Risk & Sustainable Management Group

Risk & Uncertainty Program Working Paper: R05#4

Research supported by an Australian Research Council Federation Fellowship
http://www.arc.gov.au/grant_programs/discovery_federation.htm

Consistent Bayesian updating with unconsidered propositions

Simon Grant
Rice University

and

John Quiggin
Australian Research Council Federation Fellow, University of Queensland

Schools of Economics and Political Science
University of Queensland
Brisbane, 4072
rsmg@uq.edu.au
http://www.uq.edu.au/economics/rsmg

The University of Queensland
Australia

Consistent Bayesian updating with unconsidered propositions

John Quiggin and Simon Grant

Paper presented at the 23rd Australian Economic Theory Workshop, Auckland University, Auckland, 23-24 February 2005.

Abstract

In this paper, we employ the propositional approach developed by Grant and Quiggin (2004) and consider the properties of Bayesian updating in the presence of unconsidered propositions.

1 Introduction

Bayesian decision theory and its generalizations provide a powerful set of tools for analyzing problems involving state-contingent uncertainty. In problems of this class, decision-makers begin with a complete specification of uncertainty in terms of a state space (a set of possible states of the world). The ultimate problem is choose between acts, represented as mappings from a the state space to a set of possible outcomes. In many applications, there is an intermediate stage in which the decision-maker may obtain information in the form of a signal about the state of the world, represented by a refinement of the state space. That is, any possible value of the signal means that the true state of the world must lie in some subset of the state space.

In the standard Bayesian model, the decision-maker is endowed with a prior probability distribution over the states of the world. The standard rules of probability enable the derivation of a probability for any given event, and a probability distribution for any random variable. Given the observation of a signal, Bayesian updating involves the use of conditional probabilities to derive a posterior distribution.

It has been widely recognised since the work of Ellsberg (1961) that decision-makers do not always think in terms of well-defined probabilities. Rather, some events may be regarded as being, in some sense, ambiguous. A wide range of definitions of ambiguity and proposals for modelling ambiguous preferences have been put forward. The most influential has been the multiple priors approach of Gilboa and Schmeidler (1989, 1994). In this approach, uncertainty is represented by a convex set of probability distributions. Ambiguity-averse individuals choose to evaluate acts on the basis of the least favorable probability distribution (the maxmin EU model) but a range of other decision criteria are possible.

It is possible to apply Bayesian updating to multiple priors in the obvious fashion, by deriving a posterior distribution for each element of the set However this approach raises a number of issues. First, it is possible in some circumstances to observe a signal that has probability zero for some prior distributions under consideration (this cannot occur in the standard Bayesian setting). More generally, as Halpern notes (2003, p84) an observed signal will, in general be more probable under some priors than others, and is therefore informative about the weight that should be placed on alternative priors. The procedure of updating priors separately makes no use of this information.

A more fundamental difficulty with state-contingent models of decision-making under uncertainty is that decision-makers do not possess a complete state-contingent description of the uncertainty they face. Life is full of surprises, unconsidered possibilities and so on. A number of recent studies, notably those of Modica & Rustichini (1999), Dekel, Lipman and Rusticchini (2001), Halpern (2001), Li (2003) and Heifetz, Meier & Schipper (2004) have attempted to model these problems.

In this paper, we employ the propositional approach developed by Grant and Quiggin (2004) and consider the properties of Bayesian updating in the presence of unconsidered propositions. The key idea is to take the prior distribution over a set of propositions under consideration (which is an exogenous given in the standard Bayesian model) as the conditional distribution derived from a probability distribution over a complete state space, which is not accessible to the decisionmaker. We then consider updating in the light of new information. Consistency requires that the usual Bayesian posterior over the considered propositions be the same as the prior that would be derived from the updated probability distribution over the complete space. We show that consistency is equivalent to an independence property, which justifies treating the considered propositions in isolation.

Next we consider the multiple priors model. In our approach, multiple priors are the conditional distributions derived from different implicit assumptions about unconsidered propositions. For consistent updating of multiple priors, we require not only that the independence property should hold for each prior, but that observations on considered propositions should not be informative with respect to those priors.

These are stringent conditions. However, we show by example, that in many of the standard situations in which Bayesian reasoning is applied, these conditions are reasonable, at least in the sense that they are implicit in standard approaches to such problems.

2 Setup

We consider a single individual decision-maker, denoted i, making choices over time, in a situation of uncertainty regarding the state of nature, and an incomplete description of the set of states of nature. To describe the individual's representation of the world, it is necessary to embed this representation in a more complete description.

2.1 Representation

Let the set of states of the world be Ω. We focus on the representation

$$\Omega = 2^{\mathbf{N}}$$

where $\mathbf{N} = \{1, 2, ...n, ...\}$ is supposed to be a finite or countably infinite set, indexing a family of nodes. Each node represents either an act of nature or a decision by individual i, and is associated with a specific time $t(n)$, and an elementary proposition p^n.

Each elementary proposition p^n is a statement such as 'The winner of the 2008 US Presidential election is Hillary Clinton' or 'Individual i votes for the Republican candidate in 2008'. The negation of p^n is denoted by $\neg p^n$.

At time $t(n)$, the proposition takes the truth value v^n, which will be denoted 1 (True) or -1 (False). The set of time periods is a finite or countably infinite set of the form $\mathbf{T} = 0, 1, ...$.[1]Without loss of generality, we will assume that the elements of $\mathbf{N}$ are ordered so that $n > n' \Rightarrow t(n) \geq t(n')$. Conversely, we may denote by $N(t)$ the subset $N \subseteq \mathbf{N} = \{n : t(n) = t\}$.

An exhaustive description of the state of the world, including the decisions made by individual i, consists of an evaluation of each of the elementary propositions $p^n, n \in N$. From the viewpoint of a fully informed observer, any state of the world can therefore be described by a real number $\omega \in \Omega \subseteq [0, 1]$[2], given by

$$\omega = \sum_{n \in \mathbf{N}} 2^{-(n+1)} (v^n + 1).$$

An elementary proposition p^n is true in state ω if and only if $\omega_n = 1$, where $\omega_n \in \{0, 1\}$ is the nth element in the *binary* expansion of ω. Hence, corresponding to any elementary proposition p^n is an event

$$E_{p^n} = \{\omega : \omega_n = 1\}$$

3 Propositions, histories and events

Now consider the perspective of an external observer at time t, with full knowledge of the state space Ω and of the history up to time t, given by the values v^n : for $N_t = \{n : t(n) \leq t\}$. The history at time t may be numerically represented by

$$h(t) = \sum_{n \in N_t} 2^{-(n+1)} (v^n + 1).$$

The history $h(t)$ may be viewed in three distinct, but equivalent ways. First, as the name implies, it is an element of a sequence $h(1), h(2) ...h(t)$,

[1] For simplicity, we will focus on the case when both $\mathbf{N}$ and $\mathbf{T}$ are finite, but we will not rely on this assumption in any essential fashion.

[2] If some propositions may be true in all states of the world, Ω may be a proper subset of $[0, 1]$. Alternatively, Ω may be set equal to $[0, 1]$ with some states having zero probability in all evaluations.

in which each element incorporates all its predecessors. Second, it is a time-dated event (a subset of the state space) which may be denoted E_t, consisting of all elements beginning with the N_t terms having values v^n.

$$E_t \equiv \left\{\omega : \omega_n = 2^{-(n+1)} \left(v^n + 1\right) \text{ for all } n \in N_t\right\}$$

Third, it is a compound proposition p_t, which may be defined using the $\wedge$ operator, corresponding to logical AND. For pairs of elementary propositions, let $\wedge$ be defined in terms of Ω as

$$\begin{aligned} p^n \wedge p^{n'} &= \{\omega : \omega_n = \omega_{n'} = 1\} \\ &= E_{p^n} \cap E_{p^{n'}} \end{aligned}$$

More generally, for any collection of elementary propositions indexed by $N \subseteq \mathbf{N}$, we define

$$\bigwedge_{\{n \in N\}} p^n = \bigcap_{\{n \in N\}} E_{p^n}$$

The proposition representing the event E_t associated with history $h\left(t\right)$ is then given by

$$p_t = \bigwedge_{\{n \in N_t : v^n = 1\}} p^n \bigwedge_{\{n \in N_t : v^n = -1\}} \neg p^n.$$

More generally, a compound proposition is derived by assigning truth values of 1 or -1 to all p^n where n is a member of some (possibly empty) subset $\mathbf{N}\left(p\right) \subseteq \mathbf{N}$, leaving all p^n , $n \notin \mathbf{N}\left(p\right)$ unconsidered. The set $\mathbf{N}\left(p\right)$ is referred to as the *scope* of p, and is the disjoint union of $\mathbf{N}_-\left(p\right)$, the set of elementary propositions false under p, and $\mathbf{N}_+\left(p\right)$, the set of elementary propositions true under p. The simple proposition p^n has scope $\mathbf{N}\left(p^n\right) = \{n\}$. We define the null proposition $p^{\emptyset}$ such that $p^{\emptyset}_n = 0$, $\forall n$ and do not assign a truth value to $p^{\emptyset}$.

The OR operator is

$$p^n \vee p^{n'} = \{\omega : \omega_n = 1\} \cup \{\omega : \omega_{n'} = 1\}.$$

The class of all propositions in the model is denoted by $\mathbf{P} = \{-1, 0, 1\}^{\mathbf{N}}$. It is useful to consider more general classes of propositions $P \subseteq \mathbf{P}$. To any class of propositions P, given state ω, we assign the truth value

$$t\left(P; \omega\right) = \sup_{p \in P} \left\{t\left(p; \omega\right)\right\}.$$

That is, P is true if any $p \in P$ is true, and false if all $p \in P$ are false. In terms of the logical operations defined below, the set P has the truth value derived by applying to its members the OR operator defined for elementary propositions by

$$p^n \vee p^{n'} = \{\omega : \omega_n = 1\} \cup \{\omega : \omega_{n'} = 1\}$$

and, more generally, by

$$\bigvee_{\{p \in P\}} p = \bigcup_{\{p \in P\}} E_p.$$

The relationship between events, propositions and histories may now be developed further. First, there is a 1-1 correspondence between complete histories, having a value $h(t)$ for each $t \in T$. This generates a natural correspondence between events (subsets of the state space) and collections of histories.

Any (non-null) compound proposition p corresponds, from the external viewpoint, to an event

$$E(p) = \{\omega \in [0,1] : \omega_n = 0, \forall n \in \mathbf{N}_-(p); \omega_n = 1, \forall n \in \mathbf{N}_+(p)\} \subset \Omega.$$

Since distinct compound propositions may be logically equivalent, this correspondence is not 1-1.

4 The decisionmaker's viewpoint

We consider a decision maker who is not aware of all the propositions in $\mathbf{P}$. The class of all propositions considered by individual i at time t is denoted P_t^i. This class must include all the decisions that are to made by decisionmaker i at time $t+1$ [3]and will also include a variety of propositions about the state of the world. As in the Grant and Quiggin (2004), we can define logical operations with respect to this subclass of $\mathbf{P}$.

4.1 Full rationality in a bounded domain

An important special case is one that may be described as 'full rationality in a bounded domain'. In this case, the individual has access to a fixed set of propositions $P^i \subseteq \mathbf{P}$ closed under $\vee$ and $\wedge\cdot$ Without loss of generality, it may be assumed that the set P^i is generated by a set of elementary propositions $N^i \subseteq \mathbf{N}$, that is, $P^i = 2^N$. The individual is unaware of elementary propositions in $N^{-i} = \mathbf{N} - N^i$. Let $P^{-i} = 2^{N^{-i}}$ be the set of propositions generated by elements of N^{-i}.

[3]We do not allow for 'unconscious decisions'.

At any time t, full rationality requires that individual i is aware of the history up to that time, insofar as it determines the truth value of propositions in N^i. That is, full rationality in a bounded domain precludes imperfect recall.

Thus, the event observed by individual i at time t is summed up by the truth value of elementary propositions p^n where $n \in N_t^i = N^i \cap N_t$. We have a corresponding representation of: the history

$$h^i(t) = \sum_{n \in N_t^i} 2^{-(n+1)} (v^n + 1);$$

the associated compound proposition p_t, consisting of the conjunction

$$p_t^i = \bigwedge_{\left\{n \in N_t^i : v^n = 1\right\}} p^n \bigwedge_{\left\{n \in N_t^i : v^n = -1\right\}} \neg p^n;$$

and the corresponding event E_t^i.

Given full rationality in a bounded domain, any proposition p may be written from the external viewpoint as $p^i \wedge p^{-i}$ where $p^i \in P^i, p^{-i} \in P^{-i}$. In particular, the proposition p_t characterizing the time t event E_t may be written as

$$p_t = p_t^i \wedge p_t^{-i}.$$

5 Probabilities

Suppose that we are given a measure μ on Ω, which may be taken to represent the prior beliefs that would be held by the decisionmaker in the absence of bounds on rationality, including bounds on the set of propositions under consideration. Given such a measure, the structure of the state space derived above is sufficient to give a complete characterization of Bayesian updating. For any event E, we have, at time t, a derived measure

$$\mu_t(E) = \frac{\mu(E \cap E_t)}{\mu(E_t)}.$$

Note that we can now derive $\mu_{t+1}(E)$ in two ways. First, we can repeat the definition above, yielding

$$\mu_{t+1}(E) = \frac{\mu(E \cap E_{t+1})}{\mu(E_{t+1})}.$$

Alternatively, we can apply Bayesian updating to μ_t, yielding

$$\mu_{t+1}^{t}(E) = \frac{\mu_t(E \cap E_{t+1})}{\mu_t(E_{t+1})}.$$

Now we observe that, since $E \cap E_{t+1} \subseteq E_{t+1} \subseteq E_t$,

$$\begin{aligned} \mu_t(E_{t+1}) &= \frac{\mu(E_{t+1})}{\mu(E_t)} \\ \mu_t(E \cap E_{t+1}) &= \frac{\mu(E \cap E_{t+1})}{\mu(E_t)} \end{aligned}$$

and hence,

$$\begin{aligned} \mu_{t+1}^{t}(E) &= \left(\frac{\mu(E \cap E_{t+1})}{\mu(E_t)}\right) \Big/ \left(\frac{\mu(E_{t+1})}{\mu(E_t)}\right) \\ &= \mu_{t+1}(E). \end{aligned}$$

6 Restricted Bayesianism

Given full rationality on a bounded domain, it is natural to consider μ_t^i, the restriction of the probability measure μ_t to events $E(p)$ where $p \in N^i$. That is,

$$\mu_t^i(E(p)) = \mu_t(E(p))$$

if and only if $p \in N^i$.

We now have two potential ways of deriving μ_{t+1}^i, given the observation of E_{t+1}. We can use the restriction procedure at time $t+1$ instead of t, obtaining μ_{t+1}^i as the restriction of μ_{t+1} to P^i. Alternatively, we can apply Bayesian updating directly to μ_t^i using the information obtained from E_{t+1}^i. The first approach yields, for any $p \in P^i$,

$$\begin{aligned} \mu_{t+1}^i(E(p)) &= \mu_{t+1}(E(p)) \\ &= \frac{\mu_t(E(p) \cap E_{t+1})}{\mu_t(E_{t+1})} \\ &= \frac{\mu(E(p) \cap E_{t+1})}{\mu(E_{t+1})} \end{aligned}$$

where, as shown in the previous section, the last step works because $E(p) \cap E_{t+1} \subseteq E_{t+1} \subseteq E_t$ so $\mu(E_{t+1}) = \mu_t(E_{t+1})\,\mu(E_t)$

The second approach yields

$$\begin{aligned}\mu_{t+1}^{i,t}\left(E\left(p\right)\right) &= \frac{\mu_t^i\left(E\left(p\right)\cap E_{t+1}^i\right)}{\mu_t^i\left(E_{t+1}^i\right)} \\ &= \frac{\mu\left(E\left(p\right)\cap E_{t+1}^i\right)}{\mu\left(E_{t+1}^i\right)}\end{aligned}$$

where the second step follows from the definition of μ_t. We say that *restricted Bayesian updating is consistent if,* for all i,t,p

$$\mu_{t+1}^{i,t}\left(E\left(p\right)\right)=\mu_{t+1}^{i}\left(E\left(p\right)\right).$$

Suppose that, for all p,p' such that $p\in P^i, p'\in P^{-i}$

$$\mu\left(E\left(p\wedge p'\right)\right)=\mu\left(E\left(p\right)\right)\mu\left(E\left(p'\right)\right).$$

That is, the probabilities of propositions in the restricted domain for i are independent of the probabilities of unconsidered propositions. It seems reasonable to suppose that restricted Bayesian updating will be consistent under these conditions. We now show that this is the case.

Proposition 1 *Restricted Bayesian updating is consistent if and only if for all p,p' such that $p\in P^i, p'\in P^{-i}$, and all possible histories $h\mu\left(E\left(p\wedge p'\right)\right)=\mu\left(E\left(p\right)\right)\mu\left(E\left(p'\right)\right).$*

Proof: *Suppose the condition holds. Then, for all t,*

$$\mu\left(E\left(p_t\right)\right)=\mu\left(E\left(p_t^i\right)\right)\mu\left(E\left(p_t^{-i}\right)\right).$$

In particular,

$$\mu\left(E\left(p_{t+1}\right)\right)=\mu\left(E\left(p_{t+1}^i\right)\right)\mu\left(E\left(\left(p_{t+1}^{-i}\right)\right)\right),$$

and, for $p\in P^i$,

$$\mu\left(E\left(p\right)\cap E_{t+1}\right)=\mu\left(E\left(p\right)\cap E\left(p_{t+1}^i\right)\right)\mu\left(E\left(\left(p_{t+1}^{-i}\right)\right)\right),$$

so,

$$\begin{aligned}\mu_{t+1}^{i}\left(p\right) &= \frac{\mu\left(E\left(p\right)\cap E_{t+1}\right)}{\mu\left(E_{t+1}\right)} \\ &= \frac{\mu\left(E\left(p\right)\cap E\left(p_{t+1}^i\right)\right)\mu\left(E\left(\left(p_{t+1}^{-i}\right)\right)\right)}{\mu\left(E\left(p_{t+1}^i\right)\right)\mu\left(E\left(\left(p_{t+1}^{-i}\right)\right)\right)} \\ &= \frac{\mu\left(E\left(p\right)\cap E\left(p_{t+1}^i\right)\right)}{\mu\left(E\left(p_{t+1}^i\right)\right)} \\ &= \frac{\mu\left(E\left(p\right)\cap E_{t+1}^i\right)}{\mu\left(E_{t+1}^i\right)} \\ &= \mu_{t+1}^{i,t}\left(p\right).\end{aligned}$$

On the other hand, the condition can be false only if there exist p, h, t such that

$$\mu_{t+1}^{i,t}(p) \neq \mu_{t+1}^{i}(p)$$

that is,

$$\frac{\mu_t^i\left(E(p) \cap E_{t+1}^i\right)}{\mu_t^i\left(E_{t+1}^i\right)} \neq \frac{\mu\left(E(p) \cap E_{t+1}^i\right)}{\mu\left(E_{t+1}^i\right)}$$

or

$$\frac{\mu_t^i\left(E(p) \cap E\left(p_{t+1}^i\right)\right)}{\mu_t^i\left(E\left(p_{t+1}^i\right)\right)} \neq \frac{\mu\left(E(p) \cap E\left(p_{t+1}^i\right)\right)}{\mu\left(E\left(p_{t+1}^i\right)\right)}$$

Now since

$$p_{t+1} = p_{t+1}^i \wedge p_{t+1}^{-i}$$

we have

$$\frac{\mu_t^i\left(E\left(p \wedge p_{t+1}^i \wedge p_{t+1}^{-i}\right)\right)}{\mu_t^i\left(E\left(p_{t+1}^i\right)\right)} \neq \frac{\mu\left(E\left(\left(p \wedge p_{t+1}^i \wedge p_{t+1}^{-i}\right)\right)\right)}{\mu\left(E\left(p_{t+1}^i\right)\right)}$$

which implies

$$\mu\left(E\left(\left(p \wedge p_{t+1}^i\right) \wedge p_{t+1}^{-i}\right)\right) \neq \mu\left(E\left(\left(p \wedge p_{t+1}^i\right)\right)\right) \mu\left(E\left(p_{t+1}^{-i}\right)\right)$$

6.1 Example

Consider a state space of the form $\Omega = S_1 \times S_2$ with a product measure $\mu = \mu_1\mu_2$. Suppose that variables of potential interest are x measurable with respect to S_1, z measurable with respect to S_2 and

$$y = \beta x + z,$$

where β is a (possibly unknown) parameter. Under the product measure assumption, which implies independence of x and z, restricted Bayesian updating is consistent for either S_1 or S_2 and therefore for any propositions about the values of x or about the values of z. Conversely, if x and z are not independent, restricted Bayesian updating will not apply. We consider the case when a set of propositions about x, sufficient to fully characterize the distribution of x, is considered,[4] and z represents unconsidered possible events.

Now, what about y? We have, in the absence of any hypotheses about z,

$$y = \alpha + \beta x + \varepsilon,$$

[4] For example, if it is known that x is normally distributed, a set of propositions about the mean and variance of x is required.

where $\alpha = E[z]$, $\varepsilon = z - \alpha$. Under plausible conditions, the decisionmaker may be able to make the implicit assumption $E[z] = 0$ without any detailed knowledge of γ and z. For example, suppose that x is gross domestic product and y is gross national product for some unspecified country. The difference between x and y is determined by international income flows, which necessarily sum to zero for the world as a whole. So, a decisionmaker could reasonably use the model

$$y = x + \varepsilon$$

while having no knowledge of factors determining ε. More generally, even when the value of $E[z]$ is unknown and unconsidered, a decisionmaker might have reasonable knowledge about $\beta = \partial y / \partial x$, so that Bayesian updating applied to x may be useful even when y is the variable of interest.

6.1.1 An observation on decision theory

Allthough we have not formally considered applications to decision theory, it is easy to see that for appropriately linear choice problems, restricted Bayesianism applied to x yields optimal decisions (on the assumption that no information is available about y or z). Assume for simplicity that the unconditional expectation of z is equal to zero, and consider the problem of choosing θ to maximise $E[v]$ where

$$v(\theta, y) = \theta y - c(\theta),$$

and c is a convex cost function. We have

$$\begin{aligned} E[v] &= \theta E[y] - c(\theta) \\ &= \theta \beta E[x] - c(\theta). \end{aligned}$$

Hence, the optimal policy, given an estimate of $E[x]$, is to choose θ such that

$$c'(\theta) = \beta E[x],$$

and the best estimate of $E[x]$ may be obtained by restricted Bayesian updating of the unconditional prior distribution μ_1.

In this context, then, it seems reasonable to suggest that an optimal policy could be achieved solely by considering hypotheses about β and v.

7 Multiple priors

Thus far, we have considered cases where the prior distribution μ_t on the restricted domain generated by the considered propositions is induced by a unique prior μ on the full state space, incorporating implicit probabilities for unconsidered propositions. To generate a multiple priors model, it is natural to suppose that there may be more than one such measure. An obvious way to do this is to look at the measures induced conditional on the possible truth values for one or more unconsidered propositions.

Considering any $p' \in P^{-i}$, there are two induced measures on P^i, namely

$$\mu^+ (E(p)) = \frac{\mu (E(p \wedge p'))}{\mu (E(p'))}, p \in P^i$$

and

$$\mu^- (E(p)) = \frac{\mu (E(p \wedge \neg p'))}{\mu (E(\neg p'))}, p \in P^i.$$

For a proposition p' that is independent of P^i in the sense that, for all $p \in P^i$

$$\mu (E(p \wedge p')) = \mu (E(p)) \mu (E(p')).$$

we have $\mu^+ = \mu^-$ since, for all $p \in P^i$

$$\begin{aligned}
\mu^+ (E(p)) &= \frac{\mu (E(p \wedge p'))}{\mu (E(p'))} = \frac{\mu (E(p)) \mu (E(p'))}{\mu (E(p'))} \\
&= \frac{\mu (E(p)) \mu (E(\neg p'))}{\mu (E(\neg p'))} = \frac{\mu (E(p \wedge \neg p'))}{\mu (E(\neg p'))} = \mu^- E(p)
\end{aligned}$$

In general, however, $\mu^+ E(p) \neq \mu^- E(p)$, and consideration of probability values for p' in the range $[0, 1]$ gives rise to probabilities for p in the range $[\mu^- E(p), \mu^+ E(p)]$. Thus, we can define a set of priors

$$M(p') = \left\{ \lambda \mu^+ + (1 - \lambda \mu^- : 0 \leq \lambda \leq 1 \right\}.$$

The natural interpretation here is that each element of the set of multiple priors may be derived as a conditional probability measure, given a probability number for the unconsidered proposition p'. Thus p' has a status intermediate between propositions in P^i that are under active consideration, and unconsidered propositions in the case of restricted Bayesianism. Although the decision-maker does not explicitly consider p', the range of multiple priors corresponds to the probability measure that would arise if p' were a considered proposition with probability λ.

For a more general version of the multiple priors model, let P^* be a set of unconsidered propositions, closed under $\neg$ and $\wedge$, and let Δ be the unit simplex with dimension equal to $K = card(P^*)$, having typical element $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_K)$ such that $\sum_k \lambda_k = 1$. For each $p_k \in P^*$, we have, as described above, an induced measure on P^i, which will be denoted μ_k and we define the set of priors

$$M(P^*) = \left\{ \sum_k \lambda_k \mu_k : \boldsymbol{\lambda} \in \Delta \right\}.$$

It is easy to check that this definition agrees with that given above for the case $P^* = \{p', \neg p'\}$

7.1 Consistent updating with multiple priors

The definition of consistent updating with a unique measure μ can be extended in a straightforward fashion to the case of a given $\boldsymbol{\lambda} \in M(P^*)$. For each k, i, t we may define $\mu^i_{t,k}$ as above and set

$$\boldsymbol{\lambda}^i_t = \sum_k \lambda_k \mu^i_{t,k}$$

and similarly for $\mu^{i,t}_{t+1,k}$ and $\boldsymbol{\lambda}^{i,t}_{t+1,k.}$ For all i, t, p, k,

$$\mu^{i,t}_{t+1,k}(E(p)) = \mu^i_{t+1,k}(E(p))$$

then for all i, t, p

$$\lambda^{i,t}_{t+1}(E(p)) = \lambda^i_{t+1}(E(p))$$

so that consistent Bayesian updating for each μ_k is sufficient to ensure consistent Bayesian updating for all $\boldsymbol{\lambda} \in M(P^*)$. Necessity is trivial.

Now, by the definition of μ_k, we obtain: [5]

Proposition 2 *Consistent Bayesian updating for all $\boldsymbol{\lambda} \in M(P^*)$ holds if and only if, for any p, p', p'' such that $p \in P^i$, $p' \in P^*, p'' \in P^{-i}$,*

$$\mu(p \wedge p' \wedge p'') = \mu(p \wedge p') \wedge \mu(p'')$$

[5] By additivity, the condition includes the special cases

$$\mu(p \wedge p'') = \mu(p) \wedge \mu(p'')$$

and

$$\mu(p' \wedge p'') = \mu(p') \wedge \mu(p'').$$

For the first choose $p \in P^i$, $p' \in P^*, p'' \in P^{-i}$,, apply the condition first with p, p', p'', then with $p, \neg p', p''$ and add. Similarly for the second.

7.2 The evolution of the set of priors

The fact that all elements of $M(P^*)$ may be updated consistently does not fully resolve the problems associated with updating multiple priors. Halpern (2003) makes the point that deriving the Bayesian posterior for each prior separately is problematic, given that, in general, observed signals will have different likelihood for each of the different priors. It is natural to consider whether there are conditions under which this problem will not arise, and, if so, what are the implications for the structure of knowledge.

Examining the consistency condition above, we observe that there is no problem with propositions $, p'' \in P^{-i},$. By additivity, the condition includes the special cases

$$\mu(p \wedge p'') = \mu(p) \wedge \mu(p'')$$

and

$$\mu(p' \wedge p'') = \mu(p') \wedge \mu(p'')$$

For the first choose $p \in P^i$, $p' \in P^*, p'' \in P^{-i},$, apply the condition first with p, p', p'', then with $p, \neg p', p''$ and add. Similarly for the second.

On the other hand, by hypothesis, the elements of P^* yield distinct conditional probabilities for members of P^i.Hence, we cannot expect a criterion that will apply for all possible histories and time periods.

Suppose instead, we require that for all elements $p \in P^i \cap N(t)$, $p' \in P^*$

$$\mu(p \wedge p') = \mu(p) \wedge \mu(p')$$

Then considered propositions for which the truth value is observed at time t are independent of all elements of P^*, and therefore do not affect the weighting that might be given to different priors.

Under the conditions discussed above, neither considered nor unconsidered propositions provide any information about P^* at time t. Hence, if the set of multiple priors is to evolve, it can only do so if the truth value of some $p' \in P^*$ is revealed at time t, that is, if $P^* \cap N(t)$ is nonempty.

8 Example

Suppose

$$y = \beta x + z + w$$

and that it is implicitly known that $E[w] = 0$ and that x, w and z are independent.

Now suppose there exist one or more alternative hypotheses about the value of z, each of which induces a prior distribution on y for given x. Then if we receive information on x we can update each of the conditional priors on y in a consistent fashion.

As an example suppose that x is gross domestic product y is national income, z is depreciation and w is international transfers. Given a sequence of observations on x, the decision-maker may act as a restricted Bayesian with respect to x and employ multiple priors with respect to y, each of which corresponds implicitly to an alternative hypothesis about z.

Observe in addition that there may exist unconsidered propositions p that are uninformative wrt any of w, x, y and z. Trivially, observation of the truth value of such propositions does not imply any change in the restricted Bayesian posterior, nor in the induced prior.

9 Concluding comments

In formulating more general representations of choice under uncertainty, it is highly desirable to show that, under appropriate conditions, existing representations can be derived as special cases. These conditions are often restrictive. Nevertheless, it is often the case that they may be satisfied exactly or as a reasonable approximation.

The Bayesian approach to decision theory is powerful and appealing. However, the assumption, necessary for the model to be applied, that the decision-maker possesses an exhaustive description of all possible states of the world, with an associated probability distribution, is obviously unrealistic. In this paper, we have derived necessary and sufficient conditions for the consistency of Bayesian updating when applied to a restricted set of propositions, a subset of an exhaustive propositional description of the world.

The necessary conditions are highly restrictive, suggesting that, in many cases, a multiple priors model may be more realistic. For this case also, conditions have been derived under which each element of a set of multiple priors may be updated consistently.

In general, neither of these sets of conditions may be satisfied. Typically, learning is not simply a matter of updating priors but involves new discoveries, imaginative conjectures, abandonment of previously maintained hypotheses and so on. A central task for decision theory, partially addressed by Grant and Quiggin (2004) is to develop models of these processes.

10 References

Dekel, E., Lipman, B. and Rusticchini, A. (2001), 'A unique subjective state space for unforeseen contingencies', Econometrica, 69, 891-934.

Ellsberg, D. (1961), 'Risk, ambiguity and the Savage axioms', Quarterly Journal of Economics, 75(4), 643–69.

Gilboa, I. and Schmeidler, D. (1989), 'Maxmin expected utility with non-unique prior', Journal of Mathematical Economics, 18(2), 141–53.

Gilboa, I. and Schmeidler, D. (1994), 'Updating ambiguous beliefs', Journal of Economic Theory, 59, 33–49.

Grant, S. and Quiggin, J. (2004), 'Unforeseen contingencies: a propositional approach', Paper presented at FUR Conference, Paris.

Halpern, J. (2003) Reasoning About Uncertainty, The MIT Press,