Quasi-Bayesian Behaviour: A more realistic approach to decision making?

F.J. GIRON
Universidad de Malaga
and
S. RIOS
Universidad de Madrid

SUMMARY

In this paper the theoretical and practical implications of dropping-from the basic Bayesian coherence principles- the assumption of comparability of every pair of acts is examined. The resulting theory is shown to be still perfectly coherent and has Bayesian theory as a particular case. In particular we question the need of weakening or ruling out some of the axioms that constitute the coherence principles; what are their practical implications; how this drive to the notion of partial information or partial uncertainty in a certain sense; how this partial information is combined with sample information and how this relates to Bayesian methods. We also point out the relation of this approach to rational behaviour with the more (and apparently unrelated) general notion of domination structures as applied to multicriteria decision making.

Keywords: COHERENCE PRINCIPLES; AXIOMS OF RATIONALITY; PARTIAL UNCERTAINTY; BAYESIAN PREORDERS; PARTIALLY ORDERED PROBABILITIES; UPPER AND LOWER PROBABILITIES; DOMINATION STRUCTURES.

1. INTRODUCTION

As it is well known, Bayesian coherence principles as applied to decision problems imply the existence of a utility function, unique up to a linear transformation, and what is more important from the inferential point of view, a *unique* probability measure (known as subjective or personal probability) such that in order to choose among acts that which maximizes expected utility is selected.

Thus these principles assume that in any decision problem under uncertainty the decision maker is able - by instrospection or by any other means - to assign unique probabilities to every possible event and he will choose the decision which maximizes his expected utility.

On the other hand, if nothing is known about the true state of Nature, and one does not want to stick to incoherent principles such as minimax, etc., the only solution is to turn one's attention to admissible acts or decisions by application of the dominance principle implied by the natural ordering of decision rules once utilities have been assigned.

Between these two cases: 1^{st}) - prior distribution is completely known; and 2^{nd}) - nothing is known about the prior distribution (except the trivial fact that it is a probability measure, the existence of which may be even questioned), we may place the case of *partial ignorance* or *partial uncertainty*.

What we call partial ignorance refers to the fact we represent our knowledge about states of Nature by means of a set of probability measures to which the true distributions belongs.

In a more general sense "partial ignorance" could represent information about the states of Nature **not necessarily** given in the form of probability distributions. However our axioms or rationality principles will rule out this second interpretation of partial ignorance. In other words, we shall prove that a weakening of Bayesian coherence principles characterize partial ignorance in terms of a set of probability measures and that this characterization embodies the two extreme cases (of total ignorance, and perfect knowledge of prior distribution) which are but particular cases.

The idea of representing partial ignorance by convex sets of probability measures or by means of the related concept of lower and upper probabilities is not new and dates back to Smith (1961, 1965), Good (1962) and Dempster (1968)[1], and more recently to Suppes (1974) and Ríos (1975a, 1975b, 1976). However, none of these authors give a complete characterization of partial ignorance. Smith (1961) gives a partial answer to this question for the finite case. More refined results are found in Girón (1978).

Partial ignorance may be looked at in two different ways. First, suppose the decision maker is uncertain about his *prior P* so he expresses his beliefs in the form of a statement such as *P belongs to* κ^*. The form and size of κ^* measures his relative uncertainty. It is remarkable (see theorems 3.2, 3.4 and 3.6) that if a decision maker reflects his uncertainty about states of Nature in such a way that he is not able to compare every pair of acts (Axiom A1, sec. 3) whilst other axioms hold, then his uncertainty can be measured in terms of a

[1] As early as 1940, Koopman (1940) pioneered the idea that not every pair of events are comparable. In our approach this is a result of dropping the completeness axiom *C* (see section 3).

set of probability measures and he compares acts in terms of expected utilities against the probability measures of this set.

Second interpretation runs as follows: suppose an arbitrary number of decision makers each one being perfectly coherent (that is, their preferences satisfy axioms A1 to A5, and C of section 3). Suppose further that the utilities they assign to consequences are in agreement but they differ in their preferences, that is, their personal or subjective probabilities differ. Then the intersection of their preference systems is a new preference relation satisfying axioms A1 to A5. In this case, partial information or uncertainty is represented by the convex set generated by the set of all prior distributions corresponding to the decision makers, which, in this second version, could be named the **feasible set**.

If we call coherence principles the axioms A1 to A5, and C, we shall now discuss, briefly, the implications of dropping any of them. We do not discuss the necessity of axioms A1, A2, A3 and A4 as it is well known from the literature that dispensing with any of them drive to incoherent decisions.

As to axiom A5 we could dispense with it. In this case the preference relation would be a lexicographic order that would be characterized by a multidimensional (or lexicographic) subjective probability $P = (P_1, P_2, \ldots)$.

So the principle under discussion is completeness (axiom C). In its favour one may say that whichever the decision or inferential problem one is faced at a decision has to be made, and this imply that the decision maker or statistician is able to compare every pair of acts. However in case of partial ignorance the decision maker restricts his attention to non-dominated decisions. If this set is a small one and the corresponding Bayes risks do not differ much, this might be considered as though one would be performing a sensitivity analysis in a Bayesian case (e.g., see Fishburn (1964)).

From the purely inferential view point both approaches - partial versus Bayesian knowledge - are even closer. In the Bayesian case all information is in the posterior distribution while in the quasi-Bayesian case all relevant information is in the posterior set. But this last situation can be assimilated to the first one by taking a greater sample (see, e.g., example a) of section 4).

Note the difference between dropping the completeness axiom in utility theory (Aumann (1962, 1964), Criado (1978)) and in subjective probability theory. In the first case partial knowledge of utility function is not reduced (in fact sample information is independent of utility) by sample information; yet in case of partial knowledge of prior distribution, sample information reduces uncertainty[2]. That means that our initial partial preorder converges to a

[2] See Girón (1979) for a discussion on duality between the concepts of utility and subjective probability.

complete preorder when sample size increases.

Thus dropping completeness axiom is not made for sake of mathematical generalization but to convey a rational model for the case when it is difficult to choose among decisions. The practical conclusion is: "if you feel unsure about your decisions, then take a greater sample than the one you would take if you were able to compare every pair of decisions and you will do (nearly) as well".

2. DECISION MODEL WITH PARTIAL INFORMATION

Let $(\Omega, D; L)$ be a decision problem, where Ω is a set of states of Nature or parameter space that for illustrative purposes we suppose is finite and will be denoted $\Omega = \{\theta_1, \ldots, \theta_n\}$ (later on this section this restriction will be lifted); D is a set of possible decisions, which allowing for randomization may be supposed convex, and L is a loss function (the negative of a utility function), that is:

$$L: \Omega \, x \, D \rightarrow \mathbb{R}.$$

In the Bayesian case we also have information on Ω given in the form of a single probability measure, known as "the prior distribution", which we denote by P. In our case P can be identified with a point of the n-simplex of R^n that will be denoted

$$\Omega^* = \{ (p_1, \ldots, p_n); \Sigma_{i=1}^{n} p_i = 1; \; p_i \geq 0; i = 1, \ldots, n \},$$

where it is understood that $p_i = P(\theta_i)$, so that Ω^* would be the set of all probability measures.

If K^* is a nonempty subset of Ω^*, then partial information about P (the "true" prior distribution) is to state simply that $P \in K^*$. If K^* in fact represents partial ignorance, it may be taken to be convex, for if the decision maker is uncertain about P_1 and $P_2 \in K^*$, then he is uncertain about $\alpha \, P_1 + (1\text{-}\alpha) \, P_2 \; (0 \leq \alpha \leq 1)$. So convexity of K^* is not introduced for mathematical convenience but as a fairly natural condition.[3]

[3] The topological condition of K^* being closed is not really essential for as we shall show either K^* or $\overline{K^*}$ (its closure) generate the same quasi-Bayesian preorder. Note that in the Bayesian case, K^* reduces to a point which is closed. Convexity could also be dispensed with as it can be shown that K^* and con (K^*) (convex hull) generates the **same quasi-Bayesian preorder.**

Def. 2.1. A decision model with partial information is a quadruplet $(\Omega, D; L; K^*)$ where K^* is a nonempty closed convex set of Ω^*, which will be called the **uncertainty set** or the **prior distribution set.**

As particular cases we have: 1) the case of complete ignorance, when $K^* = \Omega^*$; 2) the case of perfect knowledge of the prior distribution when $K^* = \{P\}$, that is, K^* reduces to a point, or Bayesian case.

As most of the ideas we are to set forth have simple geometrical interpretations, it will be convenient to transform the decision problem into an equivalent S-game[4] as follows:

Define the risk set S of decision problem $(\Omega, K; L)$ by

$$S = \{x = (x_1,\ldots,x_n); \exists\, d \in D;\ L(\theta_i, d) = x_i\}$$

Let us consider the simplest case of two states of Nature, that is, $\Omega = \{\theta_1, \theta_2\}$. Then the partial information about $P = (p_1, p_2)$ is given in its more general form, by inequalities

$$\alpha_1 \leq p_1 \leq \alpha_1',$$

with α_1, α_1' constants such that $0 \leq \alpha_1 \leq \alpha_1' \leq 1$.

The set K^* can be geometrically represented by the angle determined by the extreme vectors $(\alpha_1, 1-\alpha_1)$, $(\alpha_1', 1-\alpha_1')$. Let $x^* = (x_1^*, x_2^*)$ be a fixed point of the risk set. Then the Bayes risk for x^* against prior distribution $P = (p_1, p_2) \in K^*$ is

$$r(x^*; P) = p_1 x_1^* + p_2 x_2^*.$$

If we take as priors the extreme point of K^*, say $P_1 = (\alpha_1, 1-\alpha_1)$ and $P_1' = (\alpha_1', 1-\alpha_1')$, the corresponding Bayes risks are

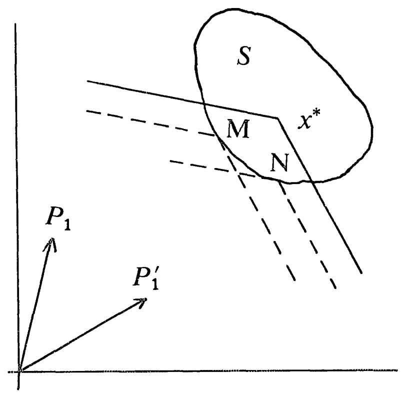
$$r\,(x^*;P_1) = \alpha_1 x_1^* + (1-\alpha_1)x_2^*$$
$$r(x^*; P_1') = \alpha_1' x_1' + (1-\alpha_1')\, x_2^*,$$

If we now consider the set of (possible) points that have smaller Bayes risk against both P_1 and P_1', we see that these points lie in the intersection of the closed half-planes given by the following inequalities

[4] For a definition of S-games see Blackwell and Girshick (1954), that suffices for the finite case. For a more general definition see Girón (1975).

$$\alpha_1 x_1 + (1-\alpha_1) x_2 \leq r(x^*, P_1)$$
$$\alpha_1' x_1 + (1-\alpha_1') x_2 \leq r(x^*, P_1') \qquad (2.1)$$

that define an angle with vertex at x^* (see figure). The most



important point to notice is that every point belonging to the angle, that is, satisfying inequalities (2.1), has smaller Bayes risk than x^* against any prior distribution $P \epsilon K^*$.

A further point to notice is that the angle does not depend on the chosen x^*, that is, if $y^* \neq x^*$ then the angle corresponding to y^* is simply a translation of the one with vertex at x^*. We shall denote this angle with vertex at origin by K. So K depends only on K^*.

This itself suggests the idea of defining a partial preorder on S (which is extended to R in an obvious way) by means of angle K and then, regard as solutions of the decision problem the maximal points (decisions) in S (in D). Thus maximal points in this weak order will coincide[5] with Bayes solutions against all probability measures belonging to K^*.

In the above figure this set is represented by the arch MN.

Note that in the case of total ignorance, that is, $\alpha_1 = 0$, $\alpha_1' = 1$, the angle defined by (2.1) is precisely the set of points $x = (x_1, x_2)$ such that

$$x_1 \leq x_1^*, x_2 \leq x_2^*;$$

[5] This is not true as stated because the relation between maximal and Bayes solutions in this case is analogous to the existing relation in the well-known case of natural ordering. For details see Rios (1976).

that is, the natural ordering of risk points in R^n.

The last and most important point to notice is the relation existing between K and K^*. In the simple case considered K is but the polar cone of set K^*. Thus partial knowledge represented by K^* induces in the space of possible decision functions a ''domination structure'' which is characterized by the polar cone of set K^*.

Recall (see definition of polar cone below) that the polar cone K is closed and that polar cones of K^* and $\overline{K^*}$ are the same. Further in the example considered the polar cone of K^* and of the set of its extremal points $\{P_1, P_2\}$ is the same.

These mathematical properties justify the hypotheses put on the set K^* of convexity and closedness. In next section these properties will be justified, through an axiomatic approach, from simple coherence principles.

Let us now return to the case of a finite number of states of nature $\{\theta_1,\ldots,\theta_n\}$. We are to define what we understand by quasi-Bayesian preference relations associated to a decision problem with partial ignorance.

Def. 2.2. Let $(\Omega, S; K^*)$ be a decision problem with partial uncertainty. We shall call **K^*-Bayesian preference** or **quasi-Bayesian preference** in S to the relation $\gtrsim_k{}^*$ defined for every $x, y \in S$ by

$$x \gtrsim_k{}^* y \text{ if and only if } x.P \leq y.P \text{ for every } P \in K^*,$$

where $x.P$ denotes dot-product.

It can be shown that $\gtrsim_k{}^*$ is a weak partial order satisfying axioms A1 to A5 of section 3. Moreover, $\gtrsim_k{}^*$ is complete (or linear if and only if K^* reduces to a point $\{P\}$: In this last case $\gtrsim_p$ is called a **Bayesian preference relation.**

Let K^* be the uncertainty set; denote by K the polar cone of set K^*, that is

$$K = \{x = (x_1,\ldots,x_n) \in \mathbb{R}^n \text{ ; } x.P \leq 0 \text{ every } P \in K^*\}.$$

K is a closed convex cone of $\mathbb{R}^n$ with vertex at origin. This defines a preference relation in $\mathbb{R}^n$ (and consequently in S) as follows

Def. 2.3. Let $x, y \in S$. x K- **dominates** y and will be denoted $x \gtrsim_k y$ if and only if $x\text{-}y \in K$.

The relation between the two definitions, which is but a consequence of duality, is the content of next result.

Theorem 2.1.

$$x \gtrsim_k{}^* y \text{ if and only if } x \gtrsim_k y.$$

It is worth mentioning that if K reduces to a point P, then K is the closed half-space defined by $\{x \epsilon \mathbb{R}^n ; x \cdot P \leq 0\}$. In case of total ignorance K is the negative orthant $D_o = \{x \epsilon \mathbb{R}^n, x_i \leq 0, i = 1,2,,\ldots,n\}$.

Def. 2.4. Let $\gtrsim$ and $\gtrsim^*$ be two weak order relations. Then, relation $\gtrsim$ **is included in** $\gtrsim^*$ if and only if $x \gtrsim^* y$ implies $x \gtrsim y$.

Theorem 2.2. Let K_1^*, and K_2^* be subsets of Ω^*, then $K^*_1 \supset K_2^*$ implies $\gtrsim_{k1}{}^*$ is included in $\gtrsim_{k2}^*$. Moreover, if K_1^*, and K_2^* are closed convex sets, the conversely statement is also true.

As a consequence of duality and theorem 2.1., we have the following

Corollary 2.1. $\gtrsim_{k1}$ is included in $\gtrsim_{k2}$ if and only if $K_1 \subset K_2$.

These partial weak orderings give rise to definitions of admissibility, complete classes and quasi-Bayesian (or K-Bayesian) decisions.

Def. 2.5. A risk point $x \epsilon S$ is **K^*-Bayes (or quasi-Bayesian)** for the problem $(\Omega, S; K^*)$ if there exists at least a $P \epsilon K^*$ such that x is Bayes against P. Accordingly $d \epsilon D$ is **K^*-Bayes** if its corresponding risk point is K^*- Bayes.

The set of all K^*-Bayes strategies will be denoted $\mathbb{B}(K^*; S)$ or $\mathbb{B}(K^*; D)$.

Relations among K-admissibility (defined in an obvious way), K-Bayesness and completeness can be found in Ríos (1976), in which the name "quasi-Bayes" was coined.

In this paper we do not discuss the computational aspects of quasi-Bayesian strategies. In the finite case, here considered, methods for finding non-dominated strategies are to be found in Leitmann (1976) and references therein. These procedures, devised for general convex domination structures, can be applied *mutatis mutandis* to the problem of finding quasi-Bayesian strategies in case K^* be a convex polyhedron by means of linear and non-linear programming technics. The general case of K^* being an arbitrary convex set may be treated by approximative methods (see reference above).

By far the most important feature of quasi-Bayesian methods is that they allow incorporation of the information provided by an experiment by use of Bayes theorem.

Let $(X, A_z; P_\theta(x))$ be an experiment, where $\Omega = \{\theta_1,\ldots,\theta_n\}$. Let $P(\theta_i | x)$ denote the posterior probability of θ_i when x has been observed and prior is $P(\theta_i)$. We define the **posterior uncertainty set** (or posterior partial information set) as the set of all posterior distributions of K^* when x is observed. This set will be represented by K^*_x. Sometimes we shall refer to this set K_x^* **as the transform of K^* through sampling when x is observed.** Properties of posterior uncertainty sets are summarized in the following.

Theorem 2.3. If K^* is a closed convex set of Ω^*, then K^*_x is also a closed convex set for every $x \in X$. Furthermore, extremal prior distributions of K^* are transformed through sampling into extremal distributions of K^*_x for any $x \in X$.

The second part of theorem usually simplifies the problem of finding the posterior uncertainty set if only we know the extremal prior distributions.

Finally, we mention the fact that the whole set of probability distributions Ω^* is invariant through sampling, that is $\Omega^*_x = \Omega^*$ for any $x \in X$. This is but a statement that total ignorance cannot be changed into partial ignorance through sampling.

Def. 2.6. Let $(\Omega, D; L; K^*)$ be a decision problem with partial information, $(X, A_x; P_\theta(x))$ an experiment. We say $\delta: (X, A_x) \rightarrow (D, A_D)$ is K^*-Bayes (or quasi-Bayes) if for every $x \in X$, $\delta(x)$ is K^*-Bayes for the problem $(\Omega, D; K^*_x)$.

Most definitions and results given in this sections are easily generalizable to the case of an infinite number of states with slight modifications except in one instance. This refers to the duality between K^* and its polar cone K that poses delicate analytical problems due, in part, to the lack of reflexivity (in the sense of functional analysis) of some of the spaces of measures under consideration, and secondly to the problem that appears in some statistical applications that D and Ω^* cannot be embeded in topological vector spaces for which one is the dual of the other one.

3. AXIOMATIC CHARACTERIZATION OF PARTIAL UNCERTAINTY

As we stated in the introduction, partial uncertainty is usually represented by a convex set of probability measures and may be considered midway between total ignorance (no knowledge of the "true" [if any] prior distribution) and, on the other hand, whole knowledge of the prior distribution (Bayesian view point).

Most axiomatic characterizations of subjective probability and, consequently, of Bayesian behaviour in the case of decisions under uncertainty are based in the ability of the decision-maker at ordering **any** pair of acts or events he is confronted with; which, as it is well known, is one of the basic principles of the so called "Bayesian coherence principles".

Here we present a variant of the above mentioned principles that still preserve the Bayesian "flavour" but have into account this possibility and, in fact, they fully characterize "partial ignorance". Basically, we follow the axioms given by Girón (1974, 1977)[6] that characterize subjective probability and the principle of maximization of expected utility.

The basic idea is the suppression of the completeness of the preference relation in the set of all possible decisions, along the lines of Aumann's contribution to utility theory [see, Aumann (1962, 1964)], which could justify the name of "subjective probability without the completeness axiom" instead of "partial ignorance".

One of the main results of this section is the characterization of **all** partial ignorance relations (this includes the Bayesian case) in terms of a class of closed convex cones in the space of decisions. This first characterization is inspired in the papers of S. Ríos (1975a, 1975b, 1976) on quasi-Bayes orders, and, on the other hand, in the work of Yu, Zeleny *et al.*[7] on domination structures.

The second characterization is the basic result we are seeking for; stated in imprecise terms it asserts that partial ignorance is characterized in terms of closed convex subsets of a space of probability measures.

In the following it will be convenient to distinguish two cases; namely: a) partial ignorance is represented in terms of σ-additive probability measures (abbreviated p.m.); b) these probabilities are only assumed to be finitely additive.

In case a) (See, e.g. Girón (1977), p. 33) a restriction on the set of states of Nature needs be imposed; namely, it is supposed to be a compact Hausdorff topological space; a further restriction is that decisions can be identified with a subset of continuous functions on such a space. However, in case b), the parametric space can be quite arbitrary and decisions or acts are only supposed to be bounded.

Case a), in spite of its apparent restrictiveness, it is not so, for many decision problems are such that the parameter space may be endowed with a metric (e.g., the intrinsic metric) which makes all acts continuous so that the only restriction would now be compactness respect to that topology.

Let Ω denote the space of states of Nature or parameter space, D a set of decisions on terminal acts, and u: $\Omega\, xD \rightarrow \mathbb{R}$ a utility function.

Def. 3.1. A decision problem under uncertainty (which, in the sequel, will be abbreviated as d.p.u.u.) is a triplet $(\Omega, D; u)$.

In case a) D can be identified with a subset S of $C\,(\Omega)$- space of all real continuous functions defined on Ω -namely

[6] These axioms were put forward in a later paper (see Girón (1978)), in a stronger form that the one given in this article.

[7] Most of their work appears in Leitmann. (1976).

$$S = \{f(\theta) \epsilon\ C(\Omega);\ \exists\ d \epsilon D; f(\theta) = u(\theta,d)\}.$$

In case b) D is identified with a subset S of B (Ω) -space of all real bounded functions defined on Ω -defined by

$$S = \{f(\theta) \epsilon B(\Omega);\ \exists\ d \epsilon D; f(\theta) = u(\theta,d)\}.$$

Further, if the decision maker or statistician allows for randomization in D, S may be regarded as a convex subset of the linear spaces, $C(\Omega)$ and $B(\Omega)$, respectively.

This suggests a new definition of a d.p.u.u.

Def. 3.2. A d.p.u.u. is a pair (Ω,S) where S is a nonempty convex subset of C (Ω) (case a) or B (Ω) (case b).

Def. 3.3. If (Ω,S) is a d.p.u.u. a **decision criterion** is a binary relation on S, which will be denoted by $\gtrsim_s$

Relation $\gtrsim_s$ is read "...at least as preferred as...". Taking $\gtrsim_s$ as the basic relation we may define the following.

$$f >_s g \text{ iff } f \gtrsim_s g \text{ and not } g \gtrsim_s f$$

$$f \sim_s g \text{ iff } f \gtrsim_s g \text{ and } g \gtrsim_s f$$

$$f \not\sim_s g \text{ iff not } f \gtrsim_s g \text{ and not } g \gtrsim_s f$$

which are read "...(strictly) preferred to...", "...indifferent to..." and "...is not comparable to...", respectively. In the sequel $g \lesssim_s f$ will mean $f \gtrsim_s g$.

The list of proposed axioms is the following, that only differs of Girón's (1977) in the first one.

A1 **(Partial preorder)**.- For every d.p.u.u. (Ω,S), $\gtrsim_s$ is reflexive and transitive.

A2 **(Strong dominance)**.- If f, $g \epsilon S$ are such that $f(\theta) > g(\theta)$ for every $\theta \epsilon \Omega$, then $f >_s g$.

A3 **(Addition of new strategies)**.- If $S \subset R$, then $f \gtrsim_s g$ implies $f \gtrsim_R g$.

A4 **(Linearity)**.- If $\lambda \epsilon (0,1), f,g,h \epsilon S$, then $f \gtrsim_s g$ if and only if

$$\lambda f + (1-\lambda)\, h \gtrsim_s \lambda g + (1-\lambda) h$$

A5 **(Continuity)**.- If f_n, g, $h \epsilon S$ for $n = 1,2,\ldots$, are such that $\{f_n\} \rightarrow f \epsilon S$,

$$f_n \gtrsim_s g,\ h \gtrsim_s f_n$$

for every $n = 1,2,\ldots$, then $f \gtrsim_s g$ and $h \gtrsim_s f$.

Convergence in this axiom is understood with respect to the usual supremum norm topology given, for both a) and b) cases,

$$\| f \| = \sup_{\theta \epsilon \Omega} |f(\theta)| .$$

Next is a completeness axiom that will only be necessary in the characterization of Bayesian behaviour.

C (Completeness). For every f, $g \epsilon S$ either $f \gtrsim_s g$ or $g \gtrsim_s f$

Axiom 3 allows us to consider the $\gtrsim$ relation as being defined on $C(\Omega)$ $[B(\Omega)]$; then relation $\gtrsim_s$ is, simply, the restriction of $\gtrsim$ to S. Further, as $C(\Omega)$ $[B(\Omega)]$ are complete normed spaces, if $\{f_n\}$ converges to f, then $f \epsilon C(\Omega)$ $[B(\Omega)]$.

In case b), as $B(\Omega)$ contains the class of indicator functions of subsets of Ω, the relation $\gtrsim$ on $B(\Omega)$ restricted to this class allows us to define a new relation, $\gtrsim^*$, on the class of all subsets of Ω, which we shall denote by $P(\Omega)$, and will be called events.

Def. 3.4. Event A **is at least as probable as event B**, and will be denoted by $A \gtrsim^* B$ if $\lambda > \mu$ implies $f \gtrsim g$, where f and g are defined by

$$f(\theta) = \begin{cases} \lambda & \text{if } \theta \epsilon A \\ \mu & \text{if } \theta \notin A \end{cases}$$

$$g(\theta) = \begin{cases} \lambda & \text{if } \theta \epsilon B \\ \mu & \text{if } \theta \notin B \end{cases}$$

It can be easily seen that if $\gtrsim$ satisfies axioms A1, A2, A3, A4, A5, definition 3.4. does not depend on λ and μ, as far as $\lambda > \mu$. This is the content of the following lemma, which could have been taken as definition.

Lema 3.1.- A $\gtrsim^* B$ if and only if $I_A \gtrsim I_B$, where I_A and I_B denote the indicator functions of sets A and B, respectively.

Furthermore, relation $\gtrsim^*$ as defined above satisfies all axioms of comparative probability (e.g., see Fine (1973) p. 17) except the comparability of every pair of acts[8]. In particular $\gtrsim^*$ satisfies

[8] Recently Fishburn (1975) and Goodman (1977) have also considered a weakening of the comparability axiom in which indifference is not assumed to be transitive.

(i) $\gtrsim^*$ is reflexive and transitive.
(ii) $A \gtrsim^* \phi$ for every event $A \in P(\Omega)$.
(iii) $\Omega \gtrsim^* \phi$
(iv) Let A,B,C be events such that $A \cap C = B \cap C = \phi$, then

$$A \gtrsim^* B \text{ if and only if } A \cup C \gtrsim^* \mathrm{BUC}.$$

As was mentioned in section 2, the natural ordering in $C(\Omega)$ [$B(\Omega)$] is the weakest partial preorder every other "reasonable" partial, or complete, preorder should be consistent with. This consistency is taken up in the formulation of axioms A2 and A5.

Def. 3.5. *f* **dominates** *g*, and will be represented by $f \gtrsim_d g$ if $f(\theta) \geq g(\theta)$ for every $\theta \epsilon \Omega$.

Relation $\gtrsim_d$ is a partial preordering satisfying axioms A1, A2, A3, A4, A5. Morover, relation $\gtrsim_d^*$ induced in $P(\Omega)$ by $\gtrsim_d$ is subset inclusion, e.i.,

$$A \gtrsim_d B \text{ if and only if } A \supseteq B.$$

Those decisions dominated by the function $I_\phi \equiv$ o will be denoted by D_o, that is,

$$D_o = \{f;\ f(\theta) \leq 0 \ \text{ for every } \theta \epsilon \Omega\}$$

Some of the results that now follow were advanced in Girón (1978)[9].

Theorem 3.1. If relation $\gtrsim$ in $C(\Omega)$ [$B(\Omega)$] satisfies A1, A2, A3, A4, A5 then there exists a unique closed convex cone K, K not being the entire space, containing D_o and with vertex at the origin, such that

$$f \gtrsim g \text{ if and only if } g\text{-}f \epsilon K \tag{3.1}$$

Conversely, every non empty closed convex K, containing D_o and with vertex at the origin, defines a partial preordering $\gtrsim$ in $C(\Omega)$ [$B(\Omega)$] by (3.1)

Furthermore, $\gtrsim$ is a complete preordering if and only if K is a closed half-space containing D_o and passing through the origin.

[9] In this paper we give new results and some refinements and amendments of results that appeared in Giron (1978). Proofs will appear in a subsequent paper.

This theorem is interesting in order to examine the structure of partial preorderings in relation to complete preorders.

Let $\gtrsim_i$ be a collection of linear preorders satisfying axioms A1, A2, A3, A4, A5, and C, where $i \epsilon I$, a certain index set. If we define relation $\gtrsim$ by

$$f \gtrsim g \text{ if and only if } f \gtrsim_i g \text{ for every } i \epsilon I,$$

then $\gtrsim$ is a partial preordering satisfying A1, A2, A3, A4, A5. This relation could be named the **intersection of the class of preorderings** $\{\gtrsim_i\} i \epsilon I$

Now, by theorem 3.1, every partial preorder is characterized by a closed convex cone K and every complete preorder by a closed half-space, so that we have as a corollary of the theorem the following.

Corollary 3.1. Every partial preorder satistying A1, A2, A3, A4, A5 is the intersection of an arbitrary collection of linear preorderings satisfying A1, A2, A3, A4, A5 and C, and conversely.

It can also be shown that the intersection of an arbitrary collection of partial preorders satisfying A1 to A5 is a partial order satisfying A1 to A5.

If we call "quasi-Bayesian preorder" then corollary 3.1 simply states that every "quasi-Bayesian preorder" is the intersection of Bayesian preorders, thus giving a precise meaning to the second interpretation of partial ignorance mentioned in the introduction.

Next theorem, and its counterpart for case b) (see theorem 3.4), characterizes a partial ignorance in terms of a set of probability measures.

Theorem 3.2. If relation $\gtrsim$ in $C(\Omega)$ satisfies A1 to A5 then there exists a unique non empty closed convex set K^* of σ-additive probability measures on the Borel field of the topological space (Ω, B_Ω) such that

$$f \gtrsim g \text{ if and only if } \int f\, d\mu \geq \int g d\mu \text{ for every } \mu \epsilon K^*$$

If $\gtrsim$ further satisfies axiom C, then K^* reduces to a single probability measure.

The second part of this theorem characterizes Bayes behaviour.

Technical note. In this theorem as well as in theorem 3.4 below, K^* is closed in the weak * topology.

Next theorem characterizes the natural ordering relation $\gtrsim$ in case a), the necessary part of the theorem being as well known result in integration theory. In fact, it is a particular case of theorem 3.2 that characterizes total ignorance.

Theorem 3.3. For every f, $g \in C(\Omega)$

$f \gtrsim g$ if and only if $\int f d\mu \geq \int g d\mu$

for every $\mu \in \Omega^*$, where Ω^* is the set of all probability measures (σ-additive) on the space (Ω, B_Ω).

The corresponding theorems for case b) are:

Theorem 3.4. If relation $\gtrsim$ on $B(\Omega)$ satisfies A1 to A5, then there exists a unique nonempty closed convex set K^* of finitely additive probability measures on $P(\Omega)$ such that

$$f \gtrsim g \text{ if and only if } \int f dP \geq \int g dP \text{ for every } P \in K^*$$

If $\gtrsim$ further satisfies axiom C, then K^* reduces to a unique probability measure.

Theorem 3.5. For every f, $g \in B(\Omega)$

$$f \gtrsim_d g \text{ if and only if } \int f dP \geq \int g dP$$

for every $P \in \Omega^*$, where Ω^* is the set of all finitely additive probability measures on the space $(\Omega, P(\Omega))$.

Next two theorems refer to the comparative probability relation $\gtrsim^*$ of definition 3.4 or lemma 4.1.

Theorem 3.6. A $\gtrsim^* B$ if and only if $P(A) \geq P(B)$ for every $P \in K^*$, where K^* is the set of theorem 3.4.

Theorem 3.7. For every pair of events $A, B \in P(\Omega)$

$A \supseteq B$ if and only if $P(A) \geq P(B)$

for every $P \in K^*$, where K^* is the set defined in theorem 3.5.

Theorem 3.6 could be used to define a system of lower and upper probabilities associated to the CP partial relation $\gtrsim$, in the following manner

$$P_*(A) = \operatorname*{Inf}_{P \in K^*} P(A),$$

$$P_*(A) = \operatorname*{Sup}_{P \in K^*} P(A).$$

Yet the properties of P_*, P^* will not be further explored in this paper, as our intention was to fully characterize partial ignorance.

This section ends with a few results referring to conditional preference. They essentially show that the intuitive ideas set forth in section 2 about the incorporation of information given by an experiment to partial prior ignorance, given in the form of a convex set of probability measures, through the use of Bayes theorem are sound and have an axiomatic foundation. It is also proven that the posterior set of probability measures is also a closed convex set, which generalizes last theorem of section 2.

Definition of conditional preference appears in a different form that the one given in Savage (1954) and Girón (1977) for the sake of mathematical tractability.

Def. 3.6. Let f and g be two given acts. f **is at least as preferred as** g **when** A **obtains,** and will denoted $f \gtrsim g$ **given** A, if and only if $I_A f \gtrsim I_A g$.

Def. 3.7. Event A **is null,** if and only if $f(\theta) > g(\theta)$ for every $\theta \in \Omega$ does not imply $f \gtrsim g$ given A.

Properties of null events derived from axioms A1 to A5 are similar to the ones given by Savage (1954). In particular we have

(i) ϕ is a null event.
(ii) If A is null and $B \in A$, then B is null.
(iii) The union of any finite number of null events is null.
(iv) Ω is not null.

In terms of the set K^* null events are characterized by the following:

Theorem 3.8. A is null if and only if there exists at least a $P \in K^*$ such that $P(A) = 0$.

Next lemma is a trivial consequence of definition 3.7., but conveys an important result in conjunction with theorem 3.4.

Lemma 3.2. If A is not null, relation $\lesssim$ given A, satisfies axiom A1 to A5.

Next theorem characterizes conditional preference.

Theorem 3.9. If axioms A1 to A5 hold and event A is not null, then there exists a unique closed convex set $K_A^* \subset \Omega^*$ such that

$$f \gtrsim g \text{ given } A, \text{ if and only if, } \int f\, dP \geq \int g\, dP$$

for every $P \in K^*_A$.

The relation between sets K^* and K^*_A of theorems 3.4 and 3.9 is given by the following theorem that shows that K^*_A is precisely the set of all conditional probability measures of K^*.

Theorem 3.10. If A is not null, then

$$K^*_A = \{P_A \epsilon \Omega^* \ ; \ \exists \ P \ \epsilon K^* \ ; \ P_A(B) = \frac{P\,(A \cap B)}{P\,(A)} \text{ for every } B \epsilon P\,(\Omega)\}$$

This has a clear behavioural interpretation in terms of intersection of orders: We know from theorems 3.4, 3.6 and corollary 3.1, that every quasi-Bayesian preorder is the intersection of quasi-Bayesian preorders. Now, suppose we are given the piece of information that "event A has obtained" and A is not null. It can be easily shown that if the partial preorder $\gtrsim$ is the intersection of $\gtrsim_i$, for $i \epsilon I$, then A is not null for $\gtrsim_i$ for every $i \epsilon I$. If $\gtrsim_i$ is characterized by subjective probability P_i and event A obtains, then P_i is changed into P_{iA} to which corresponds $\gtrsim_i$ given A, so that $\gtrsim$ given A is precisely the intersection of the $\{\gtrsim_i$ given A $\}i \epsilon I$. This is in the spirit of Bayesian behaviour: «Change your prior partial information through use of Bayes theorem into the posterior partial information and act accordingly to the principle given in theorems 3.2 and 3.4 which could be named the **principle of maximalization of expected utility**».

As was pointed out at the end of section 2 partial ignorance can be characterized by the extreme point of set K^*, for as if we denote it by K^*_e, then $K^* = \overline{\text{con}}\ (K^*_e)$, so that any possible distribution is a general mixture of extreme distributions. It can be easily shown that extremal prior distributions change into extremal posterior distributions by use of Bayes theorem.

4. ILLUSTRATIVE EXAMPLES

In the last section we give a few simple examples in order to illustrate the form of quasi-Bayesian solutions.

In case quasi-Bayesian procedures are intended only for inferential purposes the answer lies on the structure of the posterior set of probability measures, or to reduce it to a minimum, all relevant information is given by the set of extremal distribution of this set.

In the case of decision problems, a loss or utility structure is imposed upon the inferential problem, thus reducing the decision problem to the calculation of a few parameters of the posterior extremal distributions, those parameters depending on the form of the loss function.

a) Quasi-Bayesian confidence intervals in the normal case

Suppose $X_1,\ldots,X_n$ is a random sample of a normal distribution $N\,(w,r)$ where the precision r is known and the mean w is unknown. The partial information on w is given by the subset of normal distributions $N\,(\mu,\tau)$ where τ is known and $\mu\epsilon[\mu_1, \mu_2]$. (Observe that this reduces to the well-known Bayesian case when $\mu_1 = \mu_2$).

A trite calculation shows that the extremal posterior set of distributions is the subset of normal distributions $N\,(\mu',\tau')$, where

$$\mu'\epsilon\left[\frac{\tau\mu_1 + nr\bar{x}}{\tau + nr}, \frac{\tau\mu_2 + nr\bar{x}}{\tau + nr}\right]$$

and

$$\tau' = \tau + nr \quad \text{with } \bar{x} = \frac{\Sigma\, x_i}{n}$$

Then the quasi-Bayesian confidence interval for w for a given confidence coefficient p is

$$\left[\frac{\tau\mu_1 + nr\bar{x}}{\tau + nr} - \lambda_p\left(\frac{1}{\tau + nr}\right)^{1/2}, \frac{\tau\mu_2 + nr\bar{x}}{\tau + nr} + \lambda_p\left(\frac{1}{\tau + nr}\right)^{1/2}\right],$$

where

$$\Phi\,(\lambda_p) - \Phi\,(-\lambda_p) = p.$$

Observe that any of the distributions of the posterior set (not only the extremal ones which are normal) assigns to this interval a probability greater than p.

Let us now see how this interval compares with a Bayesian confidence interval for any prior distribution compatible with our partial information.

Suppose the prior distribution is $N\,(\mu,\tau)$ with $\mu\epsilon[\mu_1, \mu_2]$. For a sample size n' the Bayesian p-confidence interval is

$$\left[\frac{\tau\mu + n'r'\bar{x}'}{\tau + n'r} - \lambda_p\left(\frac{1}{\tau + nr'}\right)^{1/2}, \frac{\tau\mu + n'r\bar{x}'}{\tau + n'r} + \lambda_p\left(\frac{1}{\tau + n'r}\right)^{1/2}\right],$$

where $\bar{x}' = \dfrac{\Sigma\, x_i}{n'}$

It is evident that for the same sample size the quasi-Bayesian interval is wider than the corresponding Bayesian one. If we now equate the width of the two intervals for sample sizes n and n' respectively we obtain

$$\frac{\tau(\mu_2-\mu_1)-2\lambda_p}{\tau+nr} = \frac{2\lambda_p}{\tau+n'r}$$

It is interesting to note that the above relation does not depend on $\bar{x}$, $\bar{x}'$ and it obviously implies that $n \geq n'$. The difference n-n' could be interpreted as the "additional sample size" for which partial prior information could be considered as total prior information.

b) Quasi-Bayesian estimators for the mean of a normal distribution

Suppose the same situation of normal sampling as in example a) with the same partial information. If the loss function for this decision problem is

$$L(w, d) = (w - d)^2$$

the quasi-Bayesian estimator is seen to be

$$\delta^*(x_1,\ldots,x_n) = \left[\frac{\tau\mu_1+nr\bar{x}}{\tau+nr}, \frac{\tau\mu_2+nr\bar{x}}{\tau+nr}\right] \tag{4.1}$$

which reduces to a single point if either μ_2-$\mu_1 \to 0$, $\tau \to 0$, or $n \to \infty$

It deserves mentioning that the quasi-Bayesian estimator in this case is the union of Bayes estimators corresponding to the extremal posterior distribution. Any Bayesian estimator corresponding to a non extremal posterior distribution belongs to δ^*.

Note that if partial information reduces to the following: "Prior information is normal $N(\mu, \tau)$ with $\mu = \lambda\mu_1+(1-\lambda)\mu_2$, $0 \leq \lambda \leq 1$", the quasi-Bayesian estimator is the same as the one given by (4.1)

c) Quasi-Bayesian testing of hypotheses

In this section we consider the simplest example of testing a simple null hypotheses versus a simple alternative hypotheses, so that the two states, two actions, decision problem is,

	θ_0	θ_1
a_0	o	a
a_1	b	o

$a,b>0$

where θ_0 stands for the null hypotheses and θ_1 for the alternative; a_0 accept θ_0 and a_1 reject θ_0 (and accept θ_1, accordingly)

Partial information in this example is given in the form of a closed interval that represents the range of possible values of prior probability on the null hypotheses, that is

$$P\,[\theta_0] \in [\xi_0, \xi_1] \qquad\qquad (0 \le \xi_0 \le \xi_1 \le 1)$$

If we represent the density (with respect to some dominating measure) of a sample of size one, when θ_i $(i=0,1)$ is true by f_i, then the quasi-Bayes procedure for this decision problem when a random sample of size n is taken, which we could name "quasi-Bayesian test", is the following

$$\delta^*(x_1,\ldots,x_n) = \begin{cases} a_0 \text{ if } \Pi_{i=1}^{n} \dfrac{f_1(x_i)}{f_0(x_i)} \le \dfrac{b}{a}\,\dfrac{\xi_0}{1-\xi_0} \\ a_1 \text{ if } \Pi_{i=1}^{n} \dfrac{f_1(x_i)}{f_0(x_i)} \ge \dfrac{b}{a}\,\dfrac{\xi_1}{1-\xi_1} \\ \{a_0,a_1,\} \text{ if } \dfrac{b}{a}\,\dfrac{\xi_0}{1-\xi_0} < \dfrac{\Pi_{i=1}^{n} f_1(x_i)}{\Pi_{i=1}^{n} f_0(x_i)} < \dfrac{b}{a}\,\dfrac{\xi_1}{1-\xi_1} \end{cases}$$

This results needs some explanation: If the sample observed is such that $\delta^*(x_1,\ldots,x_n)$ equals a_0 or a_1, there is no problem, and the null hypotheses is accepted or rejected, respectively. If, however, $\delta^*(x_1,\ldots,x_n) = \{a_0, a_1\}$ then no single course of action is possible.

This means that our partial (posterior) information is not enough as to discriminate between the two actions so that new sample information is needed and a computation of the new likelihood ratio may show that $\delta^*(x_1,\ldots,x_n, x_{n+1})$ equals either a_0 or a_1 or if $\delta^*(x_1,\ldots,x_n,x_{n+1}) = \{a_0, a_1\}$ a new sample is required, and so on. This brings out the strong analogy between the quasi-bayesian test and Wald's sequential probability ratio test with barriers

$$A = \frac{b}{a}\,\frac{\xi_0}{1-\xi_0} \qquad \text{and } B = \frac{b}{a}\,\frac{\xi_1}{1-\xi_1},$$

in the case the cost of new observations is not included within the structure of the decision problem.

REFERENCES

AUMANN, R.J. (1962). Utility theory without the completeness axiom. *Econometrica* **30**, 445-462.

— (1964). Utility theory without the completeness axiom: a correction. *Econometrica* **32**, 210-212.

BLACKWELL, D. and GIRSHICK, M.A. (1954). *Theory of Games and Statistical Decisions.* Wiley: New York.

CRIADO, F. (1978). *Algunas caracterizaciones de la utilidad y extensiones.* Ph.D. Thesis. Universidad de Málaga.

DEMPSTER, A.D. (1967). Upper and lower probabilitites induced by a multivalued mapping. *Ann. Math. Statist.* **38**, 325-339.

— (1968). A generalization of Bayesian Inference. (with discussion). *J. Roy. Statist. Soc. B* **30**, 205-232.

FINE, T. (1973). *Theory of Probability: An Examination of Foundations.* Academic Press: New York.

FISHBURN, P.C. (1964). *Decision and Value Theory.* Wiley: New York.

— (1975). Weak comparative probability on infinite sets. *Ann. Prob.* **3**, 889-893.

GIRON, F.J. (1975). S-juegos generalizados. *Rev. Real Acad. Ciencias Madrid* **69**, 49-97.

— (1977). Caracterización axiomática de la regla de Bayes y la probabilidad subjetiva, *Rev. Acad. Cienc. Madrid.* **71**, 19-101.

— (1978). Una caracterización de la incertidumbre parcial. *Actas V Jornadas Luso-Espanholas de Matemática.* Aveiro, Portugal.

— (1979). Probabilidad y utilidad: conceptos duales de la teoría de la decisión. *Rev. Real Acad. Cienc. Madrid,* **73**, 225-230.

GOOD, I.J. (1962). The measure of a non-measurable set. In *Logic, Methodology and Philosophy of Science.* E. Nagel, P. Suppers and A. Tarski eds. 319-329. Standford: University Press.

GOODMAN, T.N.T. (1977). Qualitative probability and improper distributions. *J. Roy. Statist. Soc. B* **39**, 387-393.

KOOPMAN, B.O. (1940). The axioms and algebra of intuitive probabilty. *Ann. Math.* **41**, 269-292.

LEITMANN, G. (1976). *Multicriteria Decision Making and Differential Games.* London: Plenum Press.

RIOS, S. (1975a). Nuevos criterios de ordenación de reglas de decisión. *Trab. Estadist.* **26**, 5-12.

— (1975b). Ordre quasi-bayesien des regles de decision. *Proc. 40th Session of I.S.I. Warsaw,* 694-698.

— (1976). Nuevos criterios de ordenación de reglas de decisión. *Rev. Real Acad. Cienc. Madrid* **70**, 235-253.

SAVAGE, L.J. (1954). *The Foundations of Statistics.* Wiley: New York.

SMITH, C.A.B. (1961). Consistency in statistical inference and decision (with discussion). *J. Roy. Statist. Soc. B* **23**, 1-25.

— (1965). Personal probability and statistical analysis (with discussion). *J. Roy. Statist. Soc.* **A 128**, 469-499.

SUPPES, P. (1974). The measurement of belief (with discussion). *J. Roy. Statist. Soc. B* **36,** 160-175.