THE ESSENCE
of
“BAYESIAN THEORY”
written by
J. M. BERNARDO and A. F. M. SMITH

Jiro Ihara

Jan 1995-

Chapter 1

Introduction

Summary

A brief historical introduction to Bayes' theorem and its author is given, as a prelude to a statement of the perspective adopted in this volume regarding Bayesian Statistics. An overview is provided of the material to be covered in successive chapters and appendices, and a Bayesian reading list is provided.

Chapter 2

Foundations

Summary

The **concept of rationality** is explored in the context of representing beliefs or choosing actions in situations of uncertainty. An axiomatic basis, with intuitive operational appeal, is introduced for the foundations of decision theory.

The **dual** concepts of probability and utility are formaly defined and analyzed with this context.

The criterion of maximizing expected utility is shown to be the only decision criterion whcih is compartible with the axiom system.

The analysis of **sequential decision problems** is shown to reduce to successive applications of the methdology introduced.

Statistical inference is viewded as a **particular decision problem** which may be analyzed within the farmework of decision theroy.

The **logarithmic score** is established as the **natural utility function** to describe the preferences of an individual faced with a **pure inference problem**. Within this framework, the concept of discrepancy between parobability distributions and the quantification of the amount of information in new data are naturally defined in terms of expected loss and expected increase in utility, respectively.

2.1 DECISION PROBLEMS

Definition 2.1 (Decison problem)-P.18-

A decison problem is defined by the elements (**E, C, A**,$\leq$), where:

(i) **E** is an **algebra** of **relevant** events, $E_j(j \in J)$;

(ii) **C** is a set of **possible** consequences, $c_j(j \in J)$;

(iii) **A** is a set of options, or potential actions, consisting of functions which map **finite** partitions of Ω, **the certain** event in **E**, to compatibly-dimensioned, ordered sets of elements of **C** (**A** is called the *action space* .);

(iv) $\leq$ is a preference order; taking the form of a binary relation between **some** of the elements of **A**.

Note 1: An option $a \in \mathbf{A}$ consists precisely of linking of a partition of $\Omega, \{E_j, j \in J\}$, with a corresponding set of concequences, $\{c_j, j \in J\}$. To represent such a mapping we shall adopt the notation $a = \{c_j | E_j, j \in J\}$ with the interpretation that the events E_j leads to consequence $c_j, j \in J$.

More precise notation is as follows:

The option is $a_i \in \mathbf{A}$.

The partition is $\{E_{ij}, j \in J_i\}$.

The corresponding set of concequences is $\{c_{ij}, j \in J_i\}$.

$a_i = \{c_{ij} | E_{ij}, \in J_i\}$

With the interpretation that the events E_{ij} leads to consequence $c_{ij}, j \in J_i$.

Note 2: We can identify individual concsequences as special cases of options by writing $c = \{c|\Omega\}$, for any $c \in \mathbf{C}$.

We shall simply regard c as denoting either an element of **C**, or the element $\{c|\Omega\}$ of **A**.

We shall write $c_1 \leq c_2$ iff $\{c_1|\Omega\} \leq \{c_2|\Omega\}$ and say that consequence c_1 is not prefrerrded to consequence c_2.

Note 3: The basic preference relation between options, $\leq$, is conditional on the initial state of information M_0.

Definition 2.2 (Induced binary relations)-P.20-

(i) $a_1 \sim a_2 \Longleftrightarrow a_1 \leq a_2 \; and \; a_2 \leq a_1$.

(ii) $a_1 < a_2 \Longleftrightarrow a_1 \leq a_2$ and it is not true taht $a_2 \leq a_1$.

(iii) $a_1 \geq a_2 \Longleftrightarrow a_2 \leq a_1$.

(iv) $a_1 > a_2 \Longleftrightarrow a_2 < a_1$.

Definition 2.3 (Uncertainty relation)-P.21- -IMPORTANT-

$E \leq F \Longleftrightarrow$ for **all** $c_1 < c_2$, $\{c_2|E, c_1|E^c\} \leq \{c_2|F, c_1|F^c\}$;

we then say that E is **not more likely** than F.

Note 4: Since, for all $c_1 < c_2$,

$c_1 \equiv \{c_2|\Phi, c_1|\Omega\} < \{c_2|\Omega, c_1|\Phi\} \equiv c_2$,

it is always true, as one would expect, that $\Phi < \Omega$.

Note 5: *All* the order relations over $\mathbf{A} \times \mathbf{A}$, and hence over $\mathbf{C} \times \mathbf{C}$ and $\mathbf{E} \times \mathbf{E}$, are to be understood as *personal*, in the sense that , given an agreed structure for a decision problem, each individual is free to express his or her own personal preferences, in the light of his or her initial state of information M_0. Thus, for a given individual, a statement such as $E > F$ is to be interpreted as *" this individual, given the state of information described by M_0, considers event E to be more likely than event F"*. Moreover, Definition 2.3 provides such a statement with an *operational meaning* since for all $c_1 < c_2, E > F$ is equivalent to an agreement to choose option $\{c_2|E, c_1|E^c\}$ in preference to option $\{c_2|F, c_1|F^c\}$.

Definition 2.4 (Conditional preference)-P.22- -IMPORTANT-

For any $G(\in \mathbf{E}) > \Phi$

(i) $a_1 \leq_G a_2 \Longleftrightarrow$ for **all** a, $\{a_1|G, a|G^c\} \leq \{a_2|G, a|G^c\}$;

(ii) $E \leq_G F \Longleftrightarrow$ for $c_1 \leq_G c_2$, $\{c_2|E, c_1|E^c\} \leq_G \{c_2|F, c_1|F^c\}$.

Note 6: The G provides the additional information to M_0.

Note 7: Compare Definition 2.3 with Definition 2.4(ii). The former is defined "*for all* $c_1 < c_2$", but the latter "*for* $c_1 \leq_G c_2$".

See the proof of Proposition 2.14(page 39) of the book. "*for* $c_1 \leq_G c_2$" would have to be replaced with "*for all* $c_1 <_G c_2$".

2.2 COHERENCE AND QUNTFICATION

Our overall objective is developing a rational approach to choosing among options.

Our following assumptions, presented in the form of five *axioms*, can be viewded as responses to the questions: *"what rules should preference relations obey?"* and *"what events should be included in* $\mathbf{E}$*?"*

The axioms are *prescriptive*, not *descriptive.*

The axioms simply prescribe constraints which it seems to us imperative to acknowledge in those situations where an individual aspires to choose among alternatives in such a way as to **avoid** certain forms of **behavioral inconsistency**.

Axiom 1 - Axiom 3 make coherent the preferences. They are called the *coherence axioms.*

Axiom 4 and Axiom 5 may be called the *quntification axioms.*

2.2.1 Coherent Preferences

Axiom 1.(Comparability of consequences and dischotomized options)- P.23-

(i) There exist consequences c_1, c_2 such that $c_1 < c_2$.

(ii) For all consequences c_1, c_2, c_3, c_4 and events E, F,
either $\{c_2|E, c_1|E^c\} \leq \{c_4|F, c_3|F^c\}$
or $\{c_2|E, c_1|E^c\} \geq \{c_4|F, c_3|F^c\}$.

Note 8: Axiom 1(i) makes non-trivial the problems represented within the formal framework.

Note 9: Axiom 1(ii) is to be interpreted in the following way: *"If we aspire to make a rational choice between alternative options, then we must at least be willing to express preferences between simple dictomized options."*

Axioms 2.(Tansitivity of preference)-P.24-

(i) $a \leq a$.

(ii) If $a_1 \leq a_2$ and $a_2 \leq a_3$, then $a_1 \leq a_3$.

Note 10: Axiom 2(i) and Definition 2.2(i) imply $a \sim a$.

Note 11: Axiom 2(ii) is to be understood in the following sense: *"If we aspire to avoid expressing preferences whose behavioral implications are such as to lead us to the certain loss of somthing we value, then we must ensure that our preferences fit together or cohere in a transitive manner."*

Proposition 2.1 (Transitivity of uncertainties)-P.25-

(i) $E \sim E$.

(ii) If $E_1 \leq E_2$ and $E_2 \leq E_3$, then $E_1 \leq E_3$.

Note 12: Proposition 2.1 can be proved by using Definition 2.3 and Axiom 2.

Proposition 2.2 (Derived transitive properties)-P.26-

(i) If $a_1 \sim a_2$ and $a_2 \sim a_3$, then $a_1 \sim a_3$.
If $E_1 \sim E_2$ and $E_2 \sim E_3$, then $E_1 \sim E_3$.

(ii) If $a_1 < a_2$ and $a_2 \sim a_3$, then $a_1 < a_3$.
If $E_1 < E_2$ and $E_2 \sim E_3$, then $E_1 < E_3$.

Note 13: Proposition 2.2 can be proved by using Definition 2.2 and Axiom 2(ii) for the options, and the induced binary relations on events (similar to Definition 2.2) and Proposition 2.1 for the events.

Axiom 3. (Consistency of preferences)-P.26-

(i) If $c_1 \leq c_2$ then, for **all** $G > \Phi, c_1 \leq_G c_2$.

(ii) If, for **some** $c_1 < c_2, \{c_2|E, c_1|E^c\} \leq \{c_2|F, c_1|F^c\}$, then $E \leq F$.

(iii) If, for **some** c and $G > \Phi, \{a_1|G, c|G^c\} \leq \{a_2|G, c|G^c\}$, then $a_1 \leq_G a_2$.

Note 14: Axiom 3(i) formalizes the idea that preferences between pure conseqences should not be affected by the acquisition of further information regarding the uncertain events in **E**.

$c_1 \leq_G c_2 \Leftrightarrow$ for all a, $\{c_1|G, a|G^c\} \leq \{c_2|G, a|G^c\}$ by Definition 2.4(i).

Note 15: Axiom 3(ii) and Axiom 3(iii) ensure that Definitions 2.3 and 2.4(i) have operational content.

The condition of Axiom 3(ii) $\Rightarrow E \leq F \Rightarrow$ for **all** $c_1 < c_2, \{c_2|E, c_1|E^c\} \leq \{c_2|F, c_1|F^c\}$ by Definition 2.3.

The condition of Axiom 3(iii) $\Rightarrow a_1 \leq_G a_2 \Rightarrow$ for **all** a, $\{a_1|G, a|G^c\} \leq \{a_2|G, a|G^c\}$ by Definition 2.4(i).

Note 16 Axiom 3(iii) is a version of the **sure-thing principle**.

Propositon 2.3 (Invariance of preferences between consequences) -P.26-

$c_1 \leq c_2$ iff there exist $G > \Phi$ such that $c_1 \leq_G c_2$.

Note 17: Proposition 2.3 can be proved by using Axiom 3(i), Definition 2.4(i), Axiom 3(ii) and Axiom 1(ii) in this order.

Proposition 2.4 (Monotonicity) -P.27-

If $E \subseteq F$ (E logically implies F) then $E \leq F$.

Note 18: Proposition 2.4 can be proved by using Axiom 3(i) and Definition 2.3. "Definition 2.2" in the proof of the book should be read "Definition 2.3."

Note 19: Proposition 2.4 implies, as one would expect, for any even E, $\Phi \leq E \leq \Omega$.

Definition 2.5 (Significant events) -P.27-

An event E is siginificant given $G > \Phi$ if (any) $c_1 <_G c_2$ implies that $c_1 <_G \{c_2|E, c_1|E^c\} <_G c_2$. If $G = \Omega$, we shall simply say that E is significant.

Note 20: Intuitively, significant events given G are those operationally perceived by the decision-maker as "practically possible but not certain" given the information provided by G.

Proposition 2.5 (Characterization of significant events) -P.27-

An event E is significant given $G > \Phi$, iff $\underline{\Phi < E \cap G} < G$. In particular, E

is significant iff $\underline{\Phi < E} < \Omega$.

Note 21: Proposition 2.5 can be proved by using Dfinitions 2.4(i), 2.5 ,Proposition 2.3, Definition 2.3, and Axiom 3(iii) in this order.

"for all $c_1 \leq_G c_2$" in the proof of the book should be read "for all $c_1 <_G c_2$".

Definition 2.6 (Pairwise independence of events) -P.28-

We say that E and F are (pairwise) independent, denoted by $E \perp F$, iff, for all c, c_1, c_2

(i) $c \bullet \{c_2|E, c_1|E^c\} \Rightarrow c \bullet_F \{c_2|E, c_1|E^c\}$,

(ii) $c \bullet \{c_2|F, c_1|F^c\} \Rightarrow c \bullet_E \{c_2|F, c_1|F^c\}$,

where $\bullet$ is any one of the realtions $<, \sim$ *or* $>$.

2.2.2 Quantification

The *coherent axioms* (Axioms 1 to 3) provide a minimal set of rules to ensure that **qualitative** comparisons based on $\leq$ cannot have intuitively undesirable impications.

This purely qualitative framework is inadequate for serious, systematic comparisons of options.

Precison, through quantification, is achieved by introducing some form of numerical standard into a context already equipped with a coherent qualitative odering relation.

Axiom 4. (Existence of standard events) -P.29-

There exists a subalgebra $\mathbf{S}$ of $\mathbf{E}$ and a function $\mu : \mathbf{S} \to [0,\ 1]$ such that:

(i) $S_1 \leq S_2$, iff, $\mu(S_1) \leq \mu(S_2)$;

(ii) $S_1 \cap S_2 = \Phi$ implies that $\mu(S_1 \cup S_2) = \mu(S_1) + \mu(S_2)$;

(iii) for any number α in $[0,\ 1]$, and events E, F, there is a standard event S such that $\mu(S) = \alpha$, $E \perp S$ and $F \perp S$;

(iv) $S_1 \perp S_2$ implies that $\mu(S_1 \cap S_2) = \mu(S_1)\mu(S_2)$;

(v) if $E \perp S$, $F \perp S$ and $E \perp F$, then $E \sim S \Rightarrow E \sim_F S$.

Note 22: Imagine an idealized roullet. We will refer to **S** as a standard family of events in **E** and will think of **E** as the algebra generated by the relevant events in the decision problem together with the elements of **S**.

It is important to emphasize that we do not require the assumption that standard families of events actually, physically, exist, or could be **precisely** constructed in accordance with conditions (i) to (v). We only require that we **can** invoke such a set up as a mental image.

Proposition 2.6 (Collections of disjoint standard events) -P.31-

For any finite collection $\{\alpha_1, \ldots, \alpha_n\}$ of real numbers such that $\alpha_i > 0$ and $\alpha_1 + \ldots + \alpha_n \leq 1$, there exists a corresponding collection $\{S_1, \ldots, S_n\}$ of disjoint standard events such that $\mu(S_i) = \alpha_i, i = 1, \ldots, n$.

Note 23: Proposition 2.6 can be proved by using Axiom 4(iii,iv,ii)

Axiom 5. (Precise measurement of preferences and uncertainties) -P.31-

(i) If $c_1 \leq c \leq c_2$, there exists a standard event S such that $c \sim \{c_2|S, c_1|S^c\}$.

(ii) For each event E, there exists a standard event S such that $E \sim S$.

2.3 BELIEFS AND PROBABILITIES

2.3.1 Representation of beliefs

The principles of coherence and quantification by comparison with a standard will enable us to give a formal definition of **degree of belief**, thus providing a numerical measure of the uncertainty attached to each event.

The conceptual basis for this numerical measure will be seen to derive from the formal rules govering quantitative, coherent preferences, irrespective of the nature of the uncertain events under consideration.

Proposition 2.7 (Complete comparability of events) -P.33-

Either $E_1 > E_2$, or $E_1 \sim E_2$, or $E_2 > E_1$.

Note 24: Proposition 2.7 can be proved based on Axiom 5(ii), Axiom 4(i) and Proposition 2.2. "Proposition 2.1" in the proof of the book should be read "Proposition 2.2".

Note 25: Proposition 2.7 says that the uncertainty relation induced between events is **complete**, although the order relation $\leq$ between options was not assumed to be complete.

Proposition 2.8 (Additivity of uncertainty relations) -P.34-

If $A \leq B$, $C \leq D$ and $A \cap C = B \cap D = \Phi$, then $A \cup C \leq B \cup D$. Moreover, if $A < B$ or $C < D$, then $A \cup C < B \cup D$.

Note 26: Proposition 2.8 can be proved on the basis of definition2.3, Axiom 3(iii) and defintion 2.4(i).

Definition 2.7 (Measure of degree of belief) -P.34- -IMPORATANT-

Given an uncertainty relation $\leq$, the **probability** $P(E)$ of an event E is

the real number $\mu(S)$ associated with **any** standard event S such that $E \sim S$.

Note 27: For instance, the statement $P(E) = 0.5$ precisely means that E is **judged** to be **equally likely** as a standard event of 'measure' 0.5, maybe a conceptual perfect coin falling heads , or a copmuter generated 'random' integer being an odd number.

Note 28: Probabilities are always **personal degrees of belief**, in that they are **a numerical representation** of the decison-maker's **personal** uncertainty relation $\leq$ between events. **Moreover**, probabilities are alwayes conditional on the information currently available.

It makes no sense that, within the framework we are discussing, to qualify the word probability with adjectives such as "objective", "correct" or "unconditional".

Proposition 2.9 (Existence and uniqueness) -P.35-

Given an uncertainty relation $\leq$, there exists a unique probability $P(E)$ associated with each event.

Note 29: Proposition 2.9 can be proved by using Axiom 5(ii) Proposition 2.2(i) and Axiom 4(i). "Proposition 2.2(ii)" in the proof of the book should be read "Proposition 2.2(i)".

Definition 2.8 (Compatibilitry) -.P.35-

A function f:$\mathbf{E} \rightarrow \mathbf{Re}$ is said to be compatible with an order relation $\leq$ on $\mathbf{E} \times \mathbf{E}$ if, for all events,

$$E \leq F \Longleftrightarrow f(E) \leq f(F)$$

.

Proposition 2.10 (Compatibility of probability and degrees of belief) -P.35-

The probability function $P(.)$ is compatible with the uncertainty realtion $\leq$.

Note 30: Proposition 2.10 can be proved on the basis of Axiom 5(ii), Proposition 2.2(ii), Axiom 4(i) and Definition 2.7.

Proposition 2.11 (Probability structure of degrees of belief) -P.35- FUNDAMENTAL IMPORTANCE

(i) $P(\Phi) = 0$ and $P(\Omega) = 1$.

(ii) If $E \cap F = \Phi$, then $P(E \cup F) = P(E) + P(F)$.

(iii) E is significant iff $0 < P(E) < 1$.

Note 31: Proposition 2.11(i) can be proved by Definnition 2.7, Axiom 4(iii), Proposition 2.4, Proposition 2.10. Proposition 2.11(ii) can be proved by Propositions 2.8,2.6 and 2.7. Proposition 2.11(iii) can be proved by Propositions 2.5 and 2.10.

Corollay. (Finitely additivity of degrees of belief) -P.36-

(i) If $\{E_j, j \in J\}$ is a finite collection of disjoint events, then

$$P\left(\bigcup_{j \in J} E_j\right) = \sum_{j \in j} P(E_j)$$

.

(ii) For any event E, $P(E^c) = 1 - P(E)$.

Note 32: The first part follows by induction from Proposition 2.11(ii); the second part is a special case of (i) since if $\cup_j E_j = \Omega$ then, by Proposition 2.11(i), $\sum_j P(E_j) = 1$.

Note 33: Coherent degrees of belief are probabilities. The mathematical structure is merely serving as a representation of (personal) degrees of belief.

Definition 2.9 (Probability distribution) -P.26-

If $\{E_j, j \in J\}$ form a finite partition of Ω, with $P(E_j) = p_j$, then $p_j, j \in J$ is said to be a probability distribution over the partition.

Proposition 2.12 (Uniqueness of the probability MEASURE) -P.37-

P is the only probability measure **compatible** with the uncertainty realtion $\leq$.

Note 34: Proposition 2.12 can be proved by using Definition 2.8, Proposition 2.6, Axiom 4(ii), Theorem 2.63(Eichhorn, 1978) and Proposition 2.9.

Proposition 2.13 (Characterization of independence) -P.37-

$$E \perp F \Longleftrightarrow P(E \cap F) = P(E)P(F)$$

Note 35: Proposition 2.13 can be proved based on Axiom 4(iii), Axiom 4(v), Definition 24(ii), Definition 2.6, Propositions 2.2, 2.10, and Axiom 4(iv)

Note 36: Proposition 2.13 says that our operational definition of (pairwise) independence of events - Definition 2.6 - is compatible with its more standard, *ad hoc* product definition.

2.3.2 Revison of Beliefs and Bayes' Theorem

Proposition 2.14 (Properties of conditional beliefs) -P.38-

(i) $E \leq_G F \Longleftrightarrow E \cap G \leq F \cap G$.

(ii) If there exist $c_1 < c_2$ such that $\{c_2|E, c_1|E^c\} \leq_G \{c_2|F, c_1|F^c\}$, then $E \leq_G F$.

Note 37: Proposition 2.14 can be proved on the basis of Definition 2.4(ii), Proposition 2.3, Definition 2.4(i), Axiom 3(iii), Definition 2.4(i), Axiom 3(ii) and part(i) of this proposition. "$c_2 \geq c_1$" in the proof of the book should be read "$c_2 > c_1$".

Compare (ii) with Axiom 3(ii).

Definition 2.10 (Conditional measure of degree of belief) -P.39-

Given a conditional uncertainty relation $\leq_G$, $G > \Phi$, the conditional probability $P(E|G)$ of an event E given the **assumed occurence** of G is the real number $\mu(S)$ such that $E \sim_G S$.

Proposition 2.15 (Conditional probability) -P.39-

For any $G > \Phi$,

$$P(E|G) = \frac{P(E \cap F)}{P(G)}.$$

Note 38: Proposition 2.15 can be proved by using Axiom 4(iii), Propositions 2.4, 2.10, 2.13, 2.14 and Definition 2.10.

Note 39: In our formulation, $P(E|G) = P(E \cap G)/P(G)$ is a logical derivation from the axioms, **not** an *ad hoc* definition.

Proposition 2.16 (Compatibility of conditional probability and conditinal degrees of belief) -P.40-

$$E \leq_G F \Longleftrightarrow P(E|G) \leq P(F|G)$$

.

Note 40: Proposition 2.16 can be proved by Propositions 2.14(i), 2.10 and 2.15.

Proposition 2.17 (Probability structure of conditinal degrees of belief) -P.40-

For any given $G > \Phi$,

(i) $P(\Phi|G) = 0 \leq P(E|G) \leq P(\Omega|G) = 1$.

(ii) if $E \cap F \cap G = \Phi$, then $P(E \cup F|G) = P(E|G) + P(F|G)$;

(iii) E is significant given $G \Longleftrightarrow 0 < P(E|G) < 1$.

Note 41: Proposition 2.17 can be proved based on Propositions 2.15, 2.10 and 2.5.

Corollary (Finitely additive structure of conditional degrees of belief) -P.40-

For all $G > \Phi$,

(i) if $\{E_j \cap G_j, j \in J\}$ is a finite collection of disjoint events, then

$$P\left(\bigcup_{j \in J} E_j|G\right) = \sum_{j \in j} P(E_j|G).$$

(ii) for any envent E, $P(E^c|G) = 1$ - $P(E|G)$.

Proposition 2.18 (Uniqueness of the conditinal probability MEASURE) -P.41-

$P(.|G)$ is the only probability measure **compatible** with the conditional uncertainty relation $\leq_G$.

Proposition 2.19 (Bayes' theorem) -P.42-

For any finite partition $\{E_j, j \in J\}$ of Ω and $G > \Phi$,

$$P(E_i|G) = \frac{P(G|E_i)P(E_i)}{\sum_{j\in J} P(G|E_j)P(E_j)}.$$

Note 42: Bayes' theorem is a simple mathematical consequence of the fact that quantiative coherece implies that degrees of belief should obey the rules of probability. From another point of view, it may be also established (Zellner, 1988b) that, under some reasonable desiderata, Bayes's thorem is an optimal information processing system.

Note 43: Proposition 2.19 can be proved by Proposition 2.15, the Corollary to Proposition 2.11.

Definition 2.11 (Prior, posterior, and predictive probabilities) -P.43-

If $\{H_j, j \in J\}$ are exclusive and exhaustive events (hypotheses), then for any event(data) D,

(i) $P(H_j), j \in J$, are called the **prior probabilities** of the $H_j, j \in J$;

(ii) $P(D|H_j), j \in J$, are called the **likelihoods**, of the $H_j, j \in J$, given D;

(iii) $P(H_j|D), j \in J$, are called the **posterior probabilities** of the $H_j, j \in J$;

(iv) $P(D)$ is called the **predictive probability** of D implied by the likelihoods and the prior probabilities.

Note 44: The predictive probabilitiy $P(D)$, logically implied by the likelihoods and the prior probabilities, provides a basis for assessing the compatibility of the data with our beliefs. See Chapter 6.

2.3.3 Conditional Independence

Proposition 2.20 -P.45-

For all $F > \Phi$, $E \perp F \iff P(E|F) = P(E)$.

Note 45: Proposition 2.20 can be proved by Proposition 2.15.

Definition 2.12 (Mutual independence) -P.46-

Events $\{E_j, j \in J\}$ are said to be mutually independent if, for **any** $I \subseteq J$,

$$P\left(\bigcap_{i\in I} E_i\right) = \prod_{i\in I} P(E_i)$$

Definition 2.13 (Conditional independence) -P.47-

The events $\{E_j, j \in J\}$ are said to be conditionally independent given $G > \Phi$ if , for **any** $I \subseteq J$,

$$P\left(\bigcap_{i\in I} E_i|G\right) = \prod_{i\in I} P(E_i|G).$$

For any subalgebra $\mathbf{F}$ of $\mathbf{E}$, the events $\{E_j, j \in J\}$ are said to be conditionally independent given $\mathbf{F}$ iff they are conditinally indpendent given any $G > \Phi$ in $\mathbf{F}$.

2.3.4 Sequential Revision of beliefs

Note 46:

$$P(H_j|D^{(k+1)}) = \frac{P(D_{k+1}|H_j \cap D^{(k)})P(H_j|D^{(k)})}{P(D_{k+1}|D^{(k)})}$$

where $D^{(k)} = D_1 \cap D_2 \cap \ldots \cap D_k$.

Note 47:

In the case of two hypotheses, H_1, H_2, the judgement of conditional independence for $D_1, D_2, \ldots, D_n, \ldots$, given H_1 or H_2,

$$\text{posterior odds} = \text{prior odds} \times \text{likelihood ratio.}$$

2.4 ACTIONS AND UNTILITIES

2.4.1 Bounded Decision Problems

Definition 2.14 (Extreme consequences) -P.49-

The pair of consequences c_* and c^* are called, respectively, the **worst** and the **best** consequences in a decison problem if, for any other consequences $c \in \mathbf{C}$, $c_* \leq c \leq c^*$.

Definition 2.15 (Canonical utility function for consequences) -P.50-

Given a preference relation $\leq$, the utility $u(c) = u(c|c_*, c^*)$ of a consequence c, relative to the extreme consequences $c_* < c^*$, is the real number $\mu(S)$ associated with **any** standard event such that $c \sim \{c^*|S, c_*|S^c\}$. The mapping $u : \mathbf{C} \rightarrow \mathbf{Re}$ is called the **utility function**.

Proposition 2.21 (Existence and uniqueness of bounded utility) -P.50-

For any bounded decision problem $(\mathbf{E}, \mathbf{C}, \mathbf{A}, \leq)$ with extrem consequences $c_* < c^*$,

(i) for all c, $u(c|c_*, c^*)$ exists and is unique;

(ii) the value of $u(c|c_*, c^*)$ is unaffected by the **assumed occurrence** of an event $G > \Phi$;

(iii) $0 = u(c_*|c_*, c^*) \leq u(c|c_*, c^*) \leq u(c^*|c_*, c^*) = 1$.

Note 48: (i) can be proved by Axiom 5(i), Proposition 2.2(i), Axiom 3(iii) and Axiom 4(i). (ii) can be proved Axiom 4(iii), Definition 2.6. (iii) can be proved by Definition 2.15 and Axiom 4(i).

Note 49: Consider a fixed consequence c such that $c \sim \{c^*|S, c_*|S^c\}$. For some $E(c)$, if

$$c \sim \{c^*|E(c), c_*|E^c(c)\}$$

then,

$$\{c^*|S, c_*|S^c\} \sim c \sim \{c^*|E(c), c_*|E^c(c)\}.$$

This means $S \sim E(c)$. Therefore,

$$u(c|c_*, c^*) = \mu(S) = P(E(c)).$$

$P(E(c))$ is the degree of belief in the occurrence of the **best** consequence.

Note 50: Consider fixed events E and E^c associated with c^* and c_*, respectively. For some $c(E)$, if

$$c(E) \sim \{c^*|E, c_*|E^c\}$$

then, we can write

$$c(E) \sim \{c^*|E, c_*|E^c\} \sim \{c^*|S, c_*|S^c\},$$

where S is a standard event such that $E \sim S$.

Therefore,

$$u(c(E)|c_*, c^*) = \mu(S) = P(E).$$

$P(E)$ is the degree of belief in the occurrence of the **best** consequence.

Definition 2.16 (Conditional expected utility) -P.51-

For any $c_* < c^*$, $G > \Phi$ and $a \equiv \{c_j|E_j, j \in J\}$,

$$\overline{u}(a|c_*, c^*, G) = \sum_{j\in J} u(c_j|c_*, c^*)P(E_j|G)$$

is the expected utiliy of the option a, given G, with respect to the extreme consequences c_*, c^*. If $G = \Omega$, we shall simply write $\overline{u}(a|c_*, c^*)$ in place of $\overline{u}(a|c_*, c^*, \Omega)$.

Proposition 2.22 (Decision criterion for a bounded decision problem) -P.52-

For any bounded decision with extreme consequences $c_* < c^*$, and $G > \Phi$,

$$a_1 \leq_G a_2 \Longleftrightarrow \overline{u}(a_1|c_*, c^*, G) \leq \overline{u}(a_2|c_*, c^*, G).$$

Note 51: Proof revised by Ihara

(0) Preparation

Let $a_i = \{c_{ij}|E_{ij}, j = 1, \ldots, n_i\}, i = 1, 2$. By Axiom 5(i), for all (i, j) there exist S'_{ij} such that

$$c_{ij} \sim \{c^*|S'_{ij}, c_*|S'^{c}_{ij}\},$$

where $c_* \leq c_{ij} \leq c^*$ for all (i, j).

By Axiom 4(iii), for all (i,j) there exist standard events S_{ij} such that

$$P(S_{ij}) = P(S'_{ij})$$

$$(E_{ij} \cap G) \perp S_{ij}.$$

By Proposition 2.13 and $(E_{ij} \cap G) \perp S_{ij}$ we obtain

$$S_{ij} \perp (E_{ij} \cap G).$$

By Proposition 2.10 and $P(S_{ij}) = P(S'_{ij})$

$$S_{ij} \sim S'_{ij}.$$

Hence by Definition 2.3 and $c_* < c^*$,

$$\{c^*|S'_{ij}, c_*|S'^c_{ij}\} \sim \{c^*|S_{ij}, c_*|S^c_{ij}\}.$$

Hence by Proposition 2.2(i),

$$c_{ij} \sim \{c^*|S_{ij}, c_*|S^c_{ij}\}.$$

Hence by definition 2.15

$$u(c_{ij}|c_*, c^*) = P(S_{ij}).$$

So far we have obtained for all (i,j),

$$c_{ij} \sim \{c^*|S_{ij}, c_*|S^c_{ij}\},$$

$$S_{ij} \perp E_{ij} \cap G,$$

$$u(c_{ij}|c_*, c^*) = P(S_{ij}).$$

By Definition 2.6, for all (i, j)

$$c_{ij} \sim \{c^*|S_{ij}, c_*|S_{ij}^c\} \Rightarrow c_{ij} \sim_{E_{ij} \cap G} \{c^*|S_{ij}, c_*|S_{ij}^c\}$$

By Definition 2.4(i), for all (i, j) and any option a,

$$\{[c_{ij}|E_{ij} \cap G], a|(E_{ij} \cap G)^c\} \sim \{[\{c^*|S_{ij}, c_*|S_{ij}^c\}|E_{ij} \cap G], a|(E_{ij} \cap G)^c\}$$

This **MEANS** that for all (i, j),

$$[c_{ij}|E_{ij} \cap G] \sim [\{c^*|S_{ij}, c_*|S_{ij}^c\}|E_{ij} \cap G].$$

Therefore, we obtain

$$\{[c_{ij}|E_{ij} \cap G], j = 1, \ldots, n_i, a|G^c\} \sim \{[\{c^*|S_{ij}, c_*|S_{ij}^c\}|E_{ij} \cap G], j = 1, \ldots, n_i, a|G^c\},$$

which may be written as

$$\{\{c^*|A_i, c_*|B_i\}|G, a|G^c\},$$

where

$$A_i = \cup_j (S_{ij} \cap E_{ij})$$

$$B_i = \cup_j (S_{ij}^c \cap E_{ij})$$

$$A_i \cup B_i = \Omega$$
$$A_i^c = B_i.$$

Note the following:

$$\{a_i|G, a|G^c\} = \{[c_{ij}|E_{ij} \cap G], j = 1, \ldots, n_i, a|G^c\} \sim \{\{c^*|A_i, c_*|B_i\}|G, a|G^c\}$$

By Definition 2.4(i), for any option a,

$$a_1 \leq_G a_2 \Leftrightarrow \{a_1|G, a|G^c\} \leq \{a_2|G, a|G^c\}$$
$$\Leftrightarrow \{\{c^*|A_1, c_*|B_1\}|G, a|G^c\} \leq \{\{c^*|A_2, c_*|B_2\}|G, a|G^c\}.$$

Hence by using Definition 2.4(i) again, we obtain

$$a_1 \leq_G a_2 \Leftrightarrow \{c^*|A_1, c_*|B_1\} \leq_G \{c^*|A_2, c_*|B_2\}.$$

By Propostion 2.16, we obtain

$$A_1 \leq_G A_2 \Leftrightarrow P(A_1|G) \leq P(A_2|G).$$

By Propositions 2.13 and 2.15,

$$P(S_{ij} \cap E_{ij} \cap G) = P(S_{ij})P(E_{ij} \cap G) = P(S_{ij})P(E_{ij}|G)P(G).$$

By Corollary to Proposition 2.11,

$$P(A_i \cap G) = P([\cup_j(S_{ij} \cap E_{ij})] \cap G) = P(\cup_j[S_{ij} \cap E_{ij} \cap G])$$
$$= \sum_{j=1}^{n_i} P(S_{ij} \cap E_{ij} \cap G) = [\sum_{j=1}^{n_i} P(S_{ij})P(E_{ij}|G)]P(G).$$

Hence, we obtain

$$P(A_i|G) = \sum_{j=1}^{n_i} P(S_{ij})P(E_{ij}|G) = \sum_{j=1}^{n_i} u(c_{ij}|c_*, c^*)P(E_{ij}|G) = \overline{u}(a_i|c_*, c^*, G).$$

(1) $\Rightarrow$

$$a_1 \leq_G a_2 \Rightarrow \{c^*|A_1, c_*|B_1\} \leq_G \{c^*|A_2, c_*|B_2\}$$

By **Proposition 2.14(ii)** and $c_* < c^*$, we obtain

$$\{c^*|A_1, c_*|B_1\} \leq_G \{c^*|A_2, c_*|B_2\} \Rightarrow A_1 \leq_G A_2.$$

Therefore, we obtain

$$a_1 \leq_G a_2 \Rightarrow \overline{u}(a_1|c_*, c^*, G) \leq \overline{u}(a_2|c_*, c^*, G).$$

(2) $\Leftarrow$

$$\overline{u}(a_1|c_*, c^*, G) \leq \overline{u}(a_2|c_*, c^*, G) \Rightarrow A_1 \leq_G A_2$$

By **Axiom 3(i)** $c_* < c^* \Rightarrow c_* <_G c^*$. Hence by **Definition 2.4(ii)**, **for** $c_* <_G c^*$,

$$A_1 \leq_G A_2 \Rightarrow \{c^*|A_1, c_*|A_1^c\} \leq_G \{c^*|A_2, c_*|A_2^c\},$$

which may be written by using $A_i^c = B_i$ as

$$\{c^*|A_1, c_*|B_1\} \leq_G \{c^*|A_2, c_*|B_2\}.$$

Therefore, we obtain

$$\overline{u}(a_1|c_*, c^*, G) \leq \overline{u}(a_2|c_*, c^*, G) \Rightarrow a_1 \leq_G a_2.$$

Q.E.D.

Note 52: Proposition 2.22 is sometimes referred to as the **principle of maximizing expected utility**. In our development, this is clearly not an indepndent "principle", but rather an implication of our assumptions and definitions. In Summary form, the resulting prescription of quantitative, coherent decision-making is: **choose the option with the greatest expected utility**.

2.4.2 General Decision Problems

Definition 2.17 (General utility function) -P.54-

Given a preference relation $\leq$, the utility $u(c|c_1, c_2)$ of a consequence c, relative to the consequences $c_1 < c_2$, is defined to be the real number u such that

(i) if $c < c_1$ and $c_1 \sim \{c_2|S_x, c|S_x^c\}$, then $u = -x/(1-x)$;

(ii) if $c_1 \leq c \leq c_2$ and $c \sim \{c_2|S_x, c_1|S_x^c\}$, then $u = x$;

(iii) if $c > c_2$ and $c_2 \sim \{c|S_x, c_1|S_x^c\}$, then $u = 1/x$

where $x = \mu(S_x)$ is the measure associated with the standard event S_x.

Proposition 2.23 (Existence and uniqueness of utilities) -P.55-

For any decision problem, and for any pair of consequences $c_1 < c_2$,

(i) for all c, $u(c|c_1, c_2)$ exists and is unique;

(ii) the value of $u(c|c_1, c_2)$ is unaffected by the occurrence of an event $G > \Phi$;

(iii) $u(c_1|c_1, c_2) = 0$ and $u(c_2|c_1, c_2) = 1$.

Proposition 2.24 (Linearlity) -P.56-

For all $c_1 < c_2$ and $c_3 < c_4$ there exist $A > 0$ and B such that, for all c, $u(c|c_1, c_2) = Au(c|c_3, c_4) + B$.

Proposition 2.25 (General decision criterion) The Main Theorem of Chapter 2 -P.56-

For any decison problem, pair of consequences $c_1 < c_2$, and event $G > \Phi$,

$$a_1 \leq_G a_2 \Longleftrightarrow \overline{u}(a_1|c_1, c_2, G) \leq \overline{u}(a_2|c_1, c_2, G).$$

Note 53: An immediate implication of Proposition 2.25 is that all options can be compared among themselves. We recall that we did **not** directly assume that comparisons could be made between all pair of options (an **assumption** which is often critisized as unjusitified; see, for example, Fine 1973, p.221).

Instead, we merely assumed that any pair of simple dichotomized options could be compared (Axiom 1(ii)) and all consequences with the (very simply structured) standard dichotomized options (Axiom 5(i)), and that the latter could be compared among themselves (Axiom 4(i)).

Note 54: If we **begin** by defining a utility function $u : \mathbf{C} \to \mathbf{Re}$, this **induces** in turn a preference ordering which is necessarily coherent. Any function can serve as a utility function (subject only to the exsistence of the expected utility for each option, a problem which does not arise in the case of finite partitions) and the choice is personal one.

This completes our elaboration of the axiom system set out in Section 2.3. Starting from the primitive notion of preference, $\leq$,

we have shown that quantitative, coherent comparisons of options must proceed *AS IF* **a utility function has been asigned to consequences, probabilities to events and the choice of an option made on the basis of maximizing expected utility.**