P(B) \times P(C) = \left(\frac{7}{10}\right)^3 = \frac{343}{1000} \]

- Although each conjunct is more probable than not, the entire conjunction has a probability of less than \( \frac{1}{2} \).

- Bottom line: A conjunction of likely truths can itself be unlikely.

Bayes’ Theorem

- We will now focus on one important implication of our system: Bayes’ theorem.

- Named after the English theologian and mathematician Thomas Bayes (1702–1761).

- Bayes’ theorem gives us an important insight into the relationship between the evidence for a hypothesis and the hypothesis itself, hence, it promises a deeper understanding of the scientific method.

- The letter \( h \) will stand for a given hypothesis.
• The letter \( e \) will stand for a statement that summarizes the observational evidence for that hypothesis.

• Normally, \( e \) is a statement expressing the latest observational evidence for \( h \).

• So Bayes’ Theorem yields particular insight into the effect of a new piece of evidence for a hypothesis for which some body of evidence already exists.

\textit{The Derivation of Bayes’ Theorem}

• Bayes’ Theorem is actually a surprisingly simple theorem of the probability calculus.

• We start with an instance of the conditional rule (Rule 6), for a given hypothesis \( h \) and piece of evidence \( e \):

\[
P(h/e) = \frac{P(e \cdot h)}{P(e)}
\]

• A simple truth table (or proof) shows that \( e \) is logically equivalent to \((e \cdot h) \lor (e \cdot \sim h)\).

• Hence, we can replace \( e \) with \((e \cdot h) \lor (e \cdot \sim h)\) wherever we wish. Doing so in the denominator yields:

\[
P(h/e) = \frac{P(e \cdot h)}{P((e \cdot h) \lor (e \cdot \sim h))}
\]

• By the restricted disjunction rule (Rule 3),

\[
P((e \cdot h) \lor (e \cdot \sim h)) = P(e \cdot h) + P(e \cdot \sim h)
\]

• Hence:

\[
P(h/e) = \frac{P(e \cdot h)}{P(e \cdot h) + P(e \cdot \sim h)}
\]
• By the statement logic rule of commutation for • we have:

\[
P(h/e) = \frac{P(h \cdot e)}{P(h \cdot e) + P(\sim h \cdot e)}
\]

• By applying the general conjunction rule (Rule 7) three times, we arrive at Bayes’ Theorem:

\[
P(h/e) = \frac{P(h) \times P(e/h)}{[P(h) \times P(e/h)] + [P(\sim h) \times P(e/\sim h)]}
\]

**Implications and Applications of Bayes’ Theorem**

• Bayes’ theorem tells us the degree to which a given hypothesis is supported by the evidence, provided that we have three pieces of information: \(P(h), P(e/h),\) and \(P(e/\sim h).\)

  • Recall we can calculate \(P(\sim h)\) from \(P(h).\)

  • \(P(h)\) stands for the **prior probability** of the hypothesis \(h.\)

The **prior probability** of a hypothesis \(h\) is the likelihood of the hypothesis independent of any new evidence \(e.\)

• \(P(e/h)\) is the likelihood that the evidence (or phenomenon in question) would be present, assuming the hypothesis is **true**.

• \(P(e/\sim h)\) is the likelihood that the evidence (or phenomenon in question) would be present, assuming the hypothesis is **false**.
**Example 1**

- Suppose a doctor has diagnosed a patient as having *either* some minor stomach troubles *or* stomach cancer.
  - Let us assume as well that the doctor knows that the patient does not have *both* minor stomach troubles *and* stomach cancer.

- The doctor also knows that, given the symptoms, 30% of patients have stomach cancer; the rest have minor stomach troubles.

- The doctor initially suspects that the patient has only minor stomach troubles.

- But the doctor then conducts a test on the patient.

- The test has positive result = 90% chance of stomach cancer.
  - Let \( H = \text{the patient has stomach cancer} \)
  - Let \( E = \text{the test is positive} \)

- What is the probability of \( H \) given \( E \), i.e., what is \( P(H/E) \)?
  - NOTE: You might think the obvious answer is 90% but recall that the doctor has a prior hypothesis that the patient only has a 30% chance of cancer.

- \( P(H) = \text{the prior probability of } H, \text{ before } E = 30\% = .3 = 3/10 \).

- \( P(\neg H) = 70\% = .7 = 7/10 \) (by the negation rule, Rule 4).

- \( P(E/H) = 90\% = .9 = 9/10 \).

- \( P(E/\neg H) = 10\% = .1 = 1/10 \).

- Plugging these values directly into Bayes’ Theorem, we have:

\[
P(H/E) = \frac{P(H) \times P(E/H)}{P(H) \times P(E/H) + P(\neg H) \times P(E/\neg H)} = \frac{3/10 \times 9/10}{[3/10 \times 9/10] + [7/10 \times 1/10]} = \frac{27/100}{27/100 + 7/100} = \frac{27}{34}
\]

- So, the probability of the hypothesis \( H \) given the evidence \( E \) is 27/34, or approximately .79.
• I will avoid the derivation, but we note that we get an conditional analog of the negation rule (Rule 4):

\[ P(\neg h/e) = 1 - P(h/e) \]

• Bayes’ Theorem is still applicable when there are more than two hypothesis competing for our credence.

• If \( h_1, h_2, \) and \( h_3 \) are three mutually exclusive, jointly exhaustive hypotheses, then \( \neg h_1 \) is equivalent to \( h_2 \lor h_3 \).

• Hence, substituting into Bayes' Theorem, we have

\[
P(h_1/e) = \frac{P(h_1) \times P(e/h_1)}{[P(h_1) \times P(e/h_1)] + [P(h_2 \lor h_3) \times P(e/(h_2 \lor h_3))]} \]

• And this, in turn, reduces to

\[
P(h_1/e) = \frac{P(h_1) \times P(e/h_1)}{[P(h_1) \times P(e/h_1)] + [P(h_2) \times P(e/h_2)] + [P(h_3) \times P(e/h_3)]} \]

• In other words, we can accommodate as many hypotheses as we like (provided they are mutually exclusive and jointly exhaustive), simply by adding relevant clauses to the denominator.