CHAPTER ONE

LOGIC AND DATA ANALYSIS

In any scientific discipline, the goal is to **infer** underlying world-states and the rules by which those world-states change, based on observed data. In the biosciences, particularly the behavioral and neural sciences, this involves inferring the rules by which organisms acquire and process environmental inputs to produce **behavior**; that is, to infer the fundamental rules by which neural circuits and systems process environmental inputs to govern behavior. In some cases the behavior of interest is defined in terms of the organism as a whole, while in other cases our interest is in behavior at a more fine-grained level, such as the behavior of the thyroid or the behavior of the superior colliculus (a paired structure of the mammalian midbrain). In medical research, the goal is to understand the behaviors of organs and organ systems in healthy and disease states, knowledge that is applied to develop treatments; treatment efficacies are compared via the same logic used to compare scientific theories (see Box 1.1). If these are our goals as scientists, our next question must be: How are inferences made? What is the logic relating experimental data and our **hypotheses**? As we will see throughout this text, the logic relating experimental data to scientific inference permeates every aspect of the scientific process: it tells us how best to test existing theories, how to design experiments, how to improve current theories, and how to test new medical treatments. **Data analysis** is nothing but an exercise in this logic.

Much of this text will be concerned with testing scientific models (Fig. 1.1) of the behavior of organisms or their components, and for this purpose we will ultimately develop procedures for both measurement and hypothesis testing. Hypothesis testing is done in two stages. Initially, when we become interested in explaining some aspect of behavior, we have performed no formal **experiments**. We know only that there is some behavior that has certain characteristics – whatever it is that piqued our interest, whether that be some disease state or an aspect of normal economic, psychological or neurophysiological behavior that we find interesting in its own right. At this stage, we

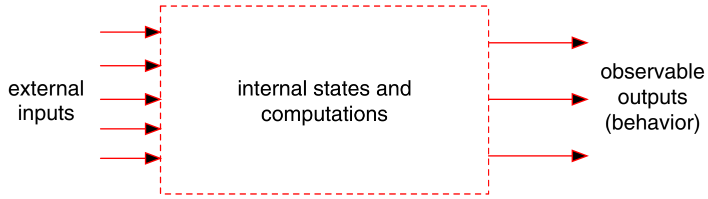

Figure 1.1 Basic anatomy of a behavioral theory. Behavioral theories must describe the interaction between observable external inputs to a system, the internal states and computations of that system, and the observable outputs (behaviors) produced by interaction of the preceding.

BOX 1.1 The Anatomy of a Scientific Theory: Examples from Psychology, Economics, Medicine, and Neuroscience

Theories are, to put it in terms of the logical arguments described in Section 1.1, statements/propositions. Scientific hypotheses are a special class of proposition, because they are testable. The basic criterion for a testable behavioral theory is that it makes predictions that can be compared to observations of behavior. A typical example from neuroscience might be the hypothesis that posits a particular population of neurons in primary motor cortex that controls arm movements. This theory is straightforward in the sense that there must be neurons that are active before an arm movement, and whose firing tells us about the type of movement that will be made (e.g., move the hand 5 cm forward); ideally, we could artificially stimulate those neurons to produce predictable arm movements.

If theories are simply statements that represent a testable guess or belief about some system's behavior, then what is the minimal structure of a testable behavioral theory? A behavioral theory must describe a behavior, and it must describe the causes/determinants of that behavior. In other words, a behavioral theory will describe the interaction of three elements (shown graphically in Fig. 1.1): external inputs to a system, the internal states and computations of that system, and an observable output or behavior of the system. Furthermore, that description must be specific enough to allow us to say which combinations of external inputs along with internal states and computations will produce the behavior, and also how changes to these precursors will affect behavior: it is this specificity that makes a theory testable.

We will encounter a range of behavioral theories in this text, from behavioral economic theories, to sensory/motor neuroscience, medicine, and theories of personality and social psychology. In these various cases we will be interested in different 'systems,' whether that be the visual system, the enteric nervous system, a single neuron, or the entire organism (human, rat, rabbit, etc.). In all of these cases, we will expect to define a set of inputs to the system (a system which may, e.g., in the case of a single-neuron system, be entirely contained within the body), a set of internal structures and computations relevant to the underlying theory, and a behavioral response mediated by those inputs and internal computations. Within the realm of motor control, we may want to examine a theory that makes predictions regarding which cells will and will not cause arm movements (primary motor neurons will, but for example primary visual neurons won't), as well as how patterns of neural activity (one type of behavior) will relate to arm movements (another type of behavior). In medicine, many theories attempt to relate treatment to behavioral outcomes that involve recovery from illness or injury. The most common examples come from drug development, such as medications designed to delay the progression of Alzheimer's disease. Here, a test of the drug (and therefore the theory that the drug has the desired effect) would involve comparison of the timecourse of disease progression in individuals taking versus not taking the drug. In the field of economics a classical theory of economic choice is the rational choice model, in which it is predicted that human economic decisions will tend to maximize expected monetary gain. Here a test of the model involves comparison of human economic choices with those of an ideal (rational) decider that maximizes expected gain (we define expected gain and an ideal decision-maker in several later chapters). Finally, within the field of personality psychology, many theories involve identifying stable characteristics of cognition and emotion that predict chronic emotional responses and social behaviors. A typical experiment might use a pencil-and-paper 'personality test' to classify individuals into categories that are subsequently used to predict questionnaire answers by the same individuals, indicating their likely emotional responses and decisions in emotionally charged circumstances.

Every testable scientific theory must make predictions, because models are compared and tested by comparing the predictions of competing models against observed data. Each experimentally observed datum contributes to our ability to discriminate among competing models, because it will be more consistent with one model over its competitors. A good behavioral theory will always posit *specific* predictable relationships between the internal states of the system (e.g., current firing of sensory neurons, chronic cognitive or emotional activation, intramuscular lactic acid levels), inputs to the system (e.g., sensory inputs, social or emotional circumstances, or monetary decision scenarios), and the behaviors produced by that system. The states of the system that mediate between inputs and outputs are characterized by both physical constants and neural response functions, both of which can be measured in terms their characteristic parameters. In the case of the visual system, one such parameter would be the focal length of the corneal lens relative to the depth of the eyeball (which is used to predict image sharpness/blur), whereas in the case of monetary decision-making an important parameter is 'risk aversion,' which describes the degree to which one will trade or forego high potential monetary gains in favor of lower gains because the lower-gain choice is also less likely to lead to a loss (predicting, among other things, the types of stock investments one is comfortable making). When we have a theory that posits these types of internal parameters, we may choose to **measure** the values of those parameters, perhaps within a certain subset of the human population (e.g., those who have or have not suffered a concussion). Measurements of those parameters represent a common type of to-be-tested proposition (i.e., 'the value of the parameter is'). These measurement problems, in addition to being useful in their own right, form the basis for more advanced calculations used within Bayesian statistics.

The most straightforward of these extensions is **model selection**, which is used to differentiate among competing models (theories) of a given behavior; thus, when we are testing a scientific theory we are performing model selection. Scientific theories progress when we have multiple competing models *of a single behavior*. Tests of these theories constitute a model selection problem, where our goal is to select the model that is most consistent with our experimental data. So whereas the solution to a parameter estimation problem takes data and uses it to provide the best estimate for the numerical value of some parameter from a single model, the solution to a model selection problem tells us which of several competing scientific theories (models) provides the most reasonable fit to our experimental data (often without worrying about the exact values of model parameters).

generate new theories by noting regularities in behavior. That is, observed regularities in behavior allow us to generate an initial set of informed guesses (hypotheses) concerning underlying rules governing behavior. This set of initial hypotheses is formed based on the fact that our initial data (observed regularities) will be logically consistent with certain underlying rules governing behavior, but not others. Once we have defined an initial set of hypotheses, one narrows the field of *most-plausible* hypotheses by performing experiments. Experiments are designed to discriminate among hypotheses. If properly designed, they will produce data that is logically *more* consistent with some hypotheses than with others. Notice that both parts of the process require us to analyze data – either to determine what initial hypotheses are tenable, given the initially observed regularities in behavior, or to later discriminate among those initial hypotheses, given the results of our experiments. The details of this practice of data analysis will occupy us for the remainder of the text.

As we will see below, there are two methods of assessing the impact of data on theory: **deduction** and **induction**. If one has previous basic coursework in logic, symbolic logic, or Boolean algebra, the topic was deduction. And while deduction is a useful and well-known method of scientific inference, it is only applicable in a very limited set of circumstances: those circumstances in which one is eliminating impossible theories that *conflict with known facts*. That is, it eliminates the *impossibilities*; it cannot, however, be used when choosing among several competing *possibilities*. When we have gone as far as we can with deduction, we must narrow the field of possibilities using inductive, rather than deductive, inferences that connect data to theory. Until recently, the **frequentist approach** was the most common method of inductive inference; but advances in the foundations of probability theory since the mid-twentieth century have shown that the only logically consistent method of induction that meets basic criteria of rationality is the approach based on probability theory. Any other method that is not ultimately identical to probability theory will be logically flawed, in the sense that it will be internally inconsistent or show other logically undesirable behavior. Here, we will demonstrate and explore the two most basic applications of probability theory to data analysis: measurement and hypothesis testing.

There are four major goals to accomplish in the remainder of this introductory chapter. The first is to provide a general introduction to scientific inference, as it relates data to hypotheses (Section 1.1). Next, in Section 1.2, we examine the concept of **uncertainty**, which is central to the distinction between inductive and deductive inference, and data analysis generally. In Section 1.3 we introduce the logic of data analysis via probability theory, whose details will occupy us for the remainder of this text. In addition to introducing you to the theoretical landscape of formal data analysis (Sections 1.1–1.3), the final goal of this chapter is to introduce the practice of **data visualization**. The critical, albeit relatively unstructured process of data visualization gives us our first look at any new dataset. The goal in data visualization is to verify that we have obtained data that, on the whole 'makes sense' given what we already know about the phenomenon under study. Data visualization allows us both to identify when errors might have been committed during collection, import, transcription, storage, retrieval, etc. of the data, as well as to identify possible unexpected mechanisms and characteristics of the object of study (e.g., a previously unrecognized input to a particular neural circuit); this is the topic of Section 1.4.

1.1 The Logic of Inference

The real-world analysis of experimental data depends on the interplay of deductive and inductive inference. **Inductive inference** is the reasoning used to select among competing alternative hypotheses when none directly conflicts with known facts; it is the topic that will occupy us for the remainder of this text. **Deductive inference**, by contrast, is used to eliminate impossible theories that are in conflict with known facts. Deduction is most applied during the initial stages of investigation when initial hypotheses are being examined.

1.1.1 Deductive Inference and Rational Belief Networks

Deductive inference occurs in two forms, the *modus ponens* and *modus tollens* types. We first state these deductive arguments (where 'A' and 'B' represent statements that could potentially be true or false, such as scientific hypotheses, usually termed **propositions** within the context of a logical argument), written both symbolically and in words:

Modus Ponens ('mode of affirmation'):

$A \Rightarrow B$ [If A is true, then B must also be true]
A [A is affirmed – that is, A is declared true]
$\therefore B$ [Therefore B is true]

Modus Tollens ('mode of denial'):

$A \Rightarrow B$ [If A is true, then B must also be true]
$\bar{B}$ [$\bar{B}$ denotes the negation[1] of B – in other words, the truth of B is denied]
$\therefore \bar{A}$ [Therefore A is false]

The logic of these deductive **arguments** is as follows: In the ***modus ponens*** case, it is first assumed that there is a logical connection whereby the truth of the proposition A guarantees the truth of the proposition B (this is stated on the first line of the argument). In addition, the second line of the argument asserts A to be true. The truth of proposition B is the logical consequence of the assertions made on lines 1 and 2, because if B is always true when A is true (line 1), and we assume that A is true (line 2), it follows that we must also assume that B is true (line 3). In the ***modus tollens*** case line 1 is identical, but the negation[1] of B is asserted on line 2. Since the rule given on line 1 requires B to *always* be true when A is true, the denial of B implies the denial of A (because B would have been true if A had been true). Notice that while both of these deductive arguments are **valid** regardless of the meanings of A and B, the truth of the conclusions in both argument forms is conditional on the truth of the statements, or **premises** given on lines 1 and 2. A valid argument form in which the premises are true is called a **sound argument**.

[1] Negation is the logical operation by which the proposition is asserted to be untrue. Thus the negation of a compound proposition, A ≡ Harvey is a blind cat, is correctly given as 'it is not true that Harvey is a blind cat.' While this could

Programming Aside: IF Else and the Modus Ponens Deduction

We can use the built-in if . . . then logical evaluation to explore the modus ponens argument. Let's create the proposition, rabbitsAreFluffy and the proposition rabbitsAreSoft, and further assert: *rabbitsAreFluffy* $\Rightarrow$ *rabbitsAreSoft*. We can write a quick program to explore this:

```
rabbitsAreFluffy=1;

%implication defined below
if rabbitsAreFluffy,
    rabbitsAreSoft=1;
else rabbitsAreSoft=nan;

%conclusion
rabbitsAreSoft
```

The proposition following *if* can either be true or false, and the next line defines the second part of the implication *rabbitsAreFluffy* $\Rightarrow$ *rabbitsAreSoft*, by telling us what is also true if rabbitsAreFluffy=true. If run as is, the final line (query of the rabbitsAreSoft proposition) will yield a 1 (true). If we change rabbitsAreFluffy to false, then the rabbitsAreSoft proposition has an undefined value based on the implication *rabbitsAreFluffy* $\Rightarrow$ *rabbitsAreSoft*, and so we assign it nan.

There are also **invalid** (and therefore **unsound**) argument forms that look quite similar to the deductive forms above. For example, the argument

$A \Rightarrow B$	[If A is true, then B must also be true]
B	[B is affirmed]
$\therefore A$	[Therefore A is fallaciously declared true]

is not deductively valid; it is a logical **fallacy**. This argument (called *affirming the consequent*) is invalid because the third line does not *necessarily* follow in all cases where the premises (lines 1–2) are true. Consider the example: A = 'we are reading a Socratic dialogue,' and B = 'we are reading ancient Greek philosophy.' In this example, further assume both A and B are true. Despite this, line 3 still doesn't logically (deductively) follow: the second line could be true reading any number of philosophers (Zeno, Pythagoreas, Parmenides, Aristotle, etc.), so this argument form (affirming the consequent), is not a valid deductive argument.

OK, now let's assume we have constructed a valid argument, such as a *modus ponens* deduction. Is this argument, being valid, guaranteed to yield a true conclusion? Perhaps surprisingly, the answer is no. A valid argument does not guarantee anything about the real-world truth of its conclusion: the conclusion may be true or untrue, despite having resulted from a valid argument. To see why this is so, and also to gain a better understanding of why deduction is nevertheless useful in the face of these facts, let's define the two true premises, A = 'Blind cats catch prey auditorily,' and B = 'Sockeye is a salmon variety,' and consider these premises in the *modus ponens* argument. We see

mean that Harvey is neither cat nor blind, it could also mean that Harvey is not a cat but is in fact blind, or it could mean that Harvey is a sighted cat. Elementary propositional logic is reviewed at the start of Chapter 2.

that because the first line of the argument may not be true (i.e., there is no logical connection between blind cats' hunting methods and salmon varieties), the conclusion of the argument may not be true (it is not *logically necessary* that it be true). However, the logical form of the argument is unaffected by these considerations: *If* the first two lines of the argument *were* true, then the conclusion *must* follow. Indeed, one could think of a deductive argument as defining a **rational belief network**, where *if* one believes A to be true, and also believes that A implies B, then a rational being *must also* believe B. The job of the scientist is to construct belief networks (both deductive and, as we will see below, inductive) involving behavioral theories, and extract experimental facts regarding the world to either support or disprove those theories via the logical connections of those networks.

Exercises

1) What can be concluded from the following arguments, if we define:
   $C = A \Rightarrow B$ (C is defined as: if A, then B)
   $D = B \Rightarrow \bar{A}$
   a. C, A, therefore ______.
   b. $C, \bar{B}$, therefore ______.
   c. D, A, therefore ______.
   d. D, B, therefore ______.
2) You are given four cards, two front-up, and two front-down. You are told that on every card with an even number on its front, there is a vowel printed on its back. There is a 2 and an 11 printed on the two front-up cards, and a 'q' and 'u' visible on the backs of the two face down cards. Which cards, if turned over, provide a test of the assertion given above concerning what is printed on the fronts and backs of individual cards?

1.1.2 Plausible Inference

All deductive conclusions are known with certainty. They are true or false, not 'likely true' or 'probably false.' This is the beauty of a deductive argument, and is a consequence of the certainty inherent in the implication (if A is true, B ***must*** also be true). Once such a 'certain implication' can be defined, we need only agree upon one of the premises (that A is true or that B is false), and from this we can determine what also *must* be true or *must* be false. There is no ambiguity or uncertainty in the conclusion of a deductive argument. Interestingly, this very same beautiful certainty is also the drawback of a deductive argument, because a deductive argument has nothing to say about 'uncertain implication'; i.e., when it is only known that if A is true it provides ***evidence*** that B is also true, but nevertheless does not *logically guarantee* that B be true. Given that our information regarding the connection between theory and data will rarely define a **certain implication**, the rules of deductive inference cannot be applied without caveat. It was therefore a breakthrough for scientific inference when Richard Cox showed that probability theory, as originally developed by Pascal and Fermat in the seventeenth century, Bernoulli and most completely by Laplace is the unique method of consistent reasoning with **uncertain** information.

BOX 1.2 What Is Probability?

Two types of deductive argument were described: the conclusion of a *sound* deductive argument describes the state of the world; the conclusion of a *valid* deductive argument is a statement of our information; that is, in a *valid* deductive argument, the conclusion tells us whether or not a proposition is true, *assuming* the statements in the argument are also true. Probabilities, and the inductive inferences they support, are of the second kind. Probabilities quantify the strength of logical support – given as a number between 0 and 1, favoring the truth of the object of the probability statement that can be asserted *assuming* conditioning statements are true. The important point is that, just as with a valid deductive argument, we do not require that conditioning statements be true in order to draw valid conclusions, only that we treat them as true for the purpose of computing probabilities; those probabilities tell us the evidence available for the object of the probability statement, *jf* the conditioning statements *were* true. This is very different from the frequency definition of probability, defined in terms of observed frequencies, which must, by definition, be true.

Probabilities defined in terms of information are an extension of the 'logical belief network' described for deductive reasoning that defines not just what a logical agent/being *must believe* (when probabilities are 0 or 1), but also what a logical agent *should believe* (what is more or less likely, defined as probabilities *between* 0 and 1). It describes the extent to which a logical agent should hold certain beliefs, given other logically related beliefs. Understood in this way, assignment of numerical values for probabilities is purely an exercise in logic; subjectivity does not enter into the calculation any more than it enters into a deductive argument. This information-based definition of probability allows the broadest use of probability, not limiting the scope of application the way the frequency definition would.

Why does the frequency definition limit the scope of probability theory? Because there is no frequency-definition way to assign probabilities to the class of proposition that is most important to a scientist: hypotheses and models. The most common example of such a proposition is meteorological: tomorrow's weather. The weather on any given day or time has only one value, because a statement such as 'it will rain tomorrow at 9 AM' is not the type of thing that can sometimes be true and sometimes be false. Tomorrow's weather is not the type of event that can have 'an infinite series of identical repetitions' (as could be theoretically achieved with coin tosses). Rather, it's truth-value (rain/no-rain) is constant and unchanging. The same is true of any scientific hypothesis, because they belong to a class of proposition that cannot, for example, be true 8 times out of 10 and false 2. Once we lose the ability to define the probabilities of hypotheses, we also lose the ability to test hypotheses in the most straightforward way – comparing their probabilities.

Cambridge University Press
978-1-108-83556-5 — Bayesian Data Analysis for the Behavioral and Neural Sciences
Todd E. Hudson
Excerpt
More Information

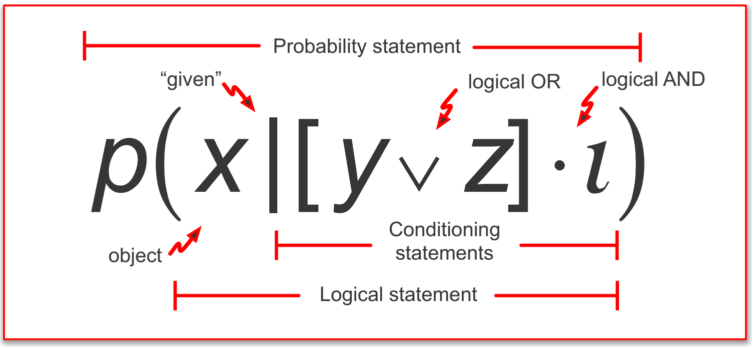

Figure 1.2 Written format for probability statements. In words, this expression represents the probability of the truth of the proposition represented by x, *given* (that is, *under the assumption*) that both the general background information about the problem (which we will represent by the Greek iota, ι) *and either* the proposition y *or* z is true.

Cox showed that any **inductive**, or **plausible inference** based on uncertain information is equivalent to a series of applications of the **sum and/or product rules of probability theory**. Probability statements are written symbolically in the form shown in Figure 1.2, and the **degrees of certainty** represented by probabilities, where each probability, $p(\bullet \mid \bullet)$, is a real number assigned to the logical statement in parentheses (the $\bullet$ are simply placeholders, indicating where logical propositions will appear).
The sum and product rules of probability theory are used to combine probabilities:

$$\text{Sum Rule:} \quad p(A \vee \bar{A} \mid \iota) = p(A \mid \iota) + p(\bar{A} \mid \iota) = 1 \tag{1.1}$$

$$\text{Product Rule:} \quad \begin{aligned} p(A \cdot B \mid \iota) &= p(A \mid \iota)p(B \mid A \cdot \iota) \\ &= p(B \mid \iota)p(A \mid B \cdot \iota) \end{aligned} \tag{1.2}$$

In this framework, complete certainty of truth (e.g., truth of the proposition x in Fig. 1.2) is represented by setting the value of the probability statement to $p(\bullet \mid \bullet) = 1$. Complete disbelief of the truth of the object is represented by $p(\bullet \mid \bullet) = 0$. The proposition preceding the vertical bar will be referred to as the **object** of the probability statement, and the statement(s) following the vertical bar are **conditioning statements**. The numerical probability value indicates the degree of certainty one can reasonably attribute to the object logical statement. Therefore, probabilities indicate how much evidence there is for asserting the truth of the proposition that is the object of the probability statement; they give the 'weight of evidence' available in the conditioning statements for supporting the truth of the object of the probability statement (see Box 1.2).

Both the object and conditioning parts of the probability statement consist of a proposition or combination of propositions connected by logical OR[2] (written '$\vee$') or logical AND[3] (written '$\cdot$'). Propositions following the vertical bar represent a state of information under which the numerical probability assignment is justified. For example, (1.1)

[2] A proposition which is itself composed of two propositions connected by logical OR is true if either one or both of the two composing propositions is true. Thus, the proposition $D = E \vee F$ is true if either the proposition E is true or the proposition F is true, or if both are true simultaneously.

[3] A compound statement composed of two propositions connected by logical AND is true only if both of the composing propositions are simultaneously true.

can be read as: The truth of one out of a pair of mutually exclusive and exhaustive possibilities (given background knowledge ι) is guaranteed. Notice that all of the probabilities listed above are conditional on (at least) the background information represented by the proposition ι. For example, in (1.1) this information would include statements indicating that the proposition A can be either true or false, as well as the conditions necessary to give a numerical value to the two individual statements, $p(A \mid \iota)$ and $p(\bar{A} \mid \iota)$. It is true in general that a probability statement lacking **conditioning statements** is not technically a meaningful expression, because the state of information justifying a particular numerical assignment has not been given. For example, writing $p(h)$ for the probability that a flipped coin will come up heads is indeterminate because as written it could be assigned a probability 0, $^1/_2$, 1, or any intermediate value, depending on whether none, one, or both of the coin's faces is known to be stamped with a head, our information about the distribution of mass within the coin, and in what manner the coin is to be flipped. In other words, the statement $p(h)$ fails to include the background information necessary to assign a unique probability.

Exercises

1) Write the following in words:
   a. $p(A \vee \bar{A} \vee B \mid C \cdot \iota) = x$
   b. $p([A \vee B] \cdot C \mid \iota) = y$
   c. $p(A \cdot \bar{A} \cdot B \mid \iota) = z$
   d. $p(A \cdot B \mid \iota) = p(A \mid \iota)p(B \mid A \cdot \iota)$
   e. $p(A \cdot B \mid \iota) = p(B \mid \iota)p(A \mid B \cdot \iota)$
2) What is the numerical value of x in 1a?
3) What is the numerical value of z in 1c?
4) The probability of the proposition B (given ι) is said to be independent of the proposition A if $p(A \cdot B \mid \iota) = p(A \mid \iota)p(B \mid \iota)$. Explain why equation (1.2) above could not (as a general rule) have been: $p(A \cdot B \mid \iota) = p(A \mid \iota)p(B \mid \iota)$.

Example: Deduction and Plausible Inference

How does probability-theory-based inference relate to the deductive forms?[4] Could one, for example, create a theory of plausible inference in which induction is a type of deduction? This would seem desirable, since it is generally agreed that deduction is *the* paradigm method of logical inference. Not only is the answer 'no,' but it turns out that the situation is reversed: deduction is contained in the rules of probability theory as a special case. First, consider the probability $p(B \mid A \cdot \iota)$, from the product rule:

$$p(B \mid A \cdot \iota) = \frac{p(A \cdot B \mid \iota)}{p(A \mid \iota)} \quad (1.3),$$

where we define the proposition ι to include the logical implication $A \Rightarrow B$, and will also assert $p(A \mid \iota)$. Under these assumptions, the probability of the **conjunction** of A and B is always 1 when the probability of A alone is 1. Thus, the ratio $\frac{p(A \cdot B|\iota)}{p(A|\iota)}$ in (1.3), and therefore $p(B \mid A \cdot \iota)$ must also be 1 in this case. This is the *modus ponens* argument – simultaneously asserting A and ι guarantees the truth of B.

[4] This example is adapted from Jaynes (2003).