Cambridge University Press
978-0-521-76454-4 — Bayesian Decision Analysis
Jim Q. Smith
Excerpt
More Information

Part I

Foundations of Decision Modelling

1
Introduction

1.0.1 Prerequisites and notation

This book will assume that the reader has a familiarity with an undergraduate mathematical course covering discrete probability theory and a first statistics course including the study of inference for continuous random variables. I will also assume a knowledge of basic mathematical proof and notation.

All observable random variables, that is all random variables whose values could at some point in the future be discovered, will be denoted by an upper case Roman letter (e.g. X) and its corresponding value by a lower case letter (e.g. x). In Bayesian inference parameters – which are usually not directly observable – are also random variables. I will use the common abuse of notation here and denote both the random variable and its value by a lower case Greek letter (e.g. θ). This is not ideal but will allow me to reserve the upper case Greek symbols (e.g. Θ) for the range of values a parameter can take. All vectors will be row vectors and denoted by bold symbols and matrices by upper case Roman symbols. I will use $=$ to symbolise a deduced equality and denote that a new quantity or variable is being defined as equal to something via the symbol $\triangleq$.

1.0.2 Bayesian decision analysis and the scope of this book

This book is about Bayesian decision analysis. Bayesian decision analysis seriously intersects with Bayesian inference but the two disciplines are distinct. A Bayesian inferential model represents the structure of a domain and its uncertainties in terms of a single probability model. In a well-built Bayesian model logical argument, science, expert judgements and evidence – for example given in terms of well-designed experiments and surveys – are all used to support this probability distribution. In their most theoretical forms these probability models simply purport to *explain observed scientific phenomena or social behaviour*. In their more applied settings it is envisaged that the analysis can be structured as a probabilistic expert system for possible used in the support decision processes whose precise details are currently unknown to the experts designing the system.

In contrast a Bayesian decision analysis is focused on solving a *given* problem or class of problems. It is of course important for a decision maker (DM) to take due regard of the expert judgements, current science and respected theories and evidence that might be summarised within a probabilistic expert system. However she needs to *apply* such domain knowledge to the actual problem she faces. She will usually only need to use a small subset of the expert information available. She therefore needs not only to draw on that small subset of the expert information that is *relevant* to her problem at hand – augmenting and complementing this as necessary with other context-specific information but also to use this probabilistic information to help her make *the best decision she can* on the basis of the information available to her. When modelling for inference it is not unusual to conclude that there is not enough information to construct a model. But this will not usually be an option for a DM. She will normally have to make do with whatever information she does have and work with this in an intelligent way to make the best decision she can in the circumstances.

The Bayesian decision analyses described in this book provide a framework that:

(1) is based on a formalism accommodating beliefs and preferences as these impact on the decision making process in a *logical* way;
(2) *draws together* sometimes diverse sources of evidence, generally acknowledged facts, underlying best science and the different objectives relevant to the analysis into a single coherent description of her given problem;
(3) provides a description that *explains* to a third party the reasons behind the judgements about the efficacy and limitations of the candidate decisions available so that these judgements can be understood, discussed and appraised;
(4) provides a framework where *conflict* of evidence and conflict of objectives can be expressed and managed appropriately.

The extent to which the foundations of Bayesian decision analysis has been explained, examined and criticised is unparalleled amongst its competitors. As stated in Edwards (2000); French *et al.* (2009) there is an enormous literature on this topic and it would be simply impossible in a single text to do justice to this. However the level of scrutiny it has attracted over the last 90 years has not only refined its application but also defined its domain of applicability. In Chapters 3, 4 and 6 I will review and develop some of this background material justifying the encoding of problems so that uncertainties are coded probabilistically and decisions are chosen to maximise expected utility.

I have therefore severely limited the scope of this book and addressed only a subset of settings and problems. This will allow me not only to present what I consider to be core material in a logical way but also to outline some important technical material in which I have a particular interest. The scope is outlined below.

(1) I will only discuss the arguments for and against a *probabilistic* framework for decision modelling. Furthermore, for practical reasons I will argue throughout the book, for a decision analysis the probabilistic reasoning assumed here is necessarily *subjective*.
(2) I consider only classes of decision problem where a single or group decision maker (DM) must find a *single agreed rationale* for her stated beliefs and preferences and it is this DM who is *responsible* for and has the *authority* to enact the decisions made. The DM will often take advice from experts to inform her beliefs. However if she admits an expert's judgement she adopts it as her own and is responsible for the judgements expressed in her decision model. Similarly

whilst acknowledging, as appropriate, the needs and aspirations of other stakeholders in the expression of her preferences, the DM will take responsibility for the propriety of any such necessary accommodation.

(3) The DM has the time and will to engage in building the type of *logical* and coherent framework that gives an *honest* representation of her problem. The model will support decision making concerning the current problem at hand in the first instance. However there will often be the promise that many aspects of the architecture and some of the expert judgements embodied in the model will be *relevant to analogous future problems* she might face.

(4) The DM is responsible for explaining the rationale behind her choice of decision in a compelling way to an *auditor*. This auditor, for example, may be an external regulator, a line manager or strategy team, a stakeholder, the DM herself or some combination of these characters. In this book we will assume that the auditor's role is to judge the plausibility of the DM's reasoning in the light of the evidence and the propriety of the scale and scope of her objectives.

(5) It is acknowledged by all players that the decision model is likely to have a limited shelf life and is intrinsically provisional. The DM simply strives to present an honest representation of her problem as she sees it at the *current* time. All accept that in the future her judgements may change in the light of new science, new surprising information and new imperatives and may later be adjusted or even discarded for its current or future analogous application.

The limited scope of this book allows us to identify various players in this process. There is the DM herself whose role is given above. There is an analyst who will support her in developing a decision model that can fulfil the tasks above as adequately as possible. There are domain experts to help her evaluate the potential effects on the objects of their expertise an enacted decision might have. Different experts may advise on different aspects of the DM's problem, but for simplicity we will assume that there is just one expert informing each domain of expertise. Throughout we will assume that the advice given by an expert will be no less refined than a probability forecast of what he believes will happen as a result of particular actions the DM might take.

Recent advances in Bayesian methodology have made it possible to support decision making in complex but highly structured domains, rich in expert judgements and informative but diverse experimental and survey evidence; see for example Cowell *et al.* (1999); Pearl (2000). Explanations and illustrations of how these advances can be implemented are presented in the second half of the book. The practical implementation of such decision modelling has its challenges. The analyst needs to guide the DM to first structure her problem by decomposing it into smaller components. Each component in the decomposition can then be linked to possibly different sources of information. The Bayesian formalism can then be used to recompose these components into a coherent description of the problem at hand. This process will be explained and illustrated throughout this book.

There are now many such qualitative frameworks developed and currently being developed, each useful for addressing a certain specific genre of problems. Perforce in this book I have had to choose a small subset if these frameworks that I have found particularly practically useful in a wide set of domains I have faced. These are the event/decision tree – discussed in Chapter 2 – the Bayesian network – discussed in Chapter 7 – and the influence diagram and causal Bayesian network discussed in Chapter 8.

In most moderate or large-scale decision making, the DM not only needs to discover good decisions and policies but also has to be able to provide reasons for her choice. The more compelling she can make this explanation the more likely it will be that she will not be inhibited in making the choices she intends to make. If her foundational rationale is accepted – and for the Bayesian one expounded below this is increasingly the case – she usually still has to convince a third party that the judgements, beliefs and objectives articulated through her decision model are appropriate to the problem she faces.

The frameworks for the decomposition of a problem discussed above are helpful in this regard because – being qualitative in nature – the judgements they embody are more likely to be shared by others. Furthermore they enable the DM to draw on any available evidence from statistical experiment and sample surveys, commonly acknowledged as being well conducted, to support as many quantitative statements she makes and use this to embellish and improve her probabilistic judgements. This draws us into an exploration of where Bayesian inference and Bayesian decision analysis intersect. In Chapter 5 we review some simple Bayesian analyses that inform the types of decision modelling discussed in this book. In Chapter 9 we discuss this issue further with respect to larger problems where significant decomposition is necessary.

One difficulty the DM faces when trying to combine evidence from different sources is when these pieces of evidence seem to give very different pictures of what is happening. When should the DM simply act as if aggregating the information and when should she choose a decision more supported by one source than another? Conflict can also arise when a problem has two competing objectives where all decisions open to the DM score well in one objective but not the other or only score moderately in both. When should the DM choose the latter type of policy and compromise and when should she concentrate in attaining high scores in just one objective? The Bayesian paradigm embodies the answers to these questions. Throughout the book I will show how various types of conflict within a given framework are being automatically managed and explained within the Bayesian methodology in the classes of problem I address.

1.0.3 The development of statistics and decision analysis

It is useful to appreciate why there has been such a growth in Bayesian methods in recent years. Some 35 years ago data-rich structures were only just beginning to be analysed using Bayesian methods. At that time inference still focused on deductions from data from a single (often designed) experiments. The influence of the physical sciences on philosophical reasoning – often through the social sciences which were striving to become more "objective" – was dominant and the complexity of inferential techniques was bounded by computational constraints. Bayesian modelling was not fashionable for a number of reasons:

(1) If decision making was to be objective then the Bayesian paradigm – based on subjective prior distributions and preferences represented via a utility function – was a poor starting point.
(2) Many of the top theoretical statisticians focused on problem formulations based on the physical and health sciences. This naturally led to the study of distributions of estimators from single

experiments that were well designed, likelihood ratio theory, simple estimation, analysis of risk functions and asymptotic inference for large data sets where distributions could be well approximated. Many foundational statistics courses in the UK still have this emphasis. In such problems where data could often be plausibly assumed to be randomly drawn from a sample distribution lying in a know parametrised family it was natural to focus inference on the development of protocols which remotely instructed the experimenter about how to draw inference over different classes of independent and structurally similar experiments. Here the obvious framework for inference was one which built on the properties of different tests and estimators which gave outputs that could be shared by any auditor. The framework of Bayesian inference, with its reliance on contextual prior information, seemed overly complicated and not particularly suited to this task.

(3) The development of stochastic numerical techniques was in its infancy. So for most large-scale problems, asymptotics were necessary. The common claim was that even if you were convinced that a Bayesian analysis should be applied in an ideal world the computations you would need to make were impossible to enact. You would therefore need to rely on large sample asymptotics to actually perform inferences. But these were exactly the conditions where frequentist approaches usually worked as well and more simply than their Bayesian analogues.

The environment had changed radically by the 21st century. In a post-modern era it is much more acceptable to acknowledge the role of the observer in the study of real processes. This acknowledgement is not just common in universities. Many outside academia now accept that a decision model *needs* to have a subjective component to be a valid framework for an inference: at least in an operational setting. Therefore when implementing an inferential paradigm for decision modelling the argument is moving away from the question of *whether* subjective elements should be introduced into decision processes on to *how* it is most appropriate to perform this task. The fact that Bayesian decision theory has attempted to answer this question over the last 90 years has made it a much more established, tested and familiar framework than its competitors. Standard Bayesian inference and decision analysis is now an operational reality in a wide range of applications, whereas alternative theories – for example those based on belief functions or fuzzy logic – whilst often providing more flexible representations – are less well developed. When looking for a subjective methodology which can systematically incorporate expert judgements and preferences the obvious prime candidate to try out first is currently the Bayesian framework.

Secondly the dominant types of decision problems have begun to shift away from small-scale repeating processes to larger-scale one-off modelling and high-dimensional business and phenomenological applications. For example in one of the examples in this book we were required to develop a decision support system for emergency protocols after the accidental release of radioactivity from a nuclear power plant. Here models of the functionality and architecture of a given nuclear plant needed to be interfaced with physical models describing the atmospheric pollutant, the deposition of radioactive waste, its passage into the food chain and into the respiratory system of humans and models of the medical consequences of different types of human behaviour. The planning of countermeasures has to take account not only of health risks and costs but also of political implications. In this type of scenario, data is sparse and often observational and not from designed experiments. Furthermore direct data-based information about many important features of the problem is simply not available. So expert judgements *have* to be elicited for at least some components of the problem. Note that to address such decision problems using a framework which embeds the plant in a sample space of similar plants appears bizarre. In particular the DM is typically concerned about the probability and extent of a *given* population adversely affected by the incident at a *given* nuclear plant, not features of the distribution of sample space of similar such plants: often the given plant and the possible emergency scenario is unique! A Bayesian analysis directly addresses the obvious issue of concern.

Thirdly the culture in which inference is applied is changing. Concurrently it is not uncommon for policy and decision making to be driven by stakeholder meetings where preferences are actively elicited from the DM body and need to be accommodated into any protocol. The necessity for a statistical model to address issues contained in the subjectivity of stakeholder preferences embeds naturally into a subjective inferential framework. Moreover businesses – especially those private companies taking over previously publicly owned utilities – now need to produce documented inferences supporting future expenditure plans. The company needs to give rational arguments incorporating expert judgements and appropriate objectives that will appear plausible and acceptable to an inferential auditor or regulator. Here again subjectivity plays an important role. The most obvious way for a company to address this need is to produce a probabilistic model of their predictions of expenditure based as far as possible on physical, structural and economics certainties, but supplemented by annotated probabilistic expert judgements where no such certainty is possible. The auditor can then scrutinise this annotated probability model and make her own judgements as to whether she believes the explanations about the process and expert judgement are credible. Note here that the auditor cannot be expected to discover whether the company's presentation is precisely true in some objective sense, but only whether what she is shown appears to be a credible working hypothesis and consistent with known facts. In the jargon of frequentist statistics by following Bayesian methods the company tries to produce a single plausible (massive) probability distribution that forms a simple null hypothesis which an auditor can then test in any way she sees fit!

Fourthly computational developments for the implementation of Bayesian methodologies have been dramatic over the last 30 years. We are now at a stage where even for straightforward modelling problems the Bayesian can usually perform her calculations more easily than the non-Bayesian. Routine but flexible analyses can now be performed using free software such as Winbugs or R and Bayesian methodology is often now taught using Bayesian methods (see e.g. Gelman and Hill (2007); Lancaster (2004)). The analysis of high-dimensional problems has been led by Bayesians using sophisticated theory developed together with probabilists to enable the approximation of posterior distributions in an enormous variety of previously intractable scenarios, provided they have enough time. The environment is now capable of supporting models for many commonly occurring multifaceted contexts and for providing the tools for calculating approximate optimal policies. So the Bayesian modeller can now implement her trade to support decision analyses that really matter.

1.1 Getting started

A decision analysis of the type discussed in this book needs to be customised. A decision analysis often begins by finding provisional answers to the following questions:

(1) What is the broad specification of the problem faced and its context? How might a decision analysis help?
(2) Who is the DM – with the authority to enact and responsible for the efficacy of any chosen policy?
(3) Who will scrutinise the DM's performance? In particular who will audit her assessment of the structure and uncertain features of her problem (sometimes of course this might be the DM herself)?
(4) What are the viable options the DM can choose between?
(5) What are the agreed facts and the uncertain features that embody a plausible description of what is happening? In particular what is the science and what are the socially accepted theories that inform the decision process? Is expert advice required on these issues and if so who should be asked?
(6) What are the features associated with the process on which the decision or policy impinges that are uncertain? How and to what extent do these uncertainties impact on the assessed efficacy of a chosen policy? How compelling will these judgements be to the auditor? Who knows about this interface?
(7) How are the intrinsic and uncertain features that determine the efficacy of any given policy related to one another? Who can advise on this? Who judgements can be drawn on?
(8) Where are the sources of information and data that might help reduce uncertainty and support any assertions the DM wants to make to an auditor? How might these sources be supplemented by expedient search or experimentation?

A Bayesian analyst will support the DM by helping her to build her own subjective probability model capturing the nature of uncertainties about features of the model which might affect her decision, helping her to annotate with supporting evidence why she chose this particular model of the underlying process. The analyst will proceed to elicit her utility function which will take due regard of the needs of stakeholders. He will then help the DM in calculating her expected utility associated with each decision viable to her. The best decisions will then be identified as those having the highest expected utility score. These terms will all be formally defined below and the theoretical justification and practical efficacy of following this methodology explored throughout this book.

1.2 A simple framework for decision making

Bayesian decision analysis has developed and been refined over many decades into a powerful and practical tool. However to appreciate some of the main aspects of such analysis it is helpful to begin by discussing the simpler methodologies. So we start by discussing problems where the responsible DM receives a single reward – usually a financial one – as a result of her chosen act. We will later show that these earlier methods were simple special cases of the fully developed theory: the scope for the efficacious use of these simple methods is, from a practical perspective, just rather restrictive. Subsequently in the book these simple techniques will be refined and elaborated to produce a broad platform on which to base a decision analysis for collections of problems of increasing complexity.

Notation 1.1. Let D – called the *decision space* – denote the space of all possible decisions d that could be chosen by the DM and Θ the space of all possible outcomes or *states of nature* θ.

In this simple scenario there is a naive way for a DM to analyse a decision problem systematically to discover good and defensible ways of acting. Before she can identify a good decision she first needs to specify two model descriptors. The first quantifies the consequences of choosing each decision $d \in D$ for each possible outcome $\theta \in \Theta$. The second quantifies her subjective probability distribution over the possible outcomes that might occur.

More specifically the two descriptors needed are:

(1) A *loss function* $L(d, \theta)$ specifying (often in monetary terms) how much she will lose if she makes a decision $d \in D$ and the future outcome is $\theta \in \Theta$. We initially restrict our attention to problems where it is possible to choose Θ big enough so that the possible consequences θ are described in sufficient detail that $L(d, \theta)$ is known by DM for all $d \in D$ and $\theta \in \Theta$. Ideally the values of the function $L(d, \theta)$ for different choices of decision and outcome will be at least plausible to an informed auditor.

(2) A *probability mass function* $p(\theta)$ on $\theta \in \Theta$ giving the probabilities of the different outcomes θ or possible states of nature just before we pick our decision d. If we have based these probabilities on a rational analysis of available data we call this mass function a *posterior mass function*. This probability mass function represents the DM's current uncertainty about the future. This will be her judgement. But if she is not the auditor herself then it will need to be annotated plausibly using facts, science, expert judgements and data summaries.

Note that if the spaces D and Θ are finite of respective dimensions r and n then $p(\theta)$ is a vector of n probabilities, whilst $\{L(d, \theta) : d \in D, \theta \in \Theta\}$ can be specified as an $r \times n$ matrix all of whose components are real numbers. If both $D = \{d_1, d_2, \ldots, d_r\}$ and $\Theta = \{\theta_1, \theta_2, \ldots \theta_n\}$ are finite sets then the losses $\{L(d_i, \theta_j) = l_{ij} : i = 1, 2, \ldots r, j = 1, 2, \ldots n\}$ can be expressed as a table called a *decision table* and shown below.

$$
\begin{array}{ccccccccc}
 & & & & \text{States} & \text{of} & \text{nature} & \\
 & & \theta_1 & \theta_2 & \cdots & \theta_j & \cdots & \theta_n \\
 & d_1 & l_{11} & l_{12} & \cdots & l_{1j} & \cdots & l_{1n} \\
 & d_2 & l_{21} & l_{22} & & l_{2j} & & l_{2n} \\
 & \vdots & & & \ddots & & & \vdots \\
\text{Decisions} & d_i & l_{i1} & l_{i2} & & l_{ij} & & l_{in} \\
 & \vdots & & & & & \ddots & \vdots \\
 & d_r & l_{r1} & l_{r2} & \cdots & l_{rj} & \cdots & l_{rn}
\end{array}
$$

Note that instead of providing a loss function the DM could equivalently provide a *pay-off* $R(d,\theta) = -L(d,\theta)$. In this book we will move freely between these two equivalent representations choosing the one with the most natural interpretation for the problem in question.

One plausible-looking strategy for choosing a good decision is to pick a decision whose associated expected loss to the DM is minimised. This strategy is the basis of one of the oldest methodologies of formal decision making. Because of its simplicity and its transparency to an auditor it is still widely used in some domains. It will be shown later that such a methodology is in fact a particular example of a full Bayesian one. It therefore provides a good starting point from which to discuss the more sophisticated approaches that are usually needed in practice.

Definition 1.2. The *expected monetary value* (EMV) strategy instructs the DM to pick that decision $d^* \in D$ minimising the expectation of her loss [or equivalently, maximising her expected payoff], this expectation being taken using DM's probability mass function over her outcome space Θ.

To follow such a strategy, the DM chooses $d \in D$ so as to minimise the function

$$\overline{L}(d) = \sum_{\theta \in \Theta} L(d,\theta)p(\theta)$$

where $\overline{L}(d)$ denotes her expected loss or, equivalently, maximises

$$\overline{R}(d) = \sum_{\theta \in \Theta} R(d,\theta)p(\theta)$$

where $\overline{R}(d)$ denotes her expected payoff.

Definition 1.3. A decision $d^* \in D$ which minimises $\overline{L}(d)$ (or equivalently maximises $\overline{R}(d)$) is called a *Bayes decision*.

Remark 1.4. As we will see later there are contexts when $p(\theta)$ may be a function of d as well as θ.

Consider first the simplest possible EMV analysis of a medical centre's treatment policies of a mild medical condition which is not painful and where the doctor – our DM – aims to treat patients so as to minimise the treatment cost. Here the centre (or her representative doctor) is the responsible DM. An auditor might be government health service officials. Note that this is a specific example where a cause of interest – here a disease – is observed indirectly through its effects – here a symptom.

Example 1.5. A patient can have one of two illnesses $I = 1, 2$ and is observed to exhibit symptom A or not, $\overline{A}$. Two treatments d_1 and d_2 are possible and the associated costs and the