

CHAPTER 1

Introduction

The evolution of modern society is driven by decisions that affect the welfare and choices of large groups of individuals. Of the scores of examples, a few will illustrate the characteristics of decisionmaking that motivate our approach:

1. A new drug has been developed in the laboratories of a private firm over a period of several years and at a cost of tens of millions of dollars. It has been tested in animals, and in increasingly larger groups of human beings in a succession of highly structured clinical trials. If the drug is approved by the Food and Drug Administration (FDA), it will be available for all licensed physicians to use at their discretion. The FDA must decide whether to approve the drug.
2. Since the mid-1980s evidence from many different sources, taken together, clearly indicates that the earth's climate is warming. The evidence that this warming is due to human activities, in particular the emission of carbon dioxide, is not as compelling but becomes stronger every year. The economic activities responsible for increases in the emission of carbon dioxide are critical to the aspirations of billions of people, and to the political order that would be needed to sustain a policy that would limit emissions. How should the evidence be presented to political leaders who are able to make and enforce decisions about emissions policy? What should their decision be?
3. A multi-billion-dollar firm is seeking to buy a firm of similar size. The two firms have documented cost reductions that will be possible because of the merger. On the other hand, joint ownership of the two firms will likely increase market power, making it in the interests of the merged firm to set higher price cost margins than did the two firms separately. How should lawyers and economists—whether disinterested or not—document and synthesize the evidence on both points for the regulatory authorities who decide whether to permit the merger? How should the regulatory authorities

make their decision? If they deny the merger, the firms must decide whether to appeal the decision to the courts.

4. A standard petroleum refining procedure produces two-thirds unleaded gasoline and one-third heating oil (or jet aviation fuel, its near equivalent). Refinery management buys crude oil, and produces and sells gasoline and heating oil. The wholesale prices of these products are volatile. Management can guarantee the difference between selling and buying prices, by means of futures contracts in which speculators (risk takers) commit to purchasing specified amounts of gasoline or heating oil, and selling agreed-on amounts of crude oil, at fixed prices. Should management lock in some or all of its net return in this way? If some, then how much?

These decisions differ in many ways. The second and third will appear prominently in the media; the first might, the last rarely will. The second is a matter of urgent global public policy, and the last is entirely private. The other two are mixtures; in each case the final decision is a matter of public policy, but in both the matter is raised to the level of public policy through a sequence of private decisions, in which anticipation of the ultimate public policy decision is quite important.

Yet these decisions have many features in common:

1. The decision must be made on the basis of less-than-perfect information. By “perfect information” is meant all the information the decisionmaker(s) would requisition if information were free, that is, immediately available at no cost in resources diverted from other uses.
2. The decision must be made at a specified time. Either waiting is prohibited by law or regulation (examples 1 and 3), is denied by the definition of the decision (example 4), or “wait” amounts to making a critical choice that may circumscribe future options (example 2).
3. The information bearing on the decision, and the consequences of the decision, are primarily quantitative. The relationship between information and outcome, mediated by working hypotheses about the connection between the two, is nondeterministic.
4. There are multiple sources of information bearing on each decision. Whether the information is highly structured and derived from controlled experiments (example 1), consists of numerous studies using different approaches and likely reaching different conclusions (examples 2 and 3), or originates in different time periods and settings whose relation to the decision at hand must be assessed repeatedly (example 4), this information must be aggregated, explicitly or implicitly, in the decision.

We will often refer to “investigators” and “clients,” terms due to Hildreth (1963). The investigator is the applied statistician or econometrician whose function is to convey quantitative information in a manner that facilitates and thereby improves decisions. The client may be the actual decisionmaker, or—more often—another scientist working to support the decision with information. The client’s identity

and preferences may be well known to the investigator (example: an expert witness hired by any interested party), or many clients may be unknown to the investigator (example: the readers of a subsequently well-cited academic paper reporting the investigator's work).

The objective of this book is to provide investigators with understanding and technical tools that will enable them to communicate effectively with clients, including decisionmakers and other investigators. Several themes emerge:

1. Make all assumptions explicit.
2. Explicitly quantify all of the essentials, including the assumptions.
3. Synthesize, or provide the means to synthesize, different approaches and models.
4. Represent the inevitable uncertainty in ways that will be useful to the client.

The understanding of effective communication is grounded in Bayesian inference and decision theory. The grounding emerges not from any single high-minded principle, but rather from the fact that this foundation is by far the most coherent and comprehensive one that presently exists. It may eventually be superseded by a superior model, but for the foreseeable future it is the foundation of economics and rational quantitative decisionmaking.

The reader grounded in non-Bayesian methods need not take any of this for granted. To these readers, the utility of the approach taken here will emerge as successive real problems succumb to effective treatment using Bayesian methods, while remaining considerably more difficult, if not entirely intractable, using non-Bayesian approaches.

Simulation methods provide an indispensable link between principles and practice. These methods, essentially unavailable before the late 1980s, represent uncertainty in terms of a large but finite number of synthetic random drawings from the distribution of unobservables (examples: parameters and latent variables), conditional on what is known (examples: data and the constraints imposed by economic theory) and the model(s) used to relate unobservables to what is known. Algorithms for the generation of the synthetic random drawings are governed by this representation of uncertainty. The investigator who masters these tools not only becomes a more fluent communicator of results but also greatly expands the choices of contexts, or models, in which to represent uncertainty and provide useful information to decisionmakers.

1.1 TWO EXAMPLES

This chapter is an overview of the chapters that follow. It provides much of what is needed for the reader to be a knowledgeable client, that is, a receiver of information communicated in the way just discussed. Being an effective investigator requires the considerably more detailed and technical understanding that the other chapters convey.

1.1.1 Public School Class Sizes

The determination of class size in public schools is a political and fiscal decision whose details vary from state to state and district to district. Regardless of the details, the decision ultimately made balances the fact that, given the number of students in the district, a lower student : teacher ratio is more costly, against the perception that a lower student : teacher ratio also increases the quality of education. Moreover, quality is difficult to measure. The most readily available measures are test scores. Changes made in federal funding of locally controlled public education since 2001 emphasize test scores as indicators of quality, and create fiscal incentives for local school boards to maintain and improve the test scores of students in their districts.

In this environment, there are several issues that decisionmaking clients must address and in which Bayesian investigation is important:

1. What is the relationship between the student : teacher ratio and test scores? Quite a few other factors, all of them measurable, may also affect test scores. We are uncertain about how to model the relationship, and for any one model there is uncertainty about the parameters in this model. Even if we were certain of both the model and the parameters, there would still be uncertainty about the resulting test scores. Full reporting and effective decisionmaking require that all these aspects of uncertainty be expressed.
2. The tradeoff between costs, on one hand, and quality of education, on the other hand, needs to be expressed. “Funding formulas” that use test scores to determine revenues available to school administrators (the clients) express at least part of this relationship quantitatively. In addition, a client may wish to see the implications of alternative valuations of educational quality, as expressed in test scores, for decisions about class size. Funding formulas may be expressed in terms of targets that make this an analytically challenging problem. The simulation methods that are an integral part of contemporary Bayesian econometrics and statistics make it practical to solve such problems routinely.
3. Another set of prospective clients consists of elected and appointed policymakers who determine funding formulas. Since these policymakers are distinct from school administrators, any funding formula anticipates (at least implicitly) the way that these administrators will handle tradeoffs between the costs of classroom staffing and the incentives created in the funding formulas. Depending on administrators’ behavior, different policies may incur higher, or lower, costs to attain the same outcome as measured by test scores.

Bayesian analysis provides a coherent and practical framework for combining information and data in a useful way in this and other decisionmaking situations. Chapters 2 and 3 take up the critical technical steps in integrating data and other sources of information and representing the values of the decisionmaking client. Chapter 4 provides the simulation methods that make it practical and routine to

undertake the required analysis. The remaining chapters return to this particular decision problem at several points.

1.1.2 Value at Risk

Financial institutions (banks, brokerage firms, insurance companies) own a variety of financial assets, often with total value in the many billions of dollars. They may include debt issued by businesses, loans to individuals, and government bonds. These firms also have financial liabilities: for example, deposit accounts in the case of private banks and life insurance policies in the case of insurance companies. Taken together, the holdings of financial assets or liabilities by a firm are known as its “portfolio.”

The value of an institution’s portfolio, or of a particular part of it, is constantly changing. This is the case even if the institution initiates no change in its holdings, because the market price of the institution’s assets or liabilities change from day to day and even minute to minute. Thus every such institution is involved in a risky business. In general, the larger the institution, the more difficult it is to assess this risk because of both the large variety of assets and liabilities and the number of individuals within the institution who have authority to change specified holdings in the institution’s portfolio.

Beginning about 1990 financial institutions, and government agencies with oversight and regulatory responsibility for these institutions, developed measures of the risk inherent in institutions’ portfolios. One of the simplest and most widely used is value at risk. To convey the essentials of the idea, let p_t be the market value of an institution’s entire portfolio, or of a defined portion of it. In the former case, p_t is the net worth of the institution—what would remain in the hypothetical situation that the institution were to sell all its assets and meet all of its liabilities. In the latter case it might be (for example) the institution’s holding of conventional mortgages, or of U.S. government bonds.

The value p_t is constantly changing. This is in part a consequence of holdings by the institution, but it is also a result of changes in market prices. Value at risk is more concerned with the latter, so p_t is taken to be the portfolio value assuming that its composition remains fixed. “Value at risk” is defined with respect to a future time period, say, t^* , relative to the current period t , where $t^* > t$ and $t^* - t$ may range from less than a day to as long as a month. A typical definition of value at risk is that it is the loss in portfolio value v_{t,t^*} that satisfies

$$P(p_t - p_{t^*} \geq v_{t,t^*}) = .05. \quad (1.1)$$

Thus value at risk is a hypothetical decline in value, such that the probability of an even greater decline is 5%. The choice of .05 appears arbitrary, since other values could be used, but .05 is by far the most common, and in fact some regulatory authorities establish limits of v_{t,t^*} in relation to p_t based on (1.1).

The precise notion of probability in (1.1) is important. Models for establishing value at risk provide a distribution for p_{t^*} , conditional on p_t and, perhaps, other

information available at time t . From this distribution we can then determine v_{t,t^*} . Most models used for this purpose are formulated in terms of the period-to-period return on the portfolio

$$r_t = (p_t - p_{t-1})/p_{t-1},$$

and statistical modeling usually directly addresses the behavior of the time series

$$y_t = \log(1 + r_t) = \log(p_t/p_{t-1}). \quad (1.2)$$

One of the simplest models is

$$y_t \sim N(\mu, \sigma^2). \quad (1.3)$$

Even this simple model leaves open a number of questions. For example, is it really intended that the same model (including the same mean and variance) pertains today for “high tech” stocks as it did in 1999, before the rapid decline in their value? In any event, the parameters μ and σ^2 are unknown, so how is this fact to be handled in the context of (1.1)? This problem is especially vexing if μ and σ^2 are subject to periodic changes, as the high-tech example suggests at least sometimes must be the case if we insist on proceeding with (1.3).

One of the biggest difficulties with (1.3) is that it is demonstrably bad as a description of returns that are relatively large in absolute value, at least with fixed μ and σ^2 . If we take as the fixed values of μ and σ^2 their conventional estimates based on daily stock price indices for the entire twentieth century, then the model implies that “crashes” like the one that occurred in October 1987, are events that are so rare as to be impossible for all practical purposes. [For the daily *Standard and Poors 500* stock returns for January 3, 1928–April 30, 1991, from Ryden et al. (1998) used in Sections 7.3 and 8.3, the mean is .000182, the standard deviation is .0135, and the largest return in absolute value is $-.228$, which is 16.9 standard deviations from the mean. If $z \sim N(0, 1)$ then $P(z \leq -16.9) = 2.25 \times 10^{-64}$. The inverse of this probability is 4.44×10^{63} . Dividing by 260 trading days in the year yields 1.71×10^{61} years. The estimated age of the universe is 1.2×10^{10} years. Chapter 8 takes up Bayesian specification analysis, which is the systematic and constructive assessment of this sort of incongruence of a model with reality.] This, of course, makes explicit the fact that we are uncertain about more than just the unknown parameters μ and σ^2 in (1.3). In fact we are also uncertain about the functional form of the distribution, and our notion of “probability” in (1.1) should account for this, too.

Section 1.4 introduces an alternative to (1.3), which is developed in detail in Section 7.3. An important variant on the value at risk problem arises when a decisionmaker (say, a vice president of an investment bank) selects the value .05, as opposed to some other probability, in (1.1). This integration of behavior with probability is the foundation of Bayesian decision theory, as well as of important parts of modern economics and finance. We shall return to this theme repeatedly, for example, in Sections 2.4 and 4.1.

1.2 OBSERVABLES, UNOBSERVABLES, AND OBJECTS OF INTEREST

A model is a simplified description of reality that is at least potentially useful in decisionmaking. Since models are simplified, they are never literally true; whatever the “data-generating process” may be, it is not the model. Since models are constructed for the purpose of decisionmaking, different decision problems can appropriately lead to different models despite the fact that the reality they simplify is the same. A well-known example is Newtonian physics, which is inadequate when applied to cosmology or subatomic interactions but works quite well in launching satellites and sending people to the moon. In the development of positron emission tomography and other kinds of imaging based on the excitation of subatomic particles, on other hand, quantum mechanics (a different model) functions quite well whereas Newtonian mechanics is inapplicable.

All scientific models have certain features in common. One is that they often reduce an aspect of reality to a few quantitative concepts that are unobservable but organize observables in a way that is useful in decisionmaking. The gravitational constant or the charge of an electron in physics, and the variance of asset returns or the equation of a demand function in the examples in the previous section are all examples of unobservables. Observables can be measured directly; the acceleration of an object when dropped, the accumulation of charge on an electrode, average test scores in different school districts, and sample means of asset returns are all examples.

A model posits certain relationships between observables and unobservables; without these relationships the concepts embodied in the unobservables would be vacuous. A scientific model takes the form “Given the values of the unobservables, the observables will behave in the following way.” The relationship may or may not be deterministic. Thus a model may be cast in the form

$$p(\mathbf{y} \mid \boldsymbol{\theta}),$$

in which $\boldsymbol{\theta}$ is a vector of unobservables and \mathbf{y} is a vector of observables. The unobservables $\boldsymbol{\theta}$ are typically parameters or latent variables. It is important to distinguish between the observables \mathbf{y} , a random vector, and their values after they are observed, which we shall denote \mathbf{y}^o and are commonly called “data.” The functional form of the probability density p gives the model some of its content. In the simple example of Section 1.1.1 the observables might be pairs of student : teacher ratios and test score averages in a sample of school districts, and the unobservables the slope and intercept parameters of a normal linear regression model linking the two. In the simple example of Section 1.1.2, the observables might be asset returns y_1, \dots, y_T , and the unobservable is $\sigma^2 = \text{var}(y_t)$.

The relationship $p(\mathbf{y} \mid \boldsymbol{\theta})$ between observables and unobservables is central, but it is not enough for decisionmaking. The relationship between the gravitational constant g and the acceleration that results when a force is applied to a mass is not enough to deliver a communications satellite into orbit—we had better know quite a lot about the value of g . Likewise, in assessing value at risk using the

simplified model of Section 1.1.2, we must know something about σ^2 . In general, the density $p(\mathbf{y} \mid \boldsymbol{\theta})$ may restrict the behavior of \mathbf{y} regardless of $\boldsymbol{\theta}$ (e.g., when dropped, everyday objects accelerate at rates that differ negligibly with their mass) but for decisionmaking we must know something about $\boldsymbol{\theta}$. (An object will fall about how many meters per second squared at sea level?) A very general way to represent knowledge about $\boldsymbol{\theta}$ is by means of a density $p(\boldsymbol{\theta})$. Formally, we may combine $p(\boldsymbol{\theta})$ and $p(\mathbf{y} \mid \boldsymbol{\theta})$ to produce information about the observables:

$$p(\mathbf{y}) = \int p(\boldsymbol{\theta}) p(\mathbf{y} \mid \boldsymbol{\theta}) d\boldsymbol{\theta}.$$

How we obtain information about $\boldsymbol{\theta}$, and how $p(\boldsymbol{\theta})$ changes in response to new information are two of the central topics of this book. In particular, we shall turn shortly to the question of how information about $\boldsymbol{\theta}$ changes when \mathbf{y} is observed.

In any decision there is typically more than one model at hand that is at least potentially useful. In fact, much of the work of actual decisionmakers lies in sorting through and weighing the implications of different models. To recognize this fact, we shall further index the relation between observables and unobservables by A to denote the model: $p(\mathbf{y} \mid \boldsymbol{\theta})$ becomes $p(\mathbf{y} \mid \boldsymbol{\theta}_A, A)$, and $p(\boldsymbol{\theta})$ becomes $p(\boldsymbol{\theta}_A \mid A)$. The vector of unobservables (in many cases, the parameters of the model A) $\boldsymbol{\theta}_A$ belongs to the set $\Theta_A \subseteq \mathbb{R}^{k_A}$. Alternative models will be denoted A_1, A_2, \dots . Note that the unobservables need not be the same in the models, but the observables $\mathbf{y} \in Y$ are. When several models have the same set of observables, and then we obtain observations (which we call “data”), it becomes possible to discriminate among models. We shall return to this topic in Section 1.5, where we will see that with a bit more effort we can actually use the data to assign probabilities to competing models.

More generally, however, the models relevant to the decision at hand need not all have the same set of observables. A classic example is the work of Friedman (1957) on the marginal propensity to consume. One model (A_1) used aggregate time series data on income and consumption, while another model (A_2) used income and consumption measures for different households at the same point in time. The sets of models addressed the same unobservable—marginal propensity to consume—but reached different conclusions. Friedman’s contribution was to show that the models A_1 and A_2 did, indeed, have different unobservables ($\boldsymbol{\theta}_{A_1}$ and $\boldsymbol{\theta}_{A_2}$), and that the differences in $\boldsymbol{\theta}_{A_1}$ and $\boldsymbol{\theta}_{A_2}$ were consistent with a third, more appropriate, concept of marginal propensity to consume. We shall denote the object of interest on which decisionmaking depends, and which all models relevant to the decision have something to say, by the vector $\boldsymbol{\omega}$. We shall denote the implications of model A for $\boldsymbol{\omega}$ by $p(\boldsymbol{\omega} \mid \mathbf{y}, \boldsymbol{\theta}_A, A)$. The models at hand *must* specify this density; if they do not, then they are not pertinent to the decision at hand.

We can apply this idea to the two examples in the previous section. In the case of the class size decision, $\boldsymbol{\omega}$ might be a $q \times 1$ vector of average test scores conditional on q alternative decisions that might be made about class size. In the case of value at risk, $\boldsymbol{\omega}$ might be a 5×1 vector, the value of the portfolio at the end of each of the next 5 business days.

In summary, we have identified three components of a *complete model*, A , involving unobservables (often parameters) θ_A , observables \mathbf{y} , and a vector of interest ω :

$$p(\theta_A | A), \quad (1.4)$$

$$p(\mathbf{y} | \theta_A, A), \quad (1.5)$$

$$p(\omega | \mathbf{y}, \theta_A, A). \quad (1.6)$$

The ordering of (1.4)–(1.6) emphasizes the fact that the model A specifies the joint distribution

$$p(\theta_A, \mathbf{y}, \omega | A) = p(\theta_A | A)p(\mathbf{y} | \theta_A, A)p(\omega | \mathbf{y}, \theta_A, A). \quad (1.7)$$

It is precisely this joint distribution that makes it possible to use data to inform decisions in an internally consistent manner, and—with more structure to be introduced in Section 1.6—addresses the question of which decision would be optimal.

Exercise 1.2.1 Conditional Probability. A test for the presence of a disease can be administered by a nurse. A result “positive” (+) indicates disease present; a result “negative” (−) indicates disease absent. However, the test is not perfect. The *sensitivity* of the test is the probability of a “positive” result conditional on the disease being present; it is .98. The *specificity* of the test is the probability of a “negative” result conditional on the disease being absent; it is .90. The *incidence* of the disease is the probability that the disease is present in a randomly selected individual; it is .005.

Denoting specificity by p , sensitivity by q , incidence by π , and test outcome by + or −, develop an expression for the probability of disease conditional on a “positive” outcome and one for the probability of disease conditional on a “negative” outcome, if the test is administered to a randomly selected individual. Evaluate these expressions using the values given above.

Exercise 1.2.2 Non-Bayesian Statistics. Suppose the model A is $y \sim N(\mu, 1)$, $\mu \geq 0$, and the sample consists of a single observation $y = y^o$.

- (a) Show that $S = (\max(y - 1.96, 0), \max(y + 1.96, 0))$ is a 95% classical confidence interval for μ , that is, $P(\mu \in S | \mu, A) = .95$.
- (b) Show that if $y^o = -2.0$ is observed, then the 95% classical confidence interval is the empty set.

Exercise 1.2.3 Ex Ante and Ex Post Tests. Let y have a uniform distribution on the interval $(\theta, \theta + 1)$, and suppose that it is desired to test the null hypothesis $H_0 : \theta = 0$ versus the alternative hypothesis $H_1 : \theta = 0.9$ (which are the only two values of θ that are possible). A single observation x is available. Consider the test that rejects H_0 if $y \geq 0.95$, and accepts H_0 otherwise.

- (a) Calculate the probabilities of type I and type II errors for this test.

- (b) Explain why it does not make common sense, for decisionmaking purposes, to accept mechanically the outcome of this test when the observed y^o lies in the interval $(0.9, 1.0)$.

1.3 CONDITIONING AND UPDATING

Because a complete model provides a joint density $p(\theta_A, \mathbf{y}, \omega | A)$, it is in principle possible to address the entire range of possible marginal and conditional distributions involving the unobservables, observables, and vector of interest. Let \mathbf{y}^o denote the actual value of the observable—the data, “ \mathbf{y} observed.” Then with the data in hand, the relevant probability density for a decision based on the model A is $p(\omega | \mathbf{y}^o, A)$. This is the single most important principle in Bayesian inference in support of decisionmaking. The principle, however, subsumes a great many details taken up in subsequent chapters.

It is useful to break up the process of obtaining $p(\omega | \mathbf{y}^o, A)$ into a number of steps, and to introduce some more terminology. The distribution corresponding to the density $p(\theta_A | A)$ is usually known as the *prior distribution* and that corresponding to $p(\mathbf{y} | \theta_A, A)$, as the *observables distribution*. The distribution of the unobservable θ_A , conditional on the observed \mathbf{y}^o , has density

$$\begin{aligned} p(\theta_A | \mathbf{y}^o, A) &= \frac{p(\theta_A, \mathbf{y}^o | A)}{p(\mathbf{y}^o | A)} = \frac{p(\theta_A | A)p(\mathbf{y}^o | \theta_A, A)}{p(\mathbf{y}^o | A)} \\ &\propto p(\theta_A | A)p(\mathbf{y}^o | \theta_A, A). \end{aligned} \quad (1.8)$$

Expression (1.8) is usually called the *posterior density* of the unobservable θ_A . The corresponding distribution is the *posterior distribution*.

The distinction between the prior and posterior distributions of θ_A is not quite as tidy as this widely used notation and terminology suggests, however. To see this, define $\mathbf{Y}'_t = (\mathbf{y}'_1, \dots, \mathbf{y}'_t)$, for $t = 0, \dots, T$ with the understanding that $\mathbf{Y}_0 = \{\emptyset\}$, and consider the decomposition of the probability density of the observables $\mathbf{y} = \mathbf{Y}_T$:

$$p(\mathbf{y} | \theta_A, A) = \prod_{t=1}^T p(\mathbf{y}_t | \mathbf{Y}_{t-1}, \theta_A, A). \quad (1.9)$$

In fact, densities of observables are usually constructed in exactly this way, because when there is dependence between observations, a recursive model is typically the natural representation.

Suppose that $\mathbf{Y}'_t = (\mathbf{y}'_1, \dots, \mathbf{y}'_t)$ is available but $(\mathbf{y}'_{t+1}, \dots, \mathbf{y}'_T)$ is not. (If “ t ” denotes time, then we are between periods t and $t + 1$). Then

$$\begin{aligned} p(\theta_A | \mathbf{Y}'_t, A) &\propto p(\theta_A | A)p(\mathbf{Y}'_t | \theta_A, A) \\ &= p(\theta_A | A) \prod_{s=1}^t p(\mathbf{y}'_s | \mathbf{Y}'_{s-1}, \theta_A, A). \end{aligned}$$

When \mathbf{y}_{t+1}^o becomes available, then

$$\begin{aligned} p(\boldsymbol{\theta}_A | \mathbf{Y}_{t+1}^o, A) &\propto p(\boldsymbol{\theta}_A | A) \prod_{s=1}^{t+1} p(\mathbf{y}_s^o | \mathbf{Y}_{s-1}^o, \boldsymbol{\theta}_A, A) \\ &\propto p(\boldsymbol{\theta}_A | \mathbf{Y}_t^o, A) p(\mathbf{y}_{t+1}^o | \mathbf{Y}_t^o, \boldsymbol{\theta}_A, A). \end{aligned} \quad (1.10)$$

The change in the distribution of $\boldsymbol{\theta}_A$ brought about by the introduction of \mathbf{y}_{t+1}^o , made clear in (1.10), is usually known as *Bayesian updating*. Comparing (1.10) with (1.8), note that $p(\boldsymbol{\theta}_A | \mathbf{Y}_t^o, A)$ plays the same role in (1.10) as does the prior density $p(\boldsymbol{\theta}_A | A)$ in (1.8), and that $p(\mathbf{y}_{t+1}^o | \mathbf{Y}_t^o, \boldsymbol{\theta}_A, A)$ plays the same role in (1.10) as does $p(\mathbf{y}^o | \boldsymbol{\theta}_A, A)$ in (1.8). Indeed, from the perspective of what happens at “time” $t + 1$, $p(\boldsymbol{\theta}_A | \mathbf{Y}_t^o, A)$ is the prior density of $\boldsymbol{\theta}_A$, and $p(\boldsymbol{\theta}_A | \mathbf{Y}_{t+1}^o, A)$ is the posterior density of $\boldsymbol{\theta}_A$. This emphasizes the fact that “prior” and “posterior” distributions (or densities, or moments, or other properties of unobservables) are always with respect to an incremental information set. In (1.8) this information is the entire data set $\mathbf{y}^o = \mathbf{Y}_T^o$, whereas in (1.10) it is \mathbf{y}_{t+1}^o .

From the posterior density (1.8), the density relevant for decisionmaking is

$$p(\boldsymbol{\omega} | \mathbf{y}^o, A) = \int_{\Theta_A} p(\boldsymbol{\theta}_A | \mathbf{y}^o, A) p(\boldsymbol{\omega} | \boldsymbol{\theta}_A, \mathbf{y}^o, A) d\boldsymbol{\theta}_A. \quad (1.11)$$

It is important to acknowledge that we are proceeding in a way that is different from most non-Bayesian statistics, generally termed “classical” statistics. The key difference between Bayesian and non-Bayesian statistics is, in fact, in conditioning. Likelihood-based non-Bayesian statistics conditions on A and $\boldsymbol{\theta}_A$, and compares the implication $p(\mathbf{y} | \boldsymbol{\theta}_A, A)$ with \mathbf{y}^o . This avoids the need for any statement about the prior density $p(\boldsymbol{\theta}_A | A)$, at the cost of conditioning on what is unknown. Bayesian statistics conditions on \mathbf{y}^o , and utilizes the full density $p(\boldsymbol{\theta}_A, \mathbf{y}, \boldsymbol{\omega} | A)$ to build up coherent tools for decisionmaking, but demands specification of $p(\boldsymbol{\theta}_A | A)$.

The strategic advantage of Bayesian statistics stems from the fact that its conditioning is driven by the actual availability of information and by its complete integration with the theory of economic behavior under uncertainty, achieved by Friedman and Savage (1948, 1952). We shall return to this point in Section 1.6 and subsequently in this book.

Two additional matters need to be addressed, as well. The first is that (1.8) and (1.11) are mere formalities as stated; actually representing the densities $p(\boldsymbol{\theta}_A | \mathbf{y}^o, A)$ and $p(\boldsymbol{\omega} | \mathbf{y}^o, A)$ in practical ways for decisionmaking is a technical challenge of high order. Indeed, the principles stated here have been recognized since at least the mid-1950s, but it was not until the application of simulation methods in the 1980s that they began to take on the practical significance that they have today. We return to these developments in Section 1.4 and Chapter 4.

The other matter ignored is explicit attention to multiple models A_1, \dots, A_J . In fact, it is not necessary to confine attention to a single model, and the developments here may be extended to several models simultaneously. We do this in Section 1.5.

Exercise 1.3.1 A Simple Posterior Distribution. Suppose that $y \sim N(\mu, 1)$ and the sample consists of a single observation y^o . Suppose that an investigator has a prior distribution for μ that is uniform on $(0, 4)$.

- (a) Derive the investigator's posterior distribution for μ .
- (b) Suppose that $y^o = -2$. Find an interval (μ_1, μ_2) such that

$$P[\mu \in (\mu_1, \mu_2) \mid y^o] = 0.95.$$

(The answer consists of a pair of real numbers.)

- (c) Do the same for the case $y^o = 1$.
- (d) Are your intervals in (b) and (c) the shortest possible in each case? (You need not use a formal argument. A sketch is enough.)

Exercise 1.3.2 Applied Conditioning and Updating. On a popular, nationally televised game show the guest is shown three doors. Behind one door there is a valuable prize (e.g., a new luxury automobile), and behind the other two doors there are trivial prizes (perhaps a new toaster). The host of the game show knows which prizes are behind which doors. The guest, who cannot see the prizes, chooses one door for the host to open. But before he opens the door selected by the guest, the host always opens one of the two doors not chosen by the guest, and this always reveals a trivial prize. (The guest and the television audience, having watched the show many times, know that this always happens.) The guest is then given the opportunity to change her selected door. After the guest makes her final choice, that door is opened and the guest receives the prize behind her chosen door.

If you were the guest, would you change your door selection when given the opportunity to do so? Would you be indifferent about changing your selection? Defend your answer with a formal probability argument.

Exercise 1.3.3 Prior Distributions. Two graduate students play the following game. An amount of money W is placed in a sealed envelope. An amount $2W$ is placed in another sealed envelope. Student A is given one envelope, and student B is given the other envelope. (The assignment of envelopes is random, and the students do not know which envelope they have received.) Before student A opens his envelope and keeps the money inside, he may exchange envelopes with student B , if B is willing to do this. (At this point, B has not opened her envelope, either; the game is symmetric.) In either case, each student keeps the money in the envelope finally accepted. Both students are rational and risk-neutral; that is, they behave so as to maximize the expected value of the money they keep at the end of the game.

Student A reasons as follows. "There is an unknown amount of money, x , in my envelope. It is just as likely that B 's envelope has $2x$ as it is that it has $x/2$. Conditional on x , my expected gain from switching envelopes is $.5(2x + .5x) - x = .25x$. Since this is positive for all x , I should offer to switch envelopes."

Student B says that the expected gain from switching envelopes is zero.

Explain the fallacy in A 's argument, and provide the details of B 's argument. In each case use the laws of probability carefully.

1.4 SIMULATORS

Decisionmaking requires specific tasks involving posterior distributions. The financial manager in Section 1.1.2 is concerned about the distribution of values of an asset 5 days from now $\omega = p_{T+5} = p_T \exp(\sum_{s=1}^5 y_{T+s})$. She has at hand observations on returns through the present time period, T , of the form $\mathbf{y}^o = (y_1^o, \dots, y_T^o)'$, and is using a model with a parameter vector θ_A . The value at risk she seeks to determine is the number c with the property

$$\int_{-\infty}^{p_T - c} p(\omega \mid \mathbf{y}^o, A) d\omega = 0.05.$$

The manager might recognize that she can decompose this problem into two parts. First, if she knows the value of θ_A —or, more precisely, if the model A specifies the value of θ_A with no uncertainty—then finding c amounts to deriving the inverse cumulative distribution function (cdf) of ω from $p(y_{T+1}, \dots, y_{T+5} \mid \mathbf{y}^o, \theta_A, A)$. This task can be completed analytically for the model (1.3) with known μ and σ^2 , but for realistic models with uncertainty about parameters this is at best tedious and in general impossible.

At this point the financial manager, or one of her staff, might point out that it is relatively easy to simulate most models of financial time series. One such model is the Markov mixture of normals model, discussed in more detail in Section 7.3, in which each y_t is drawn from one of L alternative normal distributions $N(\mu_j, \sigma_j^2)$. Each day t is characterized by an unobserved state variable s_t that assumes one of the values $1, 2, \dots$ or L , and then

$$s_t = j \Rightarrow y_t \sim N(\mu_j, \sigma_j^2). \quad (1.12)$$

The state variables themselves obey a first-order Markov process in which

$$P(s_t = j \mid s_{t-1} = i) = p_{ij}. \quad (1.13)$$

In applications to financial modeling it is reasonable that the values of σ_j^2 vary substantially depending on the state, for example, $\sigma_1^2/\sigma_2^2 \approx 3$, and the state variable is persistent as indicated by $p_{ii} \gg \sum_{j \neq i} p_{ij}$. Such a structure gives rise to episodes of high and low volatility, a feature seen in most financial returns data.

Widely available mathematical applications software makes it easy to simulate this and many other models. Given the current state $s_t = i$, the next period's state is drawn from the distribution (1.13), and then y_{t+1} is drawn from the selected normal distribution in (1.12). Our firm manager can exploit this fact if she knows the parameters of the model and the current state $s_T = j$. She repeatedly simulates

the model forward from the current day T , obtaining in simulation m the returns $y_{T+s}^{(m)}$ ($s = 1, \dots, 5$) and the corresponding simulated asset price 5 days hence, $\omega^{(m)} = p_T^o \exp(\sum_{s=1}^5 y_{T+s}^{(m)})$. At the end she can sort the M simulations of ω , and find a number $c^{(M)}$ such that 5% of the draws are below and 95% are above $p_T^o - c^{(M)}$. It turns out that $c^{(M)} \xrightarrow{\text{a.s.}} c$ as M increases.

This solves only part of the manager's problem. The model, in fact, has many unobservables, not only the unknown parameters μ_j , σ_j^2 and p_{ij} but also the states s_t . Together they constitute the unobservables vector θ_A in this model. The simulation just described requires all of the parameters and the current state s_T . Noting that

$$p(\omega | \mathbf{y}^o, A) = \int_{\Theta_A} p(\omega | \mathbf{y}^o, \theta_A, A) p(\theta_A | \mathbf{y}^o, A) d\theta_A, \quad (1.14)$$

the manager might well recognize that if she could simulate

$$\theta_A^{(m)} \sim p(\theta_A | \mathbf{y}^o, A) \quad (1.15)$$

and next apply the algorithm just described to draw

$$\omega^{(m)} \sim p(\omega | \mathbf{y}^o, \theta_A^{(m)}, A), \quad (1.16)$$

then the distribution of $\omega^{(m)}$ would be that corresponding to the density (1.14).

This strategy is valid, but producing the draws in (1.15) is much more challenging than was developing the algorithm behind (1.16). The latter simulation was relatively easy because it corresponds to the recursion in the natural expression of the model; recall (1.4)–(1.6). Given θ_A , the model tells us how \mathbf{y}_1 , then \mathbf{y}_2 , and so on, are produced, and as a consequence simulating into the future is typically straightforward. The distribution (1.15), on the other hand, asks us to reverse this process: given that a set of observables was produced by the model A , with prior distribution $p(\theta_A | A)$ and observables distribution $p(\mathbf{y} | \theta_A, A)$, make drawings from the distribution with posterior density $p(\theta_A | \mathbf{y}^o, A)$. The formal definition (1.8) is not much help in this task.

This impasse is typical if we attempt to use simulation to unravel the actual distribution corresponding to $p(\omega | \mathbf{y}^o, A)$ in a useful way. Until the late 1980s this problem had succumbed to solution in only a few simple cases, and these did not go very far beyond the even smaller set of cases that could be solved analytically from start to finish. Geweke (1989a) pointed out that importance sampling methods described in Hammersly and Handscomb (1964) could be used together with standard optimization methods to simulate $\theta_A^{(m)} \sim p(\theta_A | \mathbf{y}^o, A)$. The following year Gelfand and Smith (1990) published their discovery that methods then being used in image reconstruction could be adapted to construct a Markov chain G such that if

$$\theta_A^{(m)} \sim p(\theta_A | \theta_A^{(m-1)}, \mathbf{y}^o, G)$$

then $\theta_A^{(m)} \xrightarrow{d} p(\theta_A | \mathbf{y}^o, A)$. This work in short order burgeoned into an even more general set of procedures, known as *Markov chain Monte Carlo* (MCMC), which achieves the same result for almost any complete model. Section 7.3 shows how to apply these methods to the Markov mixture of normals model used in this example.

All of these methods, including importance sampling, produce what are known as *posterior simulators*. These algorithms make it practical to address quantitative decisionmaking problems, using a rich variety of models. Posterior simulators are the focus of Chapter 4.

1.5 MODELING

To this point we have taken the complete model (1.4)–(1.6) as given. In fact, the investigator begins with much less. Typically the vector of interest ω is specified (at least implicitly) by the client making the decision. The composition of the observables vector is sometimes obvious, but in general the question of which observables are best used to inform quantitative decisionmaking is itself an important, interesting, and sometimes difficult question.

This leaves almost all of (1.4)–(1.6) to be specified by the investigator. There is, of course, no algorithm mapping reality into models. The ability to isolate the important features of an actual decision problem, and organize them into a model that is workable and brings to bear all the important features of the decision is an acquired and well-rewarded skill. However this process does involve some specific technical steps that themselves can be cast as intermediate decision problems addressed by the investigator.

One such step is to incorporate competing models A_1, A_2, \dots, A_J in the process of inference and decisionmaking. In Section 1.2 we constructed a joint probability distribution for the unobservables θ_A , the observables \mathbf{y} , and the vector of interest ω , in the context of model A . Suppose that we have done that for each of models A_1, \dots, A_J and that the vector of observables is the same for each of these models. Then we have

$$p(\theta_{A_j} | A_j), \quad p(\mathbf{y} | \theta_{A_j}, A_j), \quad p(\omega | \theta_{A_j}, \mathbf{y}, A_j) \quad (j = 1, \dots, J).$$

If we now provide a prior probability $p(A_j)$ for each model, with $\sum_{j=1}^J p(A_j) = 1$, there is a complete probability distribution over models, unobservables, observables, and the vector of interest. Let $A = \bigcup_{j=1}^J A_j$. In each model the density (1.14), built up from (1.8) and (1.6), provides $p(\omega | \mathbf{y}^o, A_j)$. Then

$$p(\omega | \mathbf{y}, A) = \sum_{j=1}^J p(\omega | \mathbf{y}, A_j) p(A_j | \mathbf{y}, A). \quad (1.17)$$

The posterior density of ω is given by (1.17) with the data \mathbf{y}^o replacing the observable \mathbf{y} . It is a weighted average of the posterior densities of ω in the various models;

indeed, (1.17) is sometimes called *model averaging*. The weights are

$$p(A_j | \mathbf{y}^o, A) = \frac{p(A_j)p(\mathbf{y}^o | A_j)}{p(\mathbf{y}^o | A)} = \frac{p(A_j)p(\mathbf{y}^o | A_j)}{\sum_{j=1}^J p(A_j)p(\mathbf{y}^o | A_j)}. \quad (1.18)$$

The data therefore affect the weights by means of

$$\begin{aligned} p(\mathbf{y}^o | A_j) &= \int_{\Theta_{A_j}} p(\boldsymbol{\theta}_{A_j}, \mathbf{y}^o | A_j) d\boldsymbol{\theta}_{A_j} \\ &= \int_{\Theta_{A_j}} p(\boldsymbol{\theta}_{A_j} | A_j) p(\mathbf{y}^o | \boldsymbol{\theta}_{A_j}, A_j) d\boldsymbol{\theta}_{A_j}. \end{aligned} \quad (1.19)$$

The number $p(\mathbf{y}^o | A_j)$ is known as the *marginal likelihood* of model A_j . The technical obstacles to the computation, or approximation, of $p(\mathbf{y}^o | A_j)$ are at least as severe as those for simulating $\boldsymbol{\theta}_A$, but rapid progress on this problem was made during the 1990s, and this is becoming an increasingly routine procedure.

For any pair of models (A_i, A_j) , we obtain

$$\frac{p(A_i | \mathbf{y}^o)}{p(A_j | \mathbf{y}^o)} = \frac{p(A_i)}{p(A_j)} \cdot \frac{p(\mathbf{y}^o | A_i)}{p(\mathbf{y}^o | A_j)}. \quad (1.20)$$

Note that the ratio is independent of the composition of the full complement of models in A . It is therefore a useful summary of the evidence in the data \mathbf{y}^o about the relative posterior probabilities of the two models. The left side of (1.20) is known as the *posterior odds ratio*, and it is decomposed on the right side into the product of the *prior odds ratio* and the *Bayes factor*. Expressions (1.17) and (1.18) imply that providing the marginal likelihood of a model is quite useful for the subsequent work, including decisionmaking, with several models.

Expression (1.19) for the marginal likelihood makes plain that the bearing of a model on decisionmaking—its weight in the model averaging process (1.17)—depends on the prior density $p(\boldsymbol{\theta}_{A_i} | A_i)$ as well as the observables density $p(\mathbf{y} | \boldsymbol{\theta}_{A_i}, A_i)$. In particular, a model A_i may be an excellent representation of the data in the sense that for some value(s) of $\boldsymbol{\theta}_{A_i}$, $p(\mathbf{y}^o | \boldsymbol{\theta}_{A_i}, A_i)$ is large relative to the best fit $p(\mathbf{y}^o | \boldsymbol{\theta}_{A_j}, A_j)$ in other models, but if $p(\boldsymbol{\theta}_{A_i} | A_i)$ places low (even zero) probability on those values, then the posterior odds ratio (1.20) may run heavily against model A_i .

The investigator's problem in specifying $p(\boldsymbol{\theta}_{A_i} | A_i)$ is no more (or less) difficult than that of designing the observables density $p(\mathbf{y} | \boldsymbol{\theta}_{A_i}, A_i)$. The two are inseparable: $p(\boldsymbol{\theta}_{A_i} | A_i)$ has no implications for observables without $p(\mathbf{y} | \boldsymbol{\theta}_{A_i}, A_i)$, and $p(\mathbf{y} | \boldsymbol{\theta}_{A_i}, A_i)$ says little about $p(\mathbf{y} | A_i)$ until we have $p(\boldsymbol{\theta}_{A_i} | A_i)$ in hand. The first two components of any complete model, (1.4) and (1.5), combined with some relatively simple simulation, can help in these steps of the investigator's problem. Suppose that one or more aspects of the observables \mathbf{y} , which we can represent quite generally as $g(\mathbf{y})$, are thought to be important aspects of reality bearing on a decision, that therefore

should be well represented by the model. In the case of our financial decisionmaker from Section 1.1.2, one concern might focus on the model's stance on "crashes" in the value of financial assets like the one day return of worse than -20% experienced during October 1987, for many assets; then $g(\mathbf{y}) = 1$ if \mathbf{y} exhibits such a day and $g(\mathbf{y}) = 0$ if not. For any specified prior and observables densities, it is generally straightforward to simulate

$$\boldsymbol{\theta}_A^{(m)} \sim p(\boldsymbol{\theta}_A | A), \quad \mathbf{y}^{(m)} \sim p(\mathbf{y} | \boldsymbol{\theta}_A^{(m)}, A)$$

and then construct $g(\mathbf{y}^{(m)})$. The resulting $g(\mathbf{y}^{(m)})$ ($m = 1, \dots$) is an independent identically distributed (i.i.d.) sample from $p[g(\mathbf{y}) | A]$.

This process enables the investigator to understand key properties of a model A before undertaking the more demanding task of developing a posterior simulator $\boldsymbol{\theta}_A^{(m)} \sim p(\boldsymbol{\theta}_A | \mathbf{y}^o, A)$. It provides guidance in choosing the prior density $p(\boldsymbol{\theta}_A | A)$ corresponding to $p(\mathbf{y} | \boldsymbol{\theta}_A, A)$, and can reveal that an observables density $p(\mathbf{y} | \boldsymbol{\theta}_A, A)$ fails to capture important aspects of reality no matter what the value of $\boldsymbol{\theta}_A$. These tasks are all part of what is generally referred to as "model specification" in econometrics. We shall return to them in detail in Chapter 8.

1.6 DECISIONMAKING

The key property of the vector of interest $\boldsymbol{\omega}$ is that it mediates aspects of reality that are relevant for the decision that motivates the econometric or statistical modeling in the first place. To illustrate this point, return again to the decision of school administrators about the class sizes described in Section 1.1.1. School administrators prefer certain outcomes to others; for example, it is quite likely that they prefer high test scores and small teaching budgets to low test scores and large expenditures for teachers' salaries. Suppose, for sake of simplicity, that the teaching budget can be controlled with certainty by hiring more or fewer teachers. In Bayesian decision theory such a decision is known as an *action*, and represented generically by a vector \mathbf{a} . The vector of interest $\boldsymbol{\omega}$ includes all the uncertain factors that matter to administrators in evaluating the outcome; it could be a single summary of test scores, or it might disaggregate to measure test outcomes for different groups of students. The expected utility paradigm, associated with von Neumann and Morgenstern (1944), states that decisions are made so as to maximize the expected value of a *utility function* $U(\mathbf{a}, \boldsymbol{\omega})$ defined over all possible outcomes and decisions. The term "utility" is universal in economics, whereas in Bayesian decision theory the concept of "loss" prevails; the loss function $L(\mathbf{a}, \boldsymbol{\omega})$ is used in place of the utility function $U(\mathbf{a}, \boldsymbol{\omega})$. The only distinction is that the decisionmaker seeks to minimize, not maximize, $E[L(\mathbf{a}, \boldsymbol{\omega})]$. We can always take $L(\mathbf{a}, \boldsymbol{\omega}) = -U(\mathbf{a}, \boldsymbol{\omega})$.

This paradigm fits naturally into the relationship between the model A with parameter vector $\boldsymbol{\theta}_A$, the observable vector \mathbf{y} , and the vector of interest $\boldsymbol{\omega}$. Expression (1.11) provides the distribution relevant to the decision in the use of a single

model A —that is, the distribution relevant for the expectation $E[L(\mathbf{a}, \boldsymbol{\omega})]$, which therefore may be written

$$\begin{aligned} E[L(\mathbf{a}, \boldsymbol{\omega}) \mid \mathbf{y}^o, A] &= \int_{\Theta_A} L(\mathbf{a}, \boldsymbol{\omega}) p(\boldsymbol{\omega} \mid \mathbf{y}^o, A) d\boldsymbol{\omega} \\ &= \int_{\Omega} \int_{\Theta_A} L(\mathbf{a}, \boldsymbol{\omega}) p(\boldsymbol{\theta}_A \mid \mathbf{y}^o, A) p(\boldsymbol{\omega} \mid \boldsymbol{\theta}_A, \mathbf{y}^o, A) d\boldsymbol{\theta}_A d\boldsymbol{\omega}. \end{aligned}$$

Section 1.4 outlines how, in principle, we might obtain drawings $\boldsymbol{\omega}^{(m)}$ from (1.11). Typically those drawings can be used to solve the formal decision problem. In the simplest case, there are only two possible actions ($a = 0$, $a = 1$), and the drawings $\boldsymbol{\omega}^{(m)}$ are i.i.d. Then, so long as $E[L(0, \boldsymbol{\omega})]$ and $E[L(1, \boldsymbol{\omega})]$ both exist—a requirement for the expected utility paradigm to be applicable—the strong law of large numbers implies

$$M^{-1} \sum_{m=1}^M L(a, \boldsymbol{\omega}^{(m)}) \xrightarrow{\text{a.s.}} E[L(a, \boldsymbol{\omega}) \mid \mathbf{y}^o, A]$$

for $a = 0$ and $a = 1$. More generally, if \mathbf{a} is continuous and $E[L(\mathbf{a}, \boldsymbol{\omega})]$ is twice differentiable, then typically

$$M^{-1} \sum_{m=1}^M \partial L(\mathbf{a}, \boldsymbol{\omega}^{(m)}) / \partial \mathbf{a} \xrightarrow{\text{a.s.}} \partial E[L(\mathbf{a}, \boldsymbol{\omega}) \mid \mathbf{y}^o, A] / \partial \mathbf{a}$$

and this feature may be exploited to solve for the value $\mathbf{a} = \hat{\mathbf{a}}$ that minimizes expected loss, using a steepest-descent algorithm. More often, the draws $\boldsymbol{\omega}^{(m)}$ from (1.11) are serially dependent, but this complication turns out not to be essential. We revisit these issues at the level of technical detail required for their application subsequently in Chapter 4. This formalization of the decisionmaking process can be extended to the case of several competing models, using the setup developed in Sections 2.6 and 8.2.

Decisionmaking plays, or should play, an important role in modeling and inference. It focuses attention, first, on the vector of interest $\boldsymbol{\omega}$ that is relevant to the decision problem—namely, the unobservables that will ultimately drive the subjective evaluation of the decision ex post. Given $\boldsymbol{\omega}$, we may then consider the observables \mathbf{y} that are most likely to be useful in providing information about $\boldsymbol{\omega}$ before the decision is made. The observables then govern consideration of the relevant models A , their vectors of unobservables $\boldsymbol{\theta}_A$, and the associated prior densities $p(\boldsymbol{\theta}_A \mid A)$. Note that this amounts to stepping backward through the marginal–conditional decomposition (1.7), a process that is often informal.

In practice, formal decisionmaking is most useful for the structure that it places on the research endeavor from start to finish. Rarely, if ever, do decisionmakers

think and talk about decisions entirely and explicitly within the formal framework we have laid out here. However, the discipline of formal decision theory combined with Bayesian inference can, when well executed, earn the respect of real decisionmakers, and therefore a “seat at the table.” In many ways, this is the ultimate goal of Bayesian inference, and achieving it is a high reward to applied econometrics and statistics.

