UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS MATEMÁTICAS

TESIS DOCTORAL

Problemas de decisión con información parcial a priori

MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR

maría Jesús Rios Insua

DIRECTOR:

Francisco Javier Girón

Madrid, 2015

María Jesús Ríos Insúa

PROBLEMAS DE DECISION CON INFORMACION PARCIAL A PRIORI

Departamento de Estadísitca e Investigación Operativa
Facultad de Ciencias Matemáticas
Universidad Complutense de Madrid
1982

Colección Tesis Doctorales. Nº 203/82

Edita e imprime la Editorial de la Universidad
Complutense de Madrid. Servicio de Reprografía
Noviciado, 3 Madrid-8
Madrid, 1982
Xerox 9200 XB 480
Depósito Legal: M-30741-1982

PROBLEMAS DE DECISION CON INFORMACION PARCIAL A PRIORI

MARIA JESUS RIOS INSUA

Memoria para optar al grado de Doctor en Ciencias Matemáticas, realizada bajo la dirección del Dr. D. Francisco Javier Girón

Universidad Complutense de Madrid
Facultad de Matemáticas
Sección de Estadística e Investigación Operativa

Madrid, Febrero de 1981

I N D I C E

PROLOGO

El propósito de esta memoria es profundizar en el estudio del modelo de decision estadístico con información parcial.

En la primera parte, planteamos el modelo básico, partiendo de un punto de vista análogo al bayesiano, convenientemente modificado, siguiendo una idea de Smith (1961), generalizada posteriormente por Girón y Rios (1981), estableciendo que la incertidumbre o ignorancia parcial puede representarse mediante un conjunto convexo y cerrado de medidas de probabilidad sobre el espacio paramétrico.

En el primer capítulo consideramos el modelo señalado, y estudiamos aspectos particulares del mismo, derivados del hecho de que, así caracterizada la incertidumbre parcial, las decisiones óptimas serán los elementos maximales del preorden parcial engendrado en el espacio de funciones de decisión.

En el capítulo dos, estudiamos dos temas distintos aunque ambos relacionados con el modelo general visto en el capítulo uno. De una parte consideramos el problema de como incluir en el modelo la información proporcionada por el experimento, lo que nos lleva al estudio de las propiedades del que hemos denominado "operador de Bayes", en relación con el concepto de información parcial que hemos adoptado, concluyendo que el modelo básico estudiado en el capítulo uno es consistente con la aplicación del teorema de Bayes. En segundo lugar, dentro de este mismo capítulo, tratamos el prob lema de la generalización del modelo estudiado en el capítulo uno, en el sentido de considerar que la información a priori puede representarse en varias etapas, lo que lleva a la consideración de modelos jerárquicos, los cuales se establecen como consecuencia de considerar que dicha información a priori puede representarse en etapas sucesivas mediante los llamados hiperparámetros; la teoría de los baricentros de Choquet es de gran utilidad y proporciona una reducción importante en el tratamiento de dichos modelos.

En la segunda parte tratamos una serie de cuestiones relacionadas con la existencia de soluciones "cuasi-bayesianas," y su conexión con los conceptos de admisibilidad y completitud a que dan lugar los preordenes parciales asociados al concepto de información parcial adoptado en la primera parte.

En el capítulo tres, y basandonos en los S-juegos generalizados, profundizamos en el análisis de los problemas de decisión con información parcial. Una contribución importante es el concepto de polaridad estricta, basado en un teorema de Klee y que permite investigar la estructura de los problemas de decisión con incertidumbre parcial, en relación con el concepto de soporte de una distribución.

En el capítulo cuatro, consideramos los problemas de existencia de soluciones admisibles y cuasi-bayesianas, trabajando unas veces con el conjunto de incertidumbre del problema de decisión, y otras con el polar restringido del mismo. En este capítulo la condición de convexidad para el conjunto de riesgo se sustituye por la de cono-convexidad (debida a Yu), condición esta que resulta mas adecuada a la hora de estudiar una gran cantidad de problemas de decisión, facilitando así la obtención de los resultados fundamentales.

Mi sincero y profundo agradecimiento a cuantos me alen-

taron y ayudaron durante estos largos años, en especial a mi Maestro el Prof. Fco. Javier Girón por su valiosa ayuda y dirección en la elaboración de esta Memoria, así como al Prof. Sixto Rios Garcia por su estímulo incansable, al Prof. Sixto Rios Insua por su incondicional ayuda y en fin a todos los amigos y compañeros del Departamento de Estadística.

PRIMERA PARTE

CAPITULO 1.

EL MODELO DE DECISION CON INFORMACION PARCIAL

1.1.1. Introducción y aspectos generales del problema

El modelo de decisión bayesiano para tratar los problemas de decisión en ambiente de incertidumbre parte de la hipótesis de que el decisor es capaz de especificar sobre el conjunto de los poribles estados de la Naturaleza una única distribución a priori, con lo cual el problema de decisión en ambiente de incertidumbre queda reducido a un problema de decisión en ambiente de riesgo con distribución sobre los estados de la Naturaleza perfectamente conocida y la decisión óptima en ausencia de más información sobre los estados de la Naturaleza, como podría ser la que proporcionase un determinado experimento, sería aquella que maximizara la utilidad esperada.

Este paradigma bayesiano se da rara vez en la realidad, tanto desde el punto de vista subjetivista puro, que implicaría que el decisor es capaz de ordenar lineal o completamente sus posibles decisiones y además de una manera coherente (es decir, sin violar los axiomas de comportamiento racional de la teoría subjetivista, lo que permitiría obtener su probabilidad personal), como desde el punto de vista lógico que aunque también conduce a una única distribución a priori, esta suele ser impropia (medidas de Haar invariantes por la derecha, principio o regla de Jeffreys, etc.) y, aunque suele corresponder a lo que suele denominarse "distribuciones a priori no informátivas" o de "referencia" (término más apropiado), puede que ésta sea incompatible con cierta información parcial que se tenga -p.e., una distribución uniforme en la recta real, cuando se sabe que ciertos valores fuera de un cierto intervalo son improbables- aunque esta cuestión enlaza con los problemas de sensibilidad de la distribución a posteriori o de las decisiones a cambios en la a priori a través del principio de la estimación estable (véase Edwards, Lindman and Savage, (1963) y Dickey, (1976)).

Un tercer enfoque, que padece de los mismos problemas que el lógico, es hacer depender la distribución a priori de referencia del experimento a través del concepto de información.

Todos estos enfoques son útiles a la hora de aplicar la metodología bayesiana a los problemas de inferencia y decisión, en especial en lo que se refiere a la sensibilidad de la distribución a posteriori a pequeños cambios respecto de la distribución a priori de referencia.

El hecho real de que en muchos problemas de decisión e inferencia estadística suele disponerse de cierta información sobre el espacio paramétrico -aunque muchas veces sea difícil precisar la naturaleza de dicha información o de si esta pudiera representarse mediante alguna medida de incertidumbre- exige una generalización de la metodología bayesiana.

El asignar a los sucesos intervalos de credibilidad, en vez de números reales, tiene una larga historia que se remonta a Keynes (1921) y Koopman (1940) y ha sido tratada posteriormente por numerosos autores.

Varios son los enfoques propuestos para tratar el problema de la información parcial cuando ésta es de carácter probabilístico. La mayoría son métodos heurísticos o híbridos, es decir combinación de métodos bayesianos y mínimax, como los propuestos por Hodges y Lehmann (1952), Schneeweiss (1964), Menges (1964, 1966), Blum y Rosenblatt (1967) y Kudo (1967), y que no responden a una justificación axiomática alguna.

Uno de los problemas básicos a resolver a la hora de proponer un modelo de decisión o inferencia que tenga en cuenta la información parcial presente en los problemas de decisión es, precisamente, dar un significado preciso a este concepto de información parcial.

En el caso bayesiano la información que se tiene sobre los estados de la Naturaleza viene especificada por una única distribución de probabilidad como consecuencia de los principios de coherencia bayesianos.

Si se adopta un punto de vista análogo al bayesiano, es decir, se parte de unos principios de coherencia modificados de forma que den cabida al concepto de incertidumbre parcial siguiendo la línea de Smith (1961) y su generalización posterior por Girón y Ríos (1981), se llega a que la incertidumbre o ignorancia parcial puede representarse por un conjunto convexo y cerrado de medidas de probabilidad sobre el espacio paramétrico de tal manera que las decisiones óptimas o no dominadas o eficientes, como a veces se las denomina, son los elementos maximales del preórden parcial engendrado en el espacio de funciones de decisión.

En esta memoria adoptaremos esta última interpretación y los capítulos que siguen se dedicarán a aspectos particulares de este modelo.

1.1.2. El modelo básico

Comenzaremos considerando un ejemplo sencillo de problema de decisión como base para la definición del modelo de decisión con información parcial.

Supongamos que $x_1, \ldots, x_n$ es una muestra aleatoria simple de una distribución normal $N(\theta, r)$ de precisión conocida r y media desconocida θ. Se trata de estimar θ utilizando como función de pérdida una función cuadrática $L(\theta, a) = (\theta - a)^2$, si no disponemos de más información estamos ante un clásico pro blema de decisión en ambiente de incertidumbre con exprimentación asociada. En estas condiciones la <u>solución</u> del problema consistiría en la clase de las posibles decisiones o estimadores admisibles. Como este conjunto suele ser demasiado grande suele recurrirse a considerar nuevos principios que restrinjan esta clase.

En el caso bayesiano la información acerca del parámetro vendrá dada por una distribución de probabilidad sobre los estados. Si se tiene en cuenta que la muestra es de un modelo normal, la distribución a priori no suele ser cualquiera sino que, generalmente, pertenecería, por razones de conveniencia matemática, etc. (véase Raiffa y Schlaiffer, 1961) a una familia conjugada, es decir, esta información en nuestro ejemplo podría suponerse normal de parámetros conocidos.

Desde el punto de vista práctico aparece el problema de la determinación precisa de la distribución a priori que plantea los dos interrogantes. 1º ¿se puede suponer normal? 2º en caso afirmativo ¿se pueden determinar con suficiente precisión los parámetros de dicha distribución?. Esto nos llevaría a realizar un análisis de la sensibilidad del modelo respecto de la distribución inicial, suponiendo el resto de las componentes del problema de decisión perfectamente conocidos.

Otra manera de expresar nuestro conocimiento parcial acerca del parámetro θ sería especificando una cierta clase de distribuciones de probabilidad candidatas a ser la verdadera distribución a priori. Restringiéndonos, como antes, al caso de distribuciones a priori conjugadas se podría suponer que la verdadera distribución a priori sería normal $N(\mu, \sigma^2)$ con (μ, σ^2) variando en una cierta región del espacio paramétrico $K^* \subset \mathbb{R} \times \mathbb{R}^+$.

Señalamos que el hecho de suponer que la posible distribución a priori pertenezca a la familia conjugada es ya una <u>in-</u>

formación parcial. Obsérvese, también, que esta información parcial sería compatible con la no normalidad al considerar distribuciones a priori mixturas de la familia $\{N(\mu, \sigma^2);\ (\mu, \sigma^2) \in K^*\}$, como veremos más adelante.

Como ejemplo, que analizaremos más adelante, de información parcial tomaremos el caso de que $K^* = \{(\mu, \sigma^2);\ \mu \in [\mu_0, \mu_1]\ ;\ \sigma^2 = \sigma_0^2\}$ que permitirá el estudio de la sensibilidad del problema a variaciones en la media de la distribución inicial, tomando simplemente $\mu_0 = \hat{\mu} - \varepsilon$ y $\mu_1 = \hat{\mu} + \varepsilon$.

El problema fundamental que se plantea es el de determinar el conjunto de decisiones óptimas cuando se dispone de tal información a priori. En nuestro ejemplo la manera directa de proceder sería determinar todos los estimadores de Bayes respecto de todas las posibles distribuciones a priori "compatibles" con la información a priori dada.

Así en nuestro ejemplo cualquier distribución a priori compatible con la información parcial dada por K^* sería una mixtura de estas representada por una hiperdistribución $H(\mu)$ definida sobre K^*. Es decir cualquier distribución a priori tendría una función de densidad $f(\theta\,;H)$ dada por

$$f(\theta; H) = \int_{[\mu_0,\mu_1]} n(\theta; \mu, \tau_0)\, dH(\mu) \qquad (1.1)$$

donde $n(\theta; \mu, \tau_0)$ representa la función de densidad de la normal de media μ y precisión $\tau_0 = (\sigma_0^2)^{-1}$.

Obsérvese que de hecho $H(\mu)$ es una distribución sobre el hiperparámetro μ que a su vez es el parámetro de la distribución a priori; es decir se trataría de un modelo jerárquico con dos jerarquías o etapas.

De acuerdo con lo anterior tendríamos que calcular todos los estimadores de Bayes respecto de todas las distribuciones a priori cuyas densidades vendrían dadas por (1.1). Como la función de pérdida es cuadrática, es bien sabido (véase p.e., De - Groot, (1970) o Ferguson, (1967)) que el estimador de Bayes es la - media de la distribución a posteriori.

Si designamos por m (H) la media de la distribución a - posteriori cuya densidad viene dada por (1.1) se tiene, por el teorema 1.2.12. del capítulo 2, que

$$m(H) = \int_{[\mu_0, \mu_1]} m(\mu)\, d H_x(\mu)$$

donde m (μ) es la media a posteriori de las distribuciones normales de densidad $n(\theta; \mu, \tau_0)$, es decir

$$m(\mu) = \frac{\tau_0 \mu + n r \bar{x}}{\tau_0 + n r} ,$$

y H_x (μ) es la hiperdistribución a posteriori (véase la generalización de la fórmula (2.5) a continuación del teorema 1.2.12.).

Si representamos por $\bar{\mu}_{H_x}$ la media de la distribución H_x (μ) se tendrá que

$$m(H) = \frac{\tau_0 \mu_{H_x} + n r \bar{x}}{\tau_0 + n r}$$

Obsérvese de esta fórmula que si dos hiperdistribuciones H y G tienen la misma media a posteriori, el estimador de - Bayes correspondiente a ambas es el mismo. En particular todos los estimadores de Bayes, respecto de las diferentes distribuciones a priori, están en el intervalo

$$\left[\frac{\tau_0 \mu_0 + n r \bar{x}}{\tau_0 + n r} , \frac{\tau_0 \mu_1 + n r \bar{x}}{\tau_0 + n r} \right]$$

El riesgo de Bayes correspondiente a una distribución a priori dada es, como se sabe, la varianza de la distribución a posteriori. Se representamos por V (μ) la varianza de la distribución a posteriori de la n $(\vartheta\ ;\mu,\ \tau_0/x)$ y por V (H) la varianza de la mixtura dada por (1.1), entonces se tiene que

$$V(H)=\int_{[\mu_0,\mu_1]}(V(\mu)+m\ (\mu)^2)\ d\,H_x\ (\mu)\ .$$

En nuestro caso se tiene que

$$V(\mu)=\frac{1}{\tau_0+nr}$$

de donde resulta

$$V(H)=\frac{\tau_0+nr+\tau_0\sigma^2(H_x)}{(\tau_0+nr)^2}\ ,$$

siendo σ^2 (H_x) la varianza de la hiperdistribución a posteriori. Esta última expresión representaría el riesgo de Bayes R(H) asociado a la hiperdistribución H (μ), que como se observa depende exclusivamente de la varianza de H_x.

Se representamos por K^{**} al conjunto de todas las hiperdistribuciones, es decir, $K^{**}=\mathcal{M}(K^*;\ \mathcal{B}(K^*))$, entonces el riesgo de Bayes es una función de K^{**} en $\mathbb{R}$ que, como veremos más - adelante es una función cóncava. Sería interesante calcular el mínimo y el máximo de esta función con el fín de comparar lo que podríamos denominar <u>riesgos inferior y superior</u> asociados al -- problema de decisión con información parcial dado por K^*.

Claramente se tiene que

$$R_*=\underset{H\in K^{**}}{\mathrm{Inf}}\ R(H)\ =\ \frac{1}{\tau_0+nr}$$

y este ínfimo se alcanza precisamente en las distribuciones extremas de K^{**}, es decir, en las distribuciones degeneradas. Por otra parte

$$R^*=\underset{H\in K^{**}}{\mathrm{Sup}}\ R(H)=\frac{1}{\tau_0+nr}+\frac{\tau_0^2\ (\mu_1-\mu_0)^2}{4\,(\tau_0+nr)^2}$$

que se alcanza cuando (y solamente cuando) σ^2 (H_x) es máxima y esto ocurre cuando H_x es una mixtura de μ_0 y μ_1, con probabilidad $\frac{1}{2}$.

Una medida de la incertidumbre o ignorancia parcial asociada a nuestro problema de estimación, medida en términos de pérdidas sería la diferencia

$$I(K^{**}) = R^* - R_* = \frac{\tau_0^2 (\mu_1 - \mu_0)^2}{4(\tau_0 + nr)^2} .$$

En el caso de tratar el problema de análisis de sensibilidad respecto a variaciones en la media de la distribución a priori, la fórmula anterior se reduciría a la siguiente, ya que $\mu_1 - \mu_0 = 2\varepsilon$

$$I(\varepsilon) = \frac{\tau_0^2 \varepsilon^2}{(\tau_0 + nr)^2}$$

En este ejemplo hemos adoptado un punto de vista directo al tratamiento de la incertidumbre parcial. El otro punto de vista posible sería el siguiente, que además serviría para cualquier problema de decisión con incertidumbre parcial, siempre que ésta, como en el ejemplo venga representada por un cierto conjunto, que denominaremos K^*, de distribuciones a priori. Como ya dijimos en la introducción la ignorancia parcial ha sido caracterizada por Girón y Ríos (1981), de esta manera.

Sea, en principio Ω el espacio paramétrico, dotado con una estructura de σ-álgebra $\mathcal{B}$, D un cierto conjunto o espacio de decisiones y R (θ , d) la función de riesgo que en principio suponemos $\mathcal{B}$-medible para toda $d \in D$ fija. Si representamos por $\Omega^* = \mathcal{M}(\Omega, \mathcal{B})$ al conjunto de todas las medidas de probabilidad σ-aditivas definidas sobre $(\Omega, \mathcal{B})$, la ignorancia parcial vendrá representada por un cierto subconjunto no vacío K^* de Ω^* , supongamos además que el riesgo de Bayes R (P; d) existe para toda $P \in K^*$ y toda $d \in D$.

Obsérvese que si, p.e., la función de riesgo es acotada, el riesgo de Bayes siempre existe para cualquier distribución a priori $P \in \Omega^*$.

Sea $d^* \in D$ una cierta regla de decisión, que de momento supondremos fija. El conjunto de aquellas reglas de decisión que tienen riesgo de Bayes menor o igual que d^* para una distribución a priori $P \in K^*$ vendrá dado por

$$M(P, d^*) = \{ d \in D;\ R(P; d) \leq R(P; d^*) \}$$

Si el razonamiento anterior lo hacemos para todas las $P \in K^*$, el subconjunto de D, definido por

$$M(K^*, d^*) = \bigcap_{P \in K^*} M(P; d^*) =$$

$$= \{ d \in D;\ R(P; d) \leq R(P; d^*) \text{ para toda } P \in K^* \},$$

estará formado por aquellas decisiones que son mejores que la d^* en presencia de la información parcial K^*. Dicho de otro modo, la información parcial, dada por K^*, engendra en D una relación de preórden parcial, que representaremos por $\underset{K^*}{\gtrsim}$, definida del modo siguiente: si $d, d' \in D$, entonces $d \underset{K^*}{\gtrsim} d'$ si y solo si $R(P, d) \leq R(P, d')$ para todo $P \in K^*$. La solución al problema, desde este segunto punto de vista sería el determinar los elementos maximales de D respecto de la relación de preórden dada por $\underset{K^*}{\gtrsim}$.

La relación que existe entre los dos enfoques del problema se estudia con detalle y profundidad en la segunda parte de la memoria para el caso particular de que Ω sea un espacio métrico compacto y $R(\vartheta, d)$ sea una función continua de ϑ para todo $d \in D$.

El estudio del caso más general, en particular la no compacidad del espacio paramétrico suele originar problemas como ya han señalado Sacks (1963) y Farrell en relación con la admisibilidad y el no bayesianismo de ciertos estimadores como es el caso de la media muestral $\bar{x}$ en el caso normal considerado al principio.

De paso señalamos tambien que $\bar{X}$ nunca puede ser admisible si - la información a priori viene representada por un conjunto K^* compacto de Ω^* (en el cual se considera la topología débil).

La explicación profunda de estos hechos reside en que - si Ω no es compacto, el dual del espacio de las posibles funciones de riesgo, aún supuestas acotadas, no está formado exclusivamente por medidas signadas σ-finitas. Si admitimos la posibilidad de incluir como distribuciones a priori las finitamente aditivas, la situación se aclara como han demostrado recientemente Heath y Sudderth (1978).

Pasamos a continuación a dar las definiciones formales de los modelos de decisión con información parcial asociada, -- precisando los conceptos que nos han aparecido en el examen del ejemplo con que abrimos el parágrafo.

Definición 1.1.1. Un problema de decisión con información parcial es una cuaterna (Ω , D; R, K^*) donde Ω es el espacio paramétrico, D un espacio de acciones o decisiones, R (θ , d) la función de riesgo y K^* un subconjunto no vacío de $\Omega^* = \mathcal{M}(\Omega, \mathcal{B})$, conjunto de medidas de probabilidad sobre (Ω , $\mathcal{B}$).

Por razones técnicas muchas veces se introducen hipótesis adicionales como la $\mathcal{B}$-medibilidad de R (θ , d) para todo $d \in D$, la acotación de la función de riesgo, y el que $\mathcal{B}$ incluya a los subconjuntos de la forma $\{\theta\}$, con $\theta \in \mathcal{B}$, con lo cual Ω se puede considerar como un subconjunto de Ω^* . identificando $\theta \in \Omega$ con la distribución de probabilidad δ_θ definida por

$$\delta_\theta(B) = \begin{cases} 1 \text{ si } & \theta \in B \\ 0 \text{ si } & \theta \notin B \end{cases}$$

El modelo definido anteriormente engloba el caso bayesiano cuando K^* se reduce a una única distribución de probabilidad y el caso de ignorancia total cuando $K^* = \Omega$.

Definición 1.1.2. Una decisión $\tilde{d} \in D$ se dice que es K-Bayes -- (o cuasi-bayes) si $\tilde{d}$ es Bayes respecto de alguna distribución a priori $P \in K^*$.

Al conjunto de las decisiones K^*-Bayes del problema -- $(\Omega, D; R; K^*)$ lo representaremos por $\mathcal{B}(D; K)$.

Definición 1.1.3. A la relación de preórden parcial definido en D, correspondiente al problema $(\Omega, D; R; K^*)$, dado por

$$d \succsim_{K^*} d' \text{ si y solo si } R(P, d) \leq R(P, d') \text{ para todo } P \in K^*$$

lo denominaremos preórden K^*-bayesiano o cuasi-bayesiano.

Definición 1.1.4. A los elementos maximales del conjunto $(D, \succsim_{K^*})$ los denominaremos decisiones K^*-admisibles y los representaremos por $\mathcal{A}(D; K^*)$.

Como hemos señalado anteriormente, la relación entre los conjuntos $\mathcal{B}(D, K^*)$ y $\mathcal{A}(D; K^*)$ depende de la naturaleza de - los elementos que constituyen el problema de decisión, y el capítulo 4. de esta memoria trata de este problema así como del - problema de la existencia de estas clases de reglas de decisión.

Desde el punto de vista teórico es conveniente considerar la relación $\succsim_{K^*}$ definida no en D sino en el espacio de -- las posibles funciones de riesgo, que denominaremos $\mathcal{R}(\Omega, \mathcal{B})$, donde $\mathcal{R}(\Omega, \mathcal{B})$ está formado por el conjunto de todas las funciones $\mathcal{B}$-medibles de $(\Omega, \mathcal{B})$ en $(\mathbb{R}, \mathcal{B}_1)$ donde $\mathcal{B}_1$ re-- presenta la σ-álgebra de los borelianos de la recta real. De - este modo a cada regla de decisión $d \in D$ le podemos asociar -- una función $\mathcal{B}$-medible $f_d(\vartheta) \in \mathcal{R}(\Omega, \mathcal{B})$ definidas por

$$f_d(\vartheta) = R(\vartheta, d)$$

La relación $\succsim_*$ definida en D puede entenderse a $\mathcal{R}(\Omega, \mathcal{B})$ definiendo

$$f \succsim_* g \quad \text{si y solo si} \quad \int_\Omega f(\vartheta)\, P(d\vartheta) \leq \int_\Omega g(\vartheta)\, P(d\vartheta) \qquad (1.2)$$

para todo $P \in K^*$, siempre y cuando las integrales sean convergentes. Esto ocurre p.e. si la función de riesgo es acotada. Por eso a partir de ahora, supondremos que $\mathcal{R}(\Omega, \mathcal{B})$ está formado por todas las funciones $\mathcal{B}$-medibles y acotadas de $(\Omega, \mathcal{B})$ en $(\mathbb{R}, \mathcal{B}_1)$.

<u>Definición 1.1.5.</u> Una relación de preórden $\gtrsim$ definida en $\mathcal{R}(\Omega, \mathcal{B})$ es un preórden <u>vectorial</u> ó <u>cónico</u> si $f \gtrsim g$ implica $f + h \gtrsim g + h$ para todo $h \in \mathcal{R}(\Omega, \mathcal{B})$ y $\alpha f \gtrsim \alpha g$ para todo $\alpha > 0$.

El teorema siguiente, de fácil demostración, nos dice que siempre que K^* sea un subconjunto de $\mathcal{M}(\Omega, \mathcal{B})$, entonces el preórden $\underset{K^*}{\gtrsim}$ definido antes es siempre cónico.

<u>Teorema 1.1.1.</u> El preórden definido en $\mathcal{R}(\Omega, \mathcal{B})$ por (1.2) es un preórden cónico compatible con el órden natural de $\mathcal{R}(\Omega, \mathcal{B})$.

Si definimos $K = \{ f \in \mathcal{R}(\Omega, \mathcal{B}); \int f(\vartheta)\, P(d\vartheta) \leq 0 \}$, se tiene que por ser $\mathcal{M}(\Omega, \mathcal{B})$ un subconjunto del espacio dual de $\mathcal{R}(\Omega, \mathcal{B})$, K es un cono convexo con vértice en el origen, cerrado, de $\mathcal{R}(\Omega, \mathcal{B})$, y tal que

$$f \underset{K^*}{\gtrsim} g \text{ si y solo si } f - g \in K \qquad (1.3)$$

Compárense estos resultados con teoremas 3.2.1. y 3.2.3. del capítulo 3.

De este modo hemos establecido que a cada subconjunto no vacío K^* de Ω^* le corresponde un cono K con vértice en el origen, convexo y cerrado que contiene a $K_0 = \{ f \in \mathcal{R}(\Omega, \mathcal{B}); f(\vartheta) \leq 0 \text{ para todo } \vartheta \in \Omega \}$, y no son todo el espacio, de modo que por (1.3) ambos establecen el mismo preórden en el conjunto de las funciones de riesgo.

Si procedemos de manera inversa, es decir, consideramos conos K en $\mathcal{R}(\Omega, \mathcal{B})$ que satisfacen las propiedades anteriores <u>no podemos</u>, en general, asegurar la existencia de un subconjunto K^* de Ω^* para el cual se cumpla la relación (1.3).

El capítulo 3. de esta memoria da una solución satisfactoria a este problema de caracterizar preórdenes cuasi-bayesianos imponiendo ciertas restricciones a Ω y a la función de -- riesgo y de él se deduce que estos resultados no se pueden generalizar mucho más.

El teorema que sigue es una generalización de la primera parte del teorema 3.2.7., cuya demostración es trivial. La - segunda parte del teorema no se puede, sin embargo, generalizar facilmente.

Su generalización nos llevaría a considerar la convergencia débil * en el espacio dual $\mathcal{R}^*(\Omega, \mathcal{B})$ y que por lo tanto, entrañaría la consideración de medidas de probabilidad finitamente aditivas.

<u>Teorema 1.1.2.</u> K^* y co (K^*) generan en $\mathcal{R}(\Omega, \mathcal{B})$ el mismo - preórden.

Ni siquiera en el caso de que Ω fuese un espacio topológico y $\mathcal{B}$ la σ-álgebra de Borel, y en $\mathcal{M}(\Omega, \mathcal{B})$ considerásemos la topología débil se podría concluir que los preórdenes generados por K^* y $\overline{K^*}$ (donde la operación de clausura se entiende respecto de la topología débil) fuesen idénticos.

En el resto del capítulo supondremos que Ω es un espacio métrico, $\mathcal{B}$ la σ-álgebra de Borel y $R(\theta, d)$ continua y acotada en $\theta \in \Omega$ para todo $d \in D$.

Si denominamos por $R(P, d)$ el riesgo de Bayes correspondiente a la distribución a priori $P \in \Omega^*$ y a la decisión $d \in D$, se tiene de las hipótesis anteriores que $R(P, d)$ es una función continua de Ω^* dotado de la topología débil. Por lo tanto el -

riesgo de Bayes R (P) asociado al problema de decisión (Ω, D; R), definido por

$$R(P) = \operatorname*{Inf}_{d \in D} R(P, d)$$

es una función semicontinua superiormente y cóncava, como puede comprobarse facilmente (véase, p.e., De Groot pp. 125-126).

Ahora si K^* es la información parcial asociada al problema de decisión (Ω, D; R), siendo K^* convexo, podemos definir

<u>Definición 1.1.6.</u> A la cantidad $R^*(K^*)$ definida por

$$R^*(K^*) = \operatorname*{Sup}_{P \in K^*} R(P)$$

se la denomina <u>riesgo superior</u> asociado al problema de decisión con información parcial (Ω, D; R; K^*).

<u>Definición 1.1.7.</u> Se denomina <u>riesgo inferior</u> asociado al problema de decisión con información parcial (Ω, D; R; K^*) a la cantidad $R_*(K^*)$ definida por

$$R_*(K^*) = \operatorname*{Inf}_{P \in K^*} R(P)$$

Respecto de la última definición tiene importancia el siguiente lema (véase la demostración en Bourbaki (1966), p.106).

<u>Lema 1.1.1.</u> Si K^* es convexo y compacto, entonces el mínimo del riesgo de Bayes R (P) se alcanza en un punto extremo de K^*.

Una medida de la incertidumbre parcial asociada al problema de decisión (Ω, D; R; K), tomada en términos de riesgo vendrá dada por la diferencia entre el riesgo superior y el inferior.

Definición 1.1.8. Para todo problema de decisión en ambiente de incertidumbre (Ω, D; R) denominaremos riesgo de la incertidumbre parcial K^* a la cantidad I (K^*) definida por

$$I(K^*) = R^*(K^*) - R_*(K^*).$$

Obsérvese que en el caso bayesiano, es decir cuando -- $K^* = \{P\}$, siempre se tiene que $I(\{P\}) = 0$, mientras que $I(\ .\ )$ es monótona respecto de la relación de inclusión de modo que si $K^* \subset K^{*\prime}$ se tiene que $I(K^*) \leq I(K^{*\prime})$, con lo cual el máximo de la función $I(\ .\ )$ se alcanza cuando tenemos incertidumbre total, es decir, cuando $K^* = \Omega^*$.

La idea de realizar un experimento es precisamente la - de reducir el riesgo de la incertidumbre parcial que, en condiciones bastante generales, convergerá a cero cuando el tamaño - muestral se haga arbitrariamente grande.

Para terminar este capítulo introductorio vamos a considerar el papel de la experimentación cuando esta se incorpora - al modelo de decisión con información parcial. En este capítulo tratamos solamente el análisis en forma normal del modelo mientras el segundo capítulo de la memoria se dedica al estudio detallado del análisis en forma extensiva, y por consiguiente, su relación con la metodología bayesiana.

Sea $\mathcal{E} = (\mathcal{X}, \mathcal{A}; P)$ un experimento de dimensión fija tal que la familia $\{P_\theta\ ;\ \theta \in \Omega\}$ está dominada por una medida σ-finita $\nu(x)$ respecto de la cual sus densidades vienen dadas por

$$f(x/\theta) = \frac{d\,P_\theta(x)}{d\,\nu(x)} .$$

sobre la función de densidad $f(x/\theta)$ hacemos la siguiente hipótesis (compárese con la hipótesis 1.2.1. del capítulo 2.).

Hipótesis 1.1.1. $f(x/\theta)$ es una función contínua de $\theta \in \Omega$ para casi todo $x \in \mathcal{X}$ (respecto de $\nu(x)$).

El estudio en forma normal del problema de decisión original (Ω , D; R; K^*) se realiza simplemente agrandando el espacio de decisiones y, por consiguiente, extendiendo la función de riesgo a estas nuevas decisiones, dejando el resto de los componentes del problema de decisión invariantes.

Como es usual una regla de decisión, δ , es una aplicación $\mathcal{B}$-medible de ($\mathcal{X}$, $\mathcal{B}$) en (D, $\mathcal{D}$) donde $\mathcal{D}$ es una cierta σ-álgebra de subconjuntos de D que debe incluir entre sus elementos a los del tipo $\{d\}$, a fin de poder considerar el espacio de las decisiones terminales D como un subconjunto del nuevo espacio de decisiones δ , que representaremos por Δ .

La forma normal del problema de decisión con información parcial (Ω , D; R; K^*) es simplemente (Ω , Δ ; R; K^*) donde la función de riesgo se define

$$R(\vartheta, \delta) = \int_{\mathcal{X}} R(\vartheta, \delta(x))\, f(x/\vartheta)\, d\mu(x) \qquad (1.4)$$

Sería conveniente que la nueva función de riesgo fuese continua y acotada, como la inicial, con el fín de poder considerar el preórden $\underset{K^*}{\succsim}$ definido en $\mathcal{R}$ (Ω , $\mathcal{B}$). La condición impuesta a $f(x/\vartheta)$ anteriormente nos garantiza este resultado.

<u>Teorema 1.1.3.</u> Si la función de riesgo $R(\vartheta, d)$ es continua y acotada en ϑ para todo $d \in D$ y $f(x/\vartheta)$ cumple las hipótesis 1.1.1., entonces $R(\vartheta, \delta)$ es una función contínua y acotada de ϑ para todo $\delta \in \Delta$.

<u>Demostración.</u> La acotación es clara de la definición (1.4). Para demostrar la continuidad consideremos la diferencia

$$|R(\vartheta,\delta) - R(\vartheta_0,\delta)| \leq \left| \int R(\vartheta,\delta(x))\, f(x/\vartheta)\, d\mu(x) - \int R(\vartheta_0,\delta(x))\, f(x/\vartheta)\, d\mu(x) \right. \leq$$

$$\leq \int |R(\vartheta,\delta(x)) - R(\vartheta_0,\delta(x))|\, f(x/\vartheta)\, d\mu(x) + \left| \int R(\vartheta_0,\delta(x))(f(x/\vartheta) - f(x/\vartheta_0)\, d\mu(x) \right| \leq$$

$$\leq \int |R(\vartheta,\delta(x)) - R(\vartheta_0,\delta(x))|\, f(x/\vartheta)\, d\mu(x) + \int |R(\vartheta_0,\delta(x)||f(x/\vartheta) - f(x/\vartheta_0)|\, d\mu(x).$$

El primer miembro converge hacia cero cuando $\theta \longrightarrow \theta_o$, mientras que el segundo, por estar acotada $|R(\theta_o, \delta(x))|$ y por el teorema de Scheffé (1947) de que la convergencia puntual implica la convergencia en media, también converge hacia cero.

Una vez incorporada la información al modelo, se pueden calcular de nuevo las clases de estrategias K^*-Bayes, $\mathcal{B}(\Delta\ ;\ K^*)$ y la clase de decisiones K^*-admisibles $\mathcal{A}(\Delta\ ;\ K^*)$ del mismo modo que en la situación original.

CAPITULO 2.

EXTENSIONES DEL MODELO DE DECISION CON INFORMACION PARCIAL: EXPERIMENTACION. MODELOS JERARQUICOS.

1.2.1. Introducción.

En este capítulo tratamos dos problemas distintos relacionados con el modelo general expuesto en el primer capítulo, que sin embargo guardan cierta relación. En el apartado 2.2. abordamos el problema de cómo incluir la información proporcionada por un experimento en nuestro modelo, lo que conduce de modo natural al estudio de las propiedades de lo que hemos llamado "operador de Bayes" en relación con el concepto de información parcial tal como la hemos considerado en el primer capítulo. De este estudio se concluye que el modelo presentado en el primer capítulo es consistente con la aplicación del teorema de Bayes en casos más generales que los señalados por la teoría (véase concretamente Girón y Ríos, (1981)), siempre que se disponga de una formulación suficientemente general del teorema de Bayes, como la que damos en el parágrafo 2.2.

En el parágrafo 2.3. se generaliza el modelo del primer capítulo, y por consiguiente, el modelo bayesiano, al caso de considerar que la información a priori puede representarse en varias etapas, lo cual entraña la consideración de hiperparámetros en el modelo original, que se distribuyen de acuerdo con una cierta distribución, siendo a su vez estas distribuciones los nuevos parámetros en la etapa siguiente del modelo jerárquico. Esto como veremos, conllevará la consideración de espacios de medidas definidas a su vez sobre espacios de medidas, y así sucesivamente. Afortunadamente muchas de las propiedades importantes de los espacios de medidas son heredades en las sucesivas etapas del modelo jerárquico. Otra reducción importante a la hora de considerar el modelo jerárquico, se consigue mediante la utilización de la teoría de los baricentros de Choquet (1960), que será empleada también en otro problema distinto, pero relacionado con él, al final del tercer capítulo.

La conclusión más importante de este apartado es que, desde el punto de vista teórico, no es necesario considerar en los modelos jerárquicos más de una etapa.

Una posible, e importante, aplicación de esta teoría, no considerada en esta memoria, podría ser al campo de la estadística no paramétrica o más precisamente al estudio de la sensibilidad en los modelos paramétricos, cuando en estos se consideran mixturas de modelos.

También se dan en esta sección una generalización de la fórmula de Bayes a hiperdistribuciones que se consigue mediante la introducción del concepto de verosimilitud generalizada.

1.2.2. El papel de la experimentación en los problemas de decisión con información parcial.

En toda esta sección supondremos que el espacio paramétrico Ω es un espacio métrico separable. Antes de considerar la inclusión del experimento, que supondremos de dimensión fija -es decir, no secuencial- en el modelo original vamos a estudiar una serie de propiedades del espacio de las distribuciones sobre Ω .

Por Ω^* representaremos el conjunto de todas las distribuciones de probabilidad regulares sobre $(\Omega, \mathcal{B})$, donde $\mathcal{B}$ es la σ-álgebra de Borel de Ω . A veces emplearemos la notación alternativa $\mathcal{M}(\Omega, \mathcal{B})$ para designar Ω^* , y a sus elementos los representaremos por P, Q, R, ...

A Ω^* lo dotamos de la topología débil, que representaremos por (Ω^*, w) definida a continuación.

Definición 1.2.1. La red $\{P_\alpha\} \in \Omega^*$ converge débilmente a $P \in \Omega^*$, lo representaremos, $P_\alpha \xrightarrow{w} P$, si

$$\lim_{\alpha} \int f \, d \, P_{\alpha} = \int f \, d \, P.$$

para todo $f \in CA(\Omega)$, donde $CA(\Omega)$ es el espacio de todas las funciones reales, contínuas y acotadas definidas sobre Ω .

Otras caracterizaciones de la topología débil pueden -- verse en Billingsley (1968), th. 2.1. y Apéndice III y en Parthasarathy, (1967); th. 6.1., p. 40).

Aparte de la estructura topológica (Ω^*, w) posee también estructura de espacio de mixtura, consecuencia de considerar a Ω^* como un subconjunto de un espacio de medidas señaladas - σ -aditivas y finitas sobre (Ω , $\mathcal{B}$), definiendo

$$[\alpha P + (1-\alpha) Q](A) = \alpha P(A) + (1-\alpha) Q(A)$$

para todo $\alpha \in [0,1]$, $P, Q \in \Omega^*$ y $A \in \mathcal{B}$.

Lo anterior se puede generalizar a mixturas finitas, definiendo

$$\left(\sum_{i=1}^{k} \alpha_i P_i \right)(A) = \sum_{i=1}^{k} \alpha_i P(A)$$

donde $\sum_{i=1}^{n} \alpha_i = 1$, $\alpha_i \geq 0$; $i = 1, 2, \cdots, n$; $P_i \in \Omega^*$ y $A \in \mathcal{B}$

Definición 1.2.2. $K^* \subset \Omega^*$ es convexo si es cerrado por mixturas finitas.

Definición 1.2.3. La envoltura convexa de un conjunto K^*, que representarmos por $co(K^*)$, es el mínimo conjunto convexo que contiene a K^* o, equivalentemente, el conjunto formado por todas -- las mixturas finitas de elementos de K^*.

Concepto importante en ciertas aplicaciones (véase el capítulo 3.) no sólo de la Teoría de la Probabilidad, sino también de la Teoría de la Decisión es el de soporte de una medida de -- probabilidad.

Teorema 1.2.1. Sea Ω un espacio métrico y separable, y sea $P \in \Omega^*$. Entonces existe un único conjunto cerrado C_p tal que: (i) $P(C_p) = 1$, (ii) si D es cualquier otro conjunto cerrado, tal que $P(D) = 1$, entonces $C_p \subset D$. Además C_p está caracterizado por el conjunto de todos los puntos $\vartheta \in \Omega$ tales que $P(O) > 0$ para todo abierto O que contenga a ϑ.

La demostración puede verse en Parthasarathy (1967, pp. 28-29).

Este teorema sugiere como definición de soporte de una medida de probabilidad la siguiente

Definición 1.2.4. Al conjunto cerrado C_p del teorema 1.2.1. se le denomina espectro o soporte de P y lo representaremos por Sop (P).

El concepto de soporte de una distribución de probabilidad puede extenderse a espacios topológicos arbitrarios, como puede verse, p.e., en Ardanuy (1979). No obstante, para casi todas las aplicaciones a la inferencia paramétrica y no paramétrica, es suficiente el considerar espacios métricos.

Con relación a la operación de mixtura, -operación que se extiende de manera obvia al caso de mixturas numerables- es interesante el teorema siguiente, que como se verá en el parágrafo 2.3., teorema 1.2.12. es un caso particular de éste.

Teorema 1.2.2. Si $\alpha_i > 0$ para i =1,2,..., entonces

$$\mathrm{Sop}\left(\sum_{i=1}^{\infty} \alpha_i P_i\right) = \overline{\bigcup_{i=1}^{\infty} \mathrm{Sop}\,(P_i)}$$

Demostración. Como $\alpha_i > 0$, se tiene que para todo i, $\mathrm{sop}(P_i) \subset \mathrm{sop}\left(\sum_{i=1}^{\infty} \alpha_i P_i\right)$ y, por lo tanto $\bigcup_{i=1}^{\infty} \mathrm{sop}\,(P_i) \subset \mathrm{sop}\left(\sum_{i=1}^{\infty} \alpha_i P_i\right)$. Como el soporte, por definición, es cerrado, se tiene que

$$\overline{\sum_{i=1}^{\infty} \mathrm{sop}\,(P_i)} \subset \mathrm{sop}\left(\sum_{i=1}^{\infty} \alpha_i P_i\right)$$

Recíprocamente, supongamos que $\vartheta \in$ sop $(\sum_{i=1}^{\infty} \alpha_i P_i)$. Si $\vartheta \notin \overline{\bigcup_{i=1}^{\infty} \text{sop}(P_i)}$, existirá un abierto O que contiene a ϑ, tal que $O \cap \overline{\bigcup_{i=1}^{\infty} \text{sop}(P_i)} = \emptyset$. Esto implicaría que para todo i, $O \cap \text{sop}(P_i) = \emptyset$ y esto, a su vez, que $P_i\ (O) = 0$, y de aquí finalmente que

$$\sum_{i=1}^{\infty} \alpha_i \ P_i\ (O) = 0$$

lo que contradiría el que $\vartheta \in \text{sop}(\sum_{i=1}^{\infty} \alpha_i P_i)$.

A partir de ahora representaremos por F^* el subconjunto de Ω^* formado por todas las distribuciones cuyo soporte es finito, es decir, $P \in F^*$ si y todo si,

$$P = \sum_{i=1}^{n} \alpha_i \delta_{\vartheta_i}$$

con $n \in \mathbb{N}$, $\alpha_i > 0$, $\sum_{i=1}^{n} \alpha_i = 1$ y δ_{ϑ_i} es la medida de probabilidad concentrada en ϑ_i, decir,

$$\delta_{\vartheta_i}(B) = \begin{cases} 1 & si \quad \vartheta_i \in B \\ 0 & si \quad \vartheta_i \notin B \end{cases}$$

para todo $B \in \mathcal{B}$.

Claramente sop (P) $= \{\vartheta_1, \dots, \vartheta_n\}$.

Si, como es costumbre, identificamos $\delta_\vartheta \in \Omega^*$ con $\vartheta \in \Omega$ entonces Ω puede ser considerado como un subconjunto de Ω^*. En particular $\Omega \subset F^*$.

La propiedad más interesante, de F^* es que es w-denso en Ω^*, propiedad que, según señala Billingsley (1968, th.4, p. 237) es válida para espacios métricos no separables. Otra demostración puede verse en Parthasarathy (1967; th. 6.3., pp. 44-45).

<u>Teorema 1.2.3.</u> $\overline{F^*} = \Omega^*$

Este resultado nos va a ser útil a la hora de probar -- otros resultados, más adelante.

A continuación damos una propiedad importante del conjunto de todas las distribuciones de probabilidad sobre $(\Omega, \mathcal{B})$ cuyo soporte es todo el espacio Ω . A dicho conjunto lo denominaremos Ω^{*+}, según la notación del capítulo 3., y se define por

$$\Omega^{*+} = \{ P \in \Omega^{*} ; \quad sop(P) = \Omega \}$$

El teorema que damos a continuación es más general que los 2.3.13. y 2.3.14. en el sentido de que no supone que Ω sea compacto. Sin embargo, no cubre los demás casos del teorema 2.3.14.

Teorema 1.2.4. Si Ω es un espacio métrico separable, entonces $\overline{\Omega^{*+}} = \Omega^{*}$.

Demostración. Claramente $\overline{\Omega^{*+}} \subset \Omega^{*}$. Para demostrar la inclusión contraria supongamos que $P \in \Omega^{*}$ y que, sin embargo $P \notin \overline{\Omega^{*}}$.

Vamos primero a demostrar que $\Omega \subset \overline{\Omega^{*+}}$. En efecto, por ser Ω separable existirá un conjunto numerable $\Omega_0 \subset \Omega$ que es denso en Ω . Sea $\Omega_0 = \{ \theta_n \, ; \, n \in \mathbb{N} \}$. Entonces para todo $\theta_0 \in \Omega$ existirá una subsucesión $\{\theta_{n_k}\} \longrightarrow \theta_0$, con $\theta_{n_k} \in \Omega_0$. Sea $P_i = \sum_{n=1}^{\infty} p_n^i \, \delta_{\theta_n}$ con $p_n^i > 0$ para todo $i, n = 1, 2, \ldots$ y $\sum_{n=1}^{\infty} p_n^i = 1$, tales que $\lim_{\substack{i \to \infty \\ n_k \to \infty}} p_{n_k}^i = 1$. Entonces se tiene que por una parte $sop(P_i) = \Omega$, por ser Ω_0 denso en Ω , y por otro lado $\{P_i\} \rightarrow \delta_{\theta_0}$, cuando $i \rightarrow \infty$, con lo cual $\delta_{\theta_0} \in \overline{\Omega^{*+}}$.

Como $\{P\}$ y $\overline{\Omega^{*+}}$ son conjuntos convexos, cerrados y disjuntos, existirá una función continua y acotada $f \in CA(\Omega)$ y una constante c tales que

$$\int f \, dP < c$$
$$\int f \, dQ > c \quad \text{para todo} \quad Q \in \overline{\Omega^{*+}}$$

como $\delta_\theta \in \overline{\Omega^{*+}}$ para todo $\theta \in \Omega$, como acabamos de probar, se deduciría, de la segunda de las desigualdades que

$f(\theta) = \int_{\Omega} f \, d\delta_{\theta} > c$ para todo $\theta \in \Omega$ y, como $P \in \Omega^*$ y la integral es monótona, resultaría que $\int f \, dP > c$,contradiciendo la primera desitualdad.

Este teorema está emparentado con lo que Lindley (1981) ha denominado <u>ley de Cromwell</u> y que en esencia consiste en que, aparte de los axiomas de coherencia, la probabilidad subjetiva prescrita por dichos axiomas ha de ser tal que no asigne probabilidad nula a ninguno de los estados de la Naturaleza. Esta observación expresada en términos rigurosos afirma que las distribuciones a priori "coherentes" han de tener como soporte todo el espacio paramétrico. El Teorema anterior nos asegura que cualquier distribución a priori puede ser aproximada, en condiciones muy generales, tanto como queramos (en el sentido de la convergencia débil) por distribuciones cuyo soporte sea todo el espacio paramétrico.

<u>Definición 1.2.5.</u> Un subconjunto K^* de Ω^* , se dice que es <u>(uniformemente) ajustado</u> si para todo $\varepsilon > 0$ existe un compacto K_{ε} tal que $P(K_{\varepsilon}) > 1-\varepsilon$ para todo $P \in K^*$.

La relación entre subconjuntos ajustados de medidas de probabilidad y los subconjuntos de Ω^* relativamente compactos en la topología débil, viene dada por el teorema de Prohorov que damos a continuación y cuya demostración puede verse en los libros de Parthasarathy (1967) y Billingsley (1968).

<u>Teorema 1.2.4a.</u> Si K^* es uniformemente ajustado, entonces es relativamente compacto, es decir, $\overline{K^*}$ es compacto.

<u>Teorema 1.2.4b.</u> Sea Ω completo y separable. Si K^* es relativamente compacto, entonces K^* es uniformemente ajustado.

Como consecuencia del teorema de Prohorov se tiene el

<u>Teorema 1.2.5.</u> Sea Ω un espacio métrico, completo y separable. Si K^* es relativamente compacto, entonces $co(K^*)$ también lo es.

<u>Demostración.</u> Por el teorema 1.2.4b. para todo $\varepsilon > 0$ existe un compacto K_ε tal que $P(K_\varepsilon) > 1-\varepsilon$ para todo $P \in K^*$. Sea ahora -- $Q \in co(K^*)$, es decir, $Q = \sum_{i=1}^{n} \alpha_i P_i$, con $\alpha_i > 0$, $\sum_{i=1}^{n} \alpha_i = 1$, y $P_i \in K^*$. Entonces se tiene que

$$Q(K_\varepsilon) = \sum_{i=1}^{n} \alpha_i P_i(K_\varepsilon) > \sum_{i=1}^{n} \alpha_i (1-\varepsilon) = 1-\varepsilon$$

cualquiera que sea $Q \in co(K^*)$. Entonces $co(K^*)$ es uniformemente ajustado y, por el teorema 1.2.4a., $co(K^*)$ es relativamente compacto.

Como corolario de este teorema se tiene el análogo de - los teoremas de Mazur y de Krein- Šmulian para espacios de Banach.

<u>Corolario 1.2.1.</u> Sea Ω métrico, completo y separable. Si K^* es compacto, entonces $\overline{co}(K^*)$ también lo es.

Antes de incluir la experimentación en el modelo de decisión con información parcial, vamos a estudiar las propiedades del operador de Bayes, considerado éste como el que transforma probabilidades a priori en probabilidades a posteriori, una vez observado el resultado del experimento. Para ello será necesario una formulación suficientemente general del teorema de Bayes, - como la dada, p.e., por De Robertis y Hartigan (1979).

<u>Teorema 1.2.6. (de Bayes)</u> Sea $\{P_\theta;\ \theta \in \Omega\}$ una familia de medidas de probabilidad definidas en $(\mathcal{X}, \mathcal{A})$, dominada por una medida σ-finita ν y sea $f(x/\theta) = \frac{d P_\theta}{d \nu}(x)$. Si P es una distribución a priori sobre $(\Omega, \mathcal{B})$ y si $f(x/\theta)$ es $\mathcal{A} \times \mathcal{B}$ medible, entonces la única medida de probabilidad en $(\mathcal{X} \times \Omega, \mathcal{A} \times \mathcal{B})$ - que tiene como marginal P y como distribución condicionada regular $\{P_\theta;\ \theta \in \Omega\}$ dado $\mathcal{B}$ es la medida de probabilidad Q con densidad $f(x/\theta)$ respecto de $\nu \times P$. Además la distribución condicional regular de Q dado $\mathcal{A}$, tiene densidad

$$\frac{f(x/\theta)}{\int_\Omega f(x/\theta)\, dP(\theta)} \qquad (2.1)$$

respecto de P.

A dicha probabilidad a posteriori la representaremos por P_x. Nótese de (2.1) que P_x es absolutamente contínua respecto de P para casi todo $x \in \mathcal{X} (\nu)$.

A partir de ahora supondremos fijo y conocido el resultado experimental $x \in \mathcal{X}$. A la aplicación de Ω^* en Ω^* que hace corresponder a cada distribución a priori P la distribución a posteriori P_x la denominaremos <u>operador Bayes</u> y la representaremos por $B_x (P) = P_x$.

Es a las pro-piedades de este operador, en relación con el concepto de información parcial y con el de soporte de una medida de probabilidad, a lo que vamos a dedicar el resto de esta sección.

Ciertas propiedades del operador, en particular las referentes a la continuidad débil de B_x y las referentes al soporte de la distribución a posteriori, van a depender básicamente del comportamiento de la función de verosimilitud. Otras propiedades, como las de tipo algebráico, son completamente generales.

No tratamos en esta memoria, sin embargo, el caso particular de que la familia de distribuciones a priori esté dominada por una medida σ-finita, caso en el que pueden obtenerse resultados más precisos. No obstante los resultados presentados son algo más generales que los dados por Girón y Ríos (1981).

Obsérvese que en la fórmula de Bayes, (2.1) el denominador no puede anularse si la fórmula ha de tener sentido, es decir $f(x/P) \neq 0$. Esto supone que $f(X/\theta)$ no puede ser idénticamente nula para aquellos $\theta \in$ sop (P), hipótesis tácita que asumiremos de ahora en adelante. Si representamos por $\Omega_x = \{\theta ; f(x/\theta)>0\}$ la observación anterior equivale a que sop $(P) \cap \Omega_x \neq \emptyset$ y en este caso, único en el que tiene sentido la fórmula (2.1), diremos que la <u>información muestral y la a priori son compatibles</u>.

El lema que sigue es un caso particular de la fórmula de Bayes generalizada que se dará en el parágrafo 2.3. al tratar de los modelos jerárquicos, pero lo damos aquí, pues permite dar - una demostración sencilla del teorema 1.2.7.

<u>Lema 1.2.1.</u> Si $P = \sum_{i=1}^{n} \alpha_i P^i$ con $\alpha_i \geq 0$, $i = 1,\dots,n$ y $\sum_{i=1}^{n} \alpha_i = 1$, entonces

$$P_x = B_x(P) = \sum_{i=1}^{n} \alpha_i(x)\, B_x(P^i) = \sum_{i=1}^{n} \alpha_i(x)\, P_x^i$$

donde $\alpha_i(x) = \dfrac{\alpha_i f(x/P^i)}{f(x/P)} \geq 0$ son tales que

$$\sum_{i=1}^{n} \alpha_i(x) = 1 , \quad f(x/P^i) = \int_\Omega f(x/\theta)\, dP^i(\theta)$$ y

$$f(x/P) = \int_\Omega f(x/\theta)\, dP(\theta) = \sum_{i=1}^{n} \alpha_i\, f(x/P^i) .$$

<u>Demostración.</u> De la fórmula (2.1) se tiene que para todo $B \in \mathcal{B}$

$$P_x(B) = \frac{\int_B f(x/\theta)\, dP(\theta)}{f(x/P)} = \frac{\sum_{i=1}^{n} \alpha_i \int_B f(x/\theta)\, dP^i(\theta)}{f(x/P)} =$$

$$= \frac{\sum_{i=1}^{n} \alpha_i f(x/P^i)\, P_x^i(B)}{f(x/P)} = \sum_{i=1}^{n} \alpha_i(x)\, P_x^i(B).$$

De las definiciones se tiene que $\alpha_i(x) \geq 0$ y que $\sum_{i=1}^{n} \alpha_i(x) = 1.$

<u>Teorema 1.2.7.</u> B_x transforma conjuntos convexos en conjuntos convexos, es decir, si K^* es convexo $K_x^* = B_x(K^*)$ también lo es. Además B_x transforma distribuciones extremas en distribuciones extremas, es decir, si P es una distribución extrema de K^* , -- $P_x = B_x(P)$ es una distribución extrema de $K_x^* = B_x(K^*)$.

<u>Demostración.</u> Sean P_x, $Q_x \in K_x^*$ y sea $\lambda \in [0,1]$. Hay que probar que $\lambda P_x + (1-\lambda) Q_x \in K_x^*$, o lo que es lo mismo que exis te un $R \in K^*$ tal que $B_x(R) = \lambda P_x + (1-\lambda) Q_x$. Por definición existen $P,Q \in K^*$ tales que $P_x = B_x(P)$ y $Q_x = B_x(Q)$. Si ahora definimos

$$\alpha = \frac{\lambda f(x/Q)}{\lambda f(x/Q) + (1-\lambda) f(x/P)}$$

se tiene que $\alpha \in [0,1]$. Por ser K^* convexo se tiene que $\alpha P + (1-\alpha) Q \in K^*$. Si ahora definimos $R = \alpha P + (1-\alpha) Q \in K^*$, por el lema 1.2.1. se tiene que

$$B_x(R) = \lambda P_x + (1-\lambda) Q_x$$

Para demostrar la segunda parte sea $P \in K^*_e$, donde K^*_e designa el conjunto de puntos extremos de K^*, y supongamos que $P_x = B_x(P)$ no fuese extremo de K^*_x . Entonces existirían $Q_x, R_x \in K^*$ y $\lambda \in (0,1)$ con $Q_x \neq R_x$ tales que

$$P_x = \lambda Q_x + (1-\lambda) R_x$$

Por el lema 1.2.1. existiría un α definido por

$$\alpha = \frac{\lambda f(x/R)}{\lambda f(x/R) + (1-\lambda) f(x/Q)} \in (0,1)$$

donde $Q, R \in K^*$ son tales que $B_x(Q) = Q_x$, $B_x(R) = R_x$ tal que $P = \alpha Q + (1-\alpha) R$ con $R \neq Q$, lo cual implicaría que P no sería punto extremo de K^*.

Como corolario de este teorema tenemos el siguiente resultado

<u>Corolario 1.2.2.</u> El operador de Bayes B_x conmuta con la operación de envoltura convexa, es decir,

$$B_x(co(K^*)) = co(B_x(K^*)).$$

<u>Demostración.</u> Por ser $K^* \subset co(K^*)$ se tiene que $B_x(K^*) \subset B_x(co(K^*))$ y como por el teorema anterior $B_x(co(K^*))$ es convexo, de aquí que $co(B_x(K^*)) \subset B_x(co(K^*))$.

Recíprocamente sea $P_x \in B_x(co(K^*))$. Entonces existirá un $P \in co(K^*)$ tal que $P_x = B_x(P)$. Por ser $P \in co(K^*)$, existirán $P^1, \dots, P^n \in K^*$ y $\alpha_i > 0$, $i = 1,\dots,n$ con $\sum_{i=1}^{n} \alpha_i = 1$ tales que $P = \sum_{i=1}^{n} \alpha_i P^i$

Por el lema 1.2.1.

$$P_x = B_x\ (P) = \sum_{i=1}^{n} \alpha_i\ (x)\ P_x^i$$

donde $P_x^i \in B_x(K^*)$, luego $P_x \in co\ (B_x\ (K^*))$. De las dos inclusiones se sigue el corolario.

El teorema que sigue establece que, bajo restricciones muy suaves en la función de verosimilitud f (x/ϑ), B_x es un operador continuo de (Ω^*, w) en sí mismo.

<u>Hipótesis 1.2.1.</u> La verosimilitud f (x/ϑ) es una función continua y acotada de ϑ .

<u>Definición 1.2.6.</u> Al conjunto de puntos del espacio paramétrico donde la verosimilitud es estrictamente positiva lo representaremos por $\Omega_x = \{\vartheta \in \Omega\ ;\ f(x/\vartheta) > 0\}$ y a la clausura de este conjunto lo denominaremos <u>soporte de la función de verosimilitud</u>, a saber, $\overline{\Omega}_x$.

<u>Teorema 1.2.8.</u> Si f (x/ϑ) satisface las hipótesis 1.2.1., entonces B_x es un operador continuo de (Ω^*, w) en sí mismo.

<u>Demostración.</u> Si $\{P_\alpha\} \xrightarrow{w} P$ hay que probar que $B_x(P^\alpha) \xrightarrow{w} B_x(P)$ ó simplemente que $\{P_x^\alpha\} \xrightarrow{w} P_x$, donde $P_x^\alpha = B_x\ (P^\alpha)$ y $P_x = B_x\ (P)$. Esto es equivalente a probar que para toda $g \in CA(\Omega)$ se tiene que

$$\int_\Omega g\ d P_x^\alpha \longrightarrow \int_\Omega g\ d P_x$$

Por el teorema de Bayes se tiene que

$$d P_x^\alpha = \frac{f(x/\vartheta)}{f(x/P^\alpha)}\, dP^\alpha \qquad y \qquad d P_x = \frac{f(x/\vartheta)}{f(x/P)}\, dP$$

y por lo tanto todo se reduce a probar que

$$\int_\Omega \frac{g(\vartheta)\, f(x/\vartheta)}{f(x/P^\alpha)}\, dP^\alpha \longrightarrow \int_\Omega \frac{g(\vartheta)\, f(x/\vartheta)}{f(x/P)}\, dP$$

lo cual es consecuencia de las siguientes desigualdades:

$$\left| \int \frac{g(\theta) f(x/\theta)}{f(x/P^{\alpha})} dP^{\alpha} - \int \frac{g(\theta) f(x/\theta)}{f(x/P)} dP \right| =$$

$$= \left| \int \frac{g(\theta) f(x/\theta)}{f(x/P^{\alpha})} dP^{\alpha} - \int \frac{g(\theta) f(x/\theta)}{f(x/P^{\alpha})} dP + \int \frac{g(\theta) f(x/\theta)}{f(x/P^{\alpha})} dP - \int \frac{g(\theta) f(x/\theta)}{f(x/P)} dP \right| \leq$$

$$\leq \frac{1}{f(x/P^{\alpha})} \left| \int g(\theta) f(x/\theta) dP^{\alpha} - \int g(\theta) f(x/\theta) dP \right| +$$

$$+ \frac{1}{f(x/P)\, f(x/P^{\alpha})} \int | f(x/P^{\alpha}) - f(x/P) | \, | g(\theta) | \, f(x/\theta) \, dP$$

El primer miembro de la desigualdad tiende hacia cero por ser la función $g(\theta) f(x/\theta)$ continua y acotada por la hipótesis de que $\{P^{\alpha}\} \to P$, mientras que el segundo miembro también tiende hacia cero por ser $g(\theta) f(x/\theta)$ acotado y por ser $f(x/P^{\alpha}) \to f(x/P)$ ya que

$$f(x/P^{\alpha}) = \int_{\Omega} f(x/\theta) dP^{\alpha} \quad \text{y} \quad f(x/P) = \int_{\Omega} f(x/\theta) dP$$

y por ser $f(x/\theta)$ continua y $\{P^{\alpha}\} \to P$, resulta que

$$| f(x/P^{\alpha}) - f(x/P) | \longrightarrow 0 \quad .$$

<u>Corolario 1.2.3.</u> Bajo las hipótesis 1.2.1. el operador de Bayes conmuta con la operación de clausura débil para la clase de los conjuntos de Ω^* que sean relativamente compactos, es decir, si K^* es relativamente compacto, entonces $B_x(\overline{K^*}) = \overline{B_x(K^*)}$.

<u>Demostración.</u> Por ser B_x continua siempre se tiene que $B_x(\overline{K^*}) \subset \overline{B_x(K^*)}$. Por otro lado, de ser $K^* \subset \overline{K^*}$ se tiene que $B_x(K^*) \subset B_x(\overline{K^*})$ y como $\overline{K^*}$ es compacto y B_x continua $B_x(\overline{K^*})$ es compacto y por lo tanto cerrado; luego de la última inclusión se tiene que $\overline{B_x(K^*)} \subset B_x(\overline{K^*})$, y por lo tanto $B_x(\overline{K^*}) = \overline{B_x(K^*)}$.

<u>Corolario 1.2.4.</u> El operador de Bayes, bajo las hipótesis 1.2.1.,

transforma conjuntos relativamente compactos en conjuntos relativamente compactos.

<u>Demostración.</u> Si K^* es relativamente compacto, $B_x(K^*)$ también lo es, pues por el corolario anterior $\overline{B_x(K^*)} = B_x(\overline{K^*})$ y como $\overline{K^*}$ es compacto y B_x es continua $B_x(\overline{K^*})$ es compacto.

<u>Teorema 1.2.9.</u> Si $f(x/\vartheta)$ cumple las hipótesis 1.2.1. y si $\overline{\Omega}_x$ es el soporte de la verosimilitud, entonces

$$\text{sop}(P_x) = \text{sop}(P) \cap \overline{\Omega}_x$$

<u>Demostración.</u>

a) $\text{sop}(P_x) \subset \text{sop}(P) \cap \overline{\Omega}_x$

Sea $\vartheta_0 \in \text{sop}(P_x)$. Entonces $\vartheta_0 \in \text{sop}(P)$ pues caso contrario existiría un abierto de ϑ_0, $\mathcal{E}(\vartheta_0)$ tal que $\mathcal{E}(\vartheta_0) \cap \text{sop}(P) = \emptyset$ y esto implicaría que $P(\mathcal{E}(\vartheta_0)) = 0$ y como $P_x << P$ se tendría que $P_x(\mathcal{E}(\vartheta_0)) = 0$, que contradice el que $\vartheta_0 \in \text{sop}(P_x)$. Además $\vartheta_0 \in \overline{\Omega}_x$ pues caso contrario, existirá un abierto de ϑ_0, $U(\vartheta_0)$ tal que para todo $\vartheta \in U(\vartheta_0)$, $f(x/\vartheta) = 0$ y esto, a su vez implicaría, por la fórmula de Bayes (2.1) que $P_x(U(\vartheta_0)) = 0$.

b) Recíprocamente, $\text{sop}(P) \cap \overline{\Omega}_x \subset \text{sop}(P_x)$.

Sea $\vartheta_1 \in \text{sop}(P) \cap \overline{\Omega}_x$, y supongamos que $\vartheta_1 \notin \text{sop}(P_x)$. Entonces existiría un abierto $O(\vartheta_0)$ conteniendo a ϑ_0 tal que $P_x(O(\vartheta_0)) = 0$. Ahora por la fórmula (2.1.) se tendría que

$$P_x(O(\vartheta_1)) = 0 = \frac{1}{f(x/P)} \int_{O(\vartheta_1)} f(x/\vartheta)\, d P(\vartheta) = \frac{1}{f(x/P)} \int_{O(\vartheta_1)\cap\Omega_x} f(x/\vartheta)\, d P(\vartheta)$$

y como $O(\vartheta_1) \cap \Omega_x$ es un abierto, (por ser $\Omega_x = \{\vartheta;\ f(x/\vartheta) > 0\}$ y $f(x/\vartheta)$ continua), no vacío que contiene a ϑ_1 y por ser

$P(O(\vartheta_1) \cap \Omega_x) > 0$ se tendría que $f(x/\vartheta) = 0$ para todo -- $\vartheta \in O(\vartheta_1) \cap \Omega_x$ y esto implicaría que $\vartheta_1 \notin \overline{\Omega}_x$, contradicción que demuestra el teorema.

Corolario 1.2.5. Si $f(x/\vartheta)$ es continua, acotada y su soporte es todo el espacio paramétrico, es decir $\overline{\Omega}_x = \Omega$, entonces $sop(P_x) = sop(P)$.

Como consecuencia del teorema anterior también se tiene que el operador de Bayes transforma distribuciones con soporte finito en distribuciones con soporte finito. La demostración que ofrecemos no se basa en dicho teorema sino en una aplicación sencilla del lema 1.2.1.

Teorema 1.2.10. $B_x(F^*) \subset F^*$. Además si $\Omega_x = \Omega$, entonces B_x restringida a F^* es biyectiva.

Demostración. La primera parte del teorema no utiliza en absoluto la hipótesis de que $f(x/\vartheta)$ sea acotada ni continua, y es por tanto más general que la que se deducirá del teorema 1.2.9. Sea pues $P \in F^*$. Por definición de F^* se tiene que

$$P = \sum_{i=1}^{n} \alpha_i \, \delta_{\vartheta_i} \qquad \text{con} \qquad \vartheta_i \in \Omega .$$

Por el lema 1.2.1. sería

$$P_x = \sum_{i=1}^{n} \alpha_i(x) \, \delta_{\vartheta_i}$$

donde $\alpha_i(x) = \dfrac{\alpha_i f(x/\vartheta_i)}{f(x/P)}$ ya que $P_x(\delta_{\vartheta_i}) = \delta_{\vartheta_i}$ como se deduce obviamente de la fórmula (2.1) de donde $P_x \in F^*$ y además $sop(P_x) = sop(P) \cap \Omega_x$.

Si además $\Omega_x = \Omega$ entonces $f(x/\vartheta_i) > 0$ para todo $\vartheta_i \in \Omega$ con lo cual de las fórmulas

$$\alpha_i(x) = \frac{\alpha_i f(x/\vartheta_i)}{f(x/P)} \qquad i = 1, \ldots, n.$$

pueden despejarse las α_i en función de las $\alpha_i(x)$ y éstas, las α_i, serían únicas con lo cual B_x sería suprayectiva e inyectiva.

Como consecuencia de este teorema, se tiene el siguiente que pone de manifiesto el que la incertidumbre total no puede reducirse por el muestreo siempre que el soporte de la verosimilitud sea todo el espacio paramétrico, y además este sea -- compacto.

<u>Teorema 1.2.11.</u> Si Ω es compacto, entonces B_x es sobreyectiva, es decir, $B_x(\Omega^*) = \Omega^*$, siempre que la verosimilitud - sea continua, acotada y estrictamente positiva en Ω .

<u>Demostración 1.</u> Como Ω es compacto, es bien sabido que Ω^* es compacto en la topología débil. Por el teorema 1.2.3. se tiene que $\overline{F^*} = \Omega^*$ y por lo tanto F^* es relativamente compacto; si ahora aplicamos el corolario 1.2.3., se tiene que

$$B_x(\Omega^*) = B_x(\overline{F^*}) = \overline{B_x(F^*)} = \overline{F^*} = \Omega^*,$$

donde la penúltima igualdad es cierta por la segunda parte del - teorema 1.2.10. por ser B_x restringida a F^* suprayectiva.

La segunda demostración utiliza básicamente las mismas - ideas teniendo unicamente en cuenta que Ω , considerado como subconjunto de Ω^* es precisamente el conjunto de los puntos extremos de Ω^* .

<u>Demostración 2.</u> Por el teorema de Krein-Milman se tiene que -- $\Omega^* = \overline{co}(\Omega)$. Por otro lado, por ser $\Omega_x = \Omega$ se tiene del teotema de Bayes (2.1) que $B_x(\Omega) = \Omega$. De los corolarios 1.2.2. y 1.2.4. se tiene que

$$B_x(\Omega^*) = B_x(\overline{co}(\Omega)) = \overline{co}(B_x(\Omega)) = \overline{co}(\Omega) = \Omega^*.$$ "

Por último como consecuencia inmediata del corolario 1.2.5. se tiene el siguiente

<u>Corolario 1.2.6.</u> Si la función de verosimilitud cumple las hipótesis 1.2.1. y si $\Omega_x = \Omega$, entonces $B_x(\Omega^{*+}) \subset \Omega^{*+}$. Además si Ω es compacto, entonces $\overline{B_x(\Omega^{*+})} = \Omega^*$.

A continuación vamos a discutir brevemente la interpretación e importancia de los resultados anteriores referentes al operador de Bayes y al soporte de las distribuciones a priori y a posteriori en relación con la incorporación de un experimento de dimensión fija al modelo de decisión en incertidumbre parcial.

Sea, pues $(\Omega, D, R ; K^*)$ un problema de decisión con información parcial, donde Ω es un espacio métrico separable, D un conjunto arbitrario de decisiones y $R(\vartheta, d)$, la función de riesgo que supondremos continua en $\vartheta \in \Omega$, para todo $d \in D$ y K^* es un subconjunto no vacío de Ω^* .

Por $\mathcal{E} = (\mathcal{X}, \mathcal{B}; P_\vartheta)$ designamos un experimento de dimensión fija, donde $\mathcal{X}$ es el espacio muestral, que generalmente sería un subconjunto medible de $\mathbb{R}^n$, $\mathcal{B}$ la σ-álgebra de borelianos de $\mathcal{X}$ y $\{P_\vartheta ; \vartheta \in \Omega\}$ una familia de distribuciones de probabilidad que supondremos está dominada por una medida σ-finita $\nu(x)$, y las densidades respecto de esta medida las representaremos por

$$f(x/\vartheta) = \frac{dP_\vartheta(x)}{d\nu(x)}$$

Si seguimos un procedimiento análogo al análisis en forma extensiva de los problemas a priori, dada por K^*, se transformará, si el resultado del experimento es x, en la incertidumbre a posteriori $K^*_x = B_x(K^*)$ que denominaremos el <u>trasformado de K^* por el muestreo</u>. Entonces el modelo o problema original se transforma en el siguiente $(\Omega, D, R ; K^*_x)$ y la solución de este problema será, como en el problema original, el conjunto de los elementos maximales de D , respecto del preórden generado por $\succsim_{K^*}$, es decir,

$$d \succsim_{K^*} d' \text{ si y solo si } R(P_x, d) \leq R(P_x ; d')$$

para todo $P_x \in K^*$, donde $R(P,d)$ representa riesgo de Bayes de la decisión $d \in D$ respecto de la distribución a priori P , a saber

$$R(P,d) = \int_{\Omega} R(\Phi,d)\, P(d\Phi)$$

En el capítulo 1. vimos como las operaciones de envoltura convexa y de clausura debil no alteraban el preorden parcial $\succsim_{K^*}$ siempre que la función de riesgo $R(\Phi,d)$ fuese contimua en Φ y acotada para toda $d \in D$. El corolario 1.2.2. muestra que -sin necesidad de añadir hipótesis alguna sobre la función de verosimilitud- el operador Bayes conmuta con la operación de envoltura convexa, lo cual concuerda con nuestro "análisis en forma extensiva del modelo de decisión con incertidumbre parcial", ya que tanto K^*_x como $B_x(co(K^*_x))$ generan el mismo preorden parcial. Sin embargo el corolario 1.2.3. implica que la operación de clausura débil conmuta con el operador Bayes, en principio, cuando la información parcial K^* viene dada por un conjunto relativamente compacto y esto suponiendo que la función de verosimilitud cumpliese las hipótesis 1.2.1.. En general, solo se tendrá, por la continuidad del operador Bayes, que $B_x(\overline{K^*}) \subset \overline{B_x(K^*)}$ y que $B_x(\overline{K^*})$ sería adherente a $\overline{B_x(K^*)}$, y por lo tanto generarían el mismo preorden.

Esta observación muestra que el operador de Bayes no transforma, en general, cerrados en cerrados como sería de desear (compárese con el desarrollo axiomático de Girón y Rios (198) con el que la incertidumbre a posteriori dado un suceso A no nulo, viene tambien representada por un conjunto cerrado de distribuciones de probabilidad <u>pero finitamente aditivas</u>). Sin embargo si K^* es cerrado -sea o no compacto- entonces los conjuntos $B_x(K^*)$ y $\overline{B_x(K^*)}$ generan el mismo preorden a posteriori.

Por otra parte las distribuciones del conjunto $\overline{B_x(K^*)} \setminus B_x(K^*)$ <u>no</u> pueden obtenerse aplicando la fórmula de Bayes, de ninguna distribución de probabilidad a priori propia, aunque serian límites débiles de distribuciones a posteriori. Estas ideas podrían servir de fundamento teórico al tema, de importancia en estadística bayesiana, de las distribuciones a priori de referencia o no informativas impropias (vease p.e. los trabajos de Jaynes, en especial, Jaynes (1980)).

En relación con la llamada ley de Cromwell, ya expuesta en la interpretación del teorema 1.2.4., merece la pena señalar que si se cumple que sop(P) = Ω para todo $P \in K^*$, entonces siempre es aplicable la fórmula de Bayes (2.1) ya que la información a priori y la información muestral son compatibles, y ademas, por el teorema 1.2.9., siempre se tiene que sop(P) = $\overline{\Omega}_x$ para todo $P \in K^*$, y caso de que $\Omega_x = \Omega$ como ocurre, p.e., en el caso de muestras de la familia exponencial se tiene que el operador Bayes conserva el soporte de la distribución a priori lo cual es importante a la hora de saber si una decisión es admisible mediante la aplicación del teorema de Blyth (1951).

Por último, señalar que el tratamiento del modelo a posteriori (Ω, D ; R, K^*_x) es exactamente análogo al del modelo a priori (Ω, D ; R, K^*) y que no tratamos el problema de estudiar el comportamiento de K^*_x a medida que la información aumenta.

1.2.3. Modelos jerárquicos.

El considerar que la información a priori puede representarse en varias etapas mediante la introducción de los llamados hiperparámetros ha sido considerada en la literatura estadística en relación con los métodos bayesianos y con los llamados métodos empírico-bayes, en el tratamiento de problemas específicos. El objetivo de este parágrafo es simplemente el estudio teórico de estos modelos y su incidencia en los problemas de decisión con información parcial.

Consideremos pues, el caso de un problema de decisión con información parcial (Ω , D ; R , K^*) donde Ω es un espacio métrico separable y K^* es un subconjunto de Ω^* que, según es bien sabido tambien es un espacio métrico (con la métrica de Prohorov) separable (vease p.e. Parthasarathy (1967) Th. 6.2., p.43, Varadarajan (1961) o Billingsley (1968) Th.5., pp. 237-238). A lo largo de todo este parágrafo tambien supondremos que la función de riesgo $R(\vartheta ; d)$ es continua y acotada en ϑ para todo $d \in D$. Si $P \in K^*$ el riesgo asociado a esta distribución a priori y a una función de decisión $d \in D$ viene dado, como de costumbre, por

$$R(P, d) = \int_{\Omega} R(\vartheta, d)\, P(d\vartheta) \qquad (2.2)$$

La idea fundamental de aleatorizar en los problemas de decisión es simplemente convexificar el espacio correspondiente, en nuestro caso Ω , mediante la introducción del conjunto Ω^* que es convexo y cerrado. Hemos visto en el primer capítulo que las operaciones de envoltura convexa y de clausura respecto de la topología debil no alteraban el preorden generado por la información parcial K^* en el espacio de las decisiones. Así, sin pérdida de generalidad supondremos que K^* es un subconjunto convexo y cerrado de Ω^* .

Si ahora consideramos el problema de decisión en incertidumbre dado por $\Gamma_1 = (K^* , D , R_1)$, donde R_1 esta definida por (2.2) como un nuevo problema de decisión donde Ω ha sido sustituido por K^* , información parcial sobre los nuevos estados de la Naturaleza vendrá dada por un subconjunto de Ω^{**}, que tambien representaremos (véase al final del capítulo 3.) por $\mathcal{M}(\Omega^* , \mathcal{B}(\Omega^*))$, donde Ω^{**} es el subconjunto de todas las medidas de probabilidad definidas sobre el espacio medible $(\Omega^* , \mathcal{B}(\Omega^*))$, donde $\mathcal{B}(\Omega^*)$ es la σ-álgebra de Borel engendrada por los abiertos de Ω^* dotada de la métrica de Prohorov, ρ_1 , de Ω^* . Obsérvese que por los mismos teoremas anteriormente mencionados Ω^{**} es tambien un espacio métrico separable con métrica de Prohorov ρ_2 , es decir, en nuestro caso, las propiedades iniciales de Ω , es decir, el ser métrico y separable se va heredando en las sucesivas etapas.

Analogamente se definirá Ω^{***} y así sucesivamente. Por lo que a continuación veremos es suficiente considerar las dos primeras etapas de este proceso. De hecho si hemos restringido la información parcial en la primera etapa al subconjunto K^* de Ω^* , es suficiente en la segunda etapa considerar unicamente subconjuntos de $\mathcal{M}$ (K^* , $\mathcal{B}$ (K^*)), cuya significación es obvia. En este caso $\mathcal{M}$ (K^* , $\mathcal{B}$(K^*)) sería un espacio polaco si ademas imponemos a Ω la hipótesis de ser completo, propiedad que tambien es hereditaria en las sucesivas etapas del modelo jerárquico. Si representamos por $K^{**}\subset\mathcal{M}(K^*,\mathcal{B}(K^*))$ la información parcial de segundo orden, correspondiente a la segunda etapa del modelo jerárquico, tendremos el nuevo problema $\Gamma_2 = (K^{**}, D , R)$ donde R_2 viene dada por

$$R_2(\mu , d) = \int_{\Omega^*} R_1(P , d)\, \mu (d P) \qquad (2.3)$$

para toda $\mu \in K^{**}$.

Así como la función R_1 (P , d) es una extensión de R (θ , d) que es necesaria si Ω no es inicialmente convexo, la función $R_2(\mu , d)$, es una extensión de R_1 (P , d), pero que de hecho no es necesaria desde el punto de vista teórico pues la convexificación de Ω ya se ha conseguido en la primera etapa. Esto nos conduce directamente al concepto de baricentro de una distribución

Definición 1.2.6. Se denomina baricentro de la hiperdistribución $\mu\in\Omega^{**}$ a la medida de probabilidad $Q_\mu \in \Omega^*$ definida por

$$Q_\mu(B) = \int_{\Omega^*} P(B)\, \mu(d P) \qquad (2.4)$$

para todo $B \in \mathcal{B}(\Omega)$, y lo representaremos $Q_\mu = \text{bar}\,(\mu)$.

Es sabido que $P(B)$ es una función medible de P para todo $B \in \mathcal{B}(\Omega)$ (vease p.e. Jagers (1974) y final del capítulo tercero de esta memoria), con lo cual la definición (2.4) tiene sentido.

Teorema 1.2.12. (de marginalización). Si K^* es un subconjunto convexo y cerrado de Ω^* y si $\mu \in \mathcal{M}(K^*, \mathcal{B}(K^*))$ entonces $Q = \text{bar}\,(\mu) \in K^*$ y $R_2(\mu, d) = R_1(Q, d)$ para toda $d \in D$.

Demostración. Demostremos en primer lugar la segunda parte del teorema que es independiente de la primera. En efecto, se tiene que

$$R_1(Q,d) = \int_\Omega R(\vartheta,d)\, Q(d\vartheta) = \int_\Omega R(\vartheta,d) \left\{ \int_{\Omega^*} P(d\vartheta)\,\mu\,(dP) \right\} =$$

$$= \int_{\Omega^*} \left\{ \int_\Omega R(\vartheta,d)\, P(d\vartheta) \right\} \mu\,(dP) = \int_{\Omega^*} R_1(P,d)\,\mu\,(dP) = R_2(\mu,d),$$

donde el cambio de orden de integración se sigue del teorema de Fubini si $R(\vartheta, d)$ es medible $\mathcal{B} \times \mathcal{D}$.

Para demostrar la primera parte supongamos que $Q \notin K^*$, entonces los conjuntos $\{Q\}$ y K^* serían convexos, cerrados y ademas $\{Q\}$ sería compacto de donde se deduce que existe una función continua y acotada $f_0 \in CA(\Omega)$ y una constante c_0 tales que

$$\int_\Omega f_0\, dQ < c_0$$

y
$$\int_\Omega f_0\, dP > c_0$$

para todo $P \in K^*$. Ahora bien $\int_\Omega f_0(\vartheta)\, Q(d\vartheta) = \int_\Omega f_0(\vartheta) \int_{K^*} P(d\vartheta)\,\mu(dP) =$

$$= \int_{K^*} \left\{ \int_\Omega f_0(\vartheta)\, P(d\vartheta) \right\} \mu\,(dP) > \int_{K^*} c_0\,\mu\,(dP) = c_0$$

contradicción que demuestra el teorema. Observese que el cambio de orden en la integral doble es lícito en virtud del teorema de Fubini.

Este teorema pone de manifiesto que desde el punto de vista del problema de decisión considerado al principio no es necesario en los modelos jerárquicos más allá de la primera etapa pues si p.e., la información parcial asociada a un problema de decisión viene dada por la sucesión $K_1^{*}, K_2^{**}, \ldots, K_n^{**\cdots*}$ donde $K_i^{**\cdots*} \in \mathcal{M}(K_{i-1}^{*\cdots*}, \mathcal{B}(K_{i-1}^{*\cdots*}))$ basta aplicar el principio de marginalización del teorema anterior e ir sustituyendo cada conjunto $K_i^{*\cdots*}$ por el de sus baricentros comenzando por $K_n^{*\cdots*}$.

Desde el punto de vista práctico el problema anterior se ve reducido en el caso de que los subconjuntos de información parcial del modelo jerárquico $\{K_1^{*}, K_2^{**}, \ldots, K_n^{*\cdots*}\}$ fuesen compactos, aplicando la teoría de los baricentros de Choquet segun la cual en la etapa i-ésima unicamente habría que considerar el conjunto $\mathcal{M}(K_{ie}^{*\cdots*}, \mathcal{B}(K_{ie}^{*\cdots*}))$ donde $K_{ie}^{*\cdots*}$ es el conjunto de los puntos extremos de $K_i^{*\cdots*}$.

En relación con los modelos jerárquicos y el concepto de soporte de una hiperdistribución, tiene interes el teorema siguiente, que es una generalización del teorema 1.2.2.

<u>Teorema 1.2.13</u>. Si Q es el baricentro de μ, entonces se verifica que

$$\operatorname{sop}(Q) = \overline{\bigcup_{P \in \operatorname{sop}(\mu)} \operatorname{sop}(P)}$$

<u>Demostración</u>. Si $P_0 \in \operatorname{sop}(\mu)$ entonces $\operatorname{sop}(P) \subset \operatorname{sop}(Q)$. En efecto, sea $Q_0 \in \operatorname{sop}(P_0)$, Esto implica que para todo abierto $\mathcal{E}(\vartheta_0)$ es $P(\mathcal{E}(\vartheta_0)) > 0$. Entonces se tiene que $Q(\mathcal{E}(\vartheta_0)) = \int_{\Omega^*} P(\mathcal{E}(\vartheta_0))\mu(dP) \geq \int_{O(P_0)} P(\mathcal{E}(\vartheta_0))\mu(dP) > 0,$

donde $O(P_0)$ es cualquier abierto que contiene a P_0 y por hipótesis $\mu(O(P_0)) > 0$, de donde se sigue que $Q(\mathcal{E}(\theta_0)) > 0$ y por lo tanto $\theta_0 \in$ sop (Q).

De aquí se deduce que $\bigcup_{P \in \text{sop}(\mu)}$ sop $(P) \subset$ sop (Q) y como sop (Q) es cerrado se sigue que

$$\overline{\bigcup_{P \in \text{sop}(\mu)} \text{sop}\,(P)} \subset \text{sop}\,(Q)$$

Reciprocamente, si $\theta \in$ sop (Q) y se tuviese que $\theta \notin \overline{\bigcup_{P \in \text{sop}(\mu)} \text{sop}\,(P)}$, existiría un entorno $\mathcal{E}_0(\theta)$ tal que

$\mathcal{E}_0(\theta) \cap \overline{\bigcup_{P \in \text{sop}(\mu)} \text{sop}\,(P)} = \emptyset$, y esto implicaría el que

$P(\mathcal{E}_0(\theta)) = 0$ para toda $P \in$ sop (μ), y de aquí que

$$Q(\mathcal{E}_0(\theta)) = \int_{\Omega^*} P(\mathcal{E}_0(\theta))\,\mu(dP) = \int_{\text{sop}(\mu)} P(\mathcal{E}_0(\theta))\,\mu(dP) = 0$$

contradiccion que demuestra el teorema.

Consideremos ahora, como la incorporación de un experimento $\mathcal{E} = (\mathcal{X}, \mathcal{B}, P_\theta ; \theta \in \Omega)$ con $\{P_\theta ; \theta \in \Omega\}$ dominada por la medida σ-finita ν, siendo $f(x/\theta)$ las densidades de P_θ respecto de ν, afecta a los modelos jerárquicos. Por lo que acabamos de ver, vamos a considerar unicamente dos etapas, siendo extensiones de nuestros resultados a más etapas análogas a las dadas. Como hemos supuesto en todo el capítulo, supondremos fijo el resultado experimental $x \in \mathcal{X}$.

El modelo inicial $(\Omega ; D, R, K^*)$ se transforma en el modelo a posteriori (Ω, D, R, K^*_x) donde $K^*_x = B_x(K^*)$ y el operador B_x queda caracterizado por la verosimilitud $f(x/\theta)$ a través de la fórmula (2.1).

Veamos que ocurre con el modelo jerárquico de segundo orden $\Gamma_2 = (K^{**}, D, R_2)$ tras la observación del resultado experimental $x \in \mathcal{X}$.

La primera cuestión es extender la función de verosimilitud de Ω a Ω^* de un modo coherente. Aquí proponemos la siguiente definición, que es una generalización por mixturas de la considerada en inferencia cuando se consideran modelos contaminados.

Definición 1.2.7. Si $\mathcal{E} = (\mathcal{X}, \mathcal{B}; P_\theta, \theta \in \Omega)$ es un experimento y $f(x/\theta)$ es la función de verosimilitud, se define la función de verosimilitud generalizada y la representaremos por $f(x/P)$ a la función definida en Ω^* por

$$f(x/P) = \int_\Omega f(x/\theta)\, P(d\theta) \qquad (2.4)$$

Observese que si $f(x/\theta)$ satisface las hipótesis 1.2.1., entonces $f(x/P)$ es tambien una función acotada y continua de P respecto de la topología débil. Ademas si $P = \delta_\theta$ entonces $f(x/\delta_\theta) = f(x/\theta)$.

Si ahora consideramos la segunda etapa del modelo jerárquico $\Gamma_2 = (K^{**}, D, R_2)$, este se transformará en $\Gamma_{2x} = (K^{**}_x, D, R_2)$ donde K^{**}_x se obtendrá de la fórmula de Bayes aplicada ahora tomando como espacio paramétrico Ω^* y como nueva función de verosimilitud $f(x/P)$.

Según esto si la hiperdistribución $\mu \in K^{**}$, la hiperdistribución a posteriori μ^*_x vendrá dada por la fórmula

$$\frac{d\mu_x(P)}{d\mu(P)} = \frac{f(x/P)}{\int_{\Omega^*} f(x/P)\,\mu(dP)} \qquad (2.5)$$

Observese que el denominador de la fórmula (2.5) se podría escribir, extendiendo de nuevo la función de verosimilitud de Ω^* a Ω^{**} por un procedimiento identico, como

$$f(x/\mu) = \int_{\Omega^*} f(x/P)\mu(dP)$$

y así sucesivamente en las diversas etapas del modelo.

La pregunta que surge de forma natural es la siguiente: Por el teorema de marginalización 1.2.12. podemos sustituir el modelo de la segunda etapa dado por K^{**} por el modelo de la primera etapa representado por $K^* = \text{bar}\ (K^{**})$. Analogamente podríamos hacer con K_x^{**}, sustituyendolo por bar (K_x^{**}) ¿ Es $K_x^* = \text{bar}\ (K_x^{**})$?

Otra pregunta más concreta en el mismo sentido sería la siguiente: ¿ Si el baricentro de la hiperdistribución a priori μ es Q , es el baricentro de la hiperdistribución a posteriori μ_x , dada por (2.5), $B_x(Q)$?

La respuesta a ambas preguntas es negativa. No obstante existe una relación sencilla entre $B_x(Q)$ y μ_x dada por el teorema siguiente que junto con (2.5) puede considerarse como una generalización del teorema de Bayes. Observese tambien que si la respuesta a la segunda pregunta fuese afirmativa, tambien lo sería a la primera. Hacemos notar por último, que este teorema es una generalización del lema 1.2.1..

<u>Teorema 1.2.14</u>. Si Q es el baricentro de la hiperdistribución a priori μ , es decir $Q = \text{bar}\ (\mu)$, y μ_x es la hiperdistribución a posteriori de μ , dada por (2.5), entonces

$$B_x(Q) = \int_{\Omega^*} B_x(P)\mu_x(dP) \quad .$$

<u>Demostración</u>.Sea $Q_x = B_x(Q)$. Por el teorema de Bayes (2.1) tenemos que para todo $B \in \mathcal{B}$

$$Q_x(B) = \frac{\int_B f(x/\vartheta)Q(d\vartheta)}{f(x/Q)} = \frac{\int_B f(x/\vartheta)\left(\int_{\Omega^*} P(d\vartheta)\mu(dP)\right)}{f(x/Q)} =$$

$$= \frac{\int_{\Omega^*}\left(\int_B f(x/\vartheta)\, P(d\vartheta)\mu(dP)\right)}{f(x/Q)} = \frac{\int_{\Omega^*} P_x(B)\, f(x/P)\, \mu(dP)}{f(x/Q)}$$

como $f(x/\mu) = \int_{\Omega^*} f(x/P)\mu(dP) = \int_{\Omega^*}\left(\int_{\Omega} f(x/\vartheta)\, P(d\vartheta)\right)\mu(dP) =$

$$= \int_{\Omega} f(x/\vartheta)\left(\int_{\Omega^*} P(d\vartheta)\mu(dP)\right) = \int_{\Omega} f(x/\vartheta)\; Q(d\vartheta) = f(x/Q)$$

sustituyendo arriba y teniendo en cuenta (2.5), resulta finalmente que

$$Q_x(B) = \int_{\Omega^*} P_x(B)\mu_x(dP) = \int_{\Omega^*} B_x(P)(B)\mu_x(dP).$$

Para obtener de este teorema el lema 1.2.1., basta tomar μ tal que sop $(\mu) = \{P^1, \ldots\ldots, P^n\}$ con $\mu(\{P^i\}) = \alpha_i$ y $P = \text{bar}(\mu)$, ya que la fórmula (2.5) para este caso se reduciría a

$$\alpha_i(x) = \frac{\alpha_i\, f(x/P^i)}{f(x/P)} \quad .$$

Otra generalización importante, desde el punto de vista de las aplicaciones, es el caso en que

$$K^* = \{ P(\cdot\ , \gamma);\ \gamma \in \Gamma \}$$

donde γ es un parámetro que generalmente varía en un subconjunto de Borel de un espacio euclideo de dimensión finita. En este caso una hiperdistribución a priori quedará caracterizada por una distribución sobre Γ, $H(\Gamma)$, en cuyo caso la fórmula (2.5) adoptará la expresión

$$\frac{dH_x(\gamma)}{dH(\gamma)} = \frac{f(x/\gamma)}{\int_\Gamma f(x/\gamma)\, dH(\gamma)}$$

donde $H_x(\gamma)$ es la hiperdistribución a posteriori del parámetro γ y

$$f(x/\gamma) = \int_{\Omega} f(x/\vartheta)\, P(d\vartheta;\gamma)$$

La distribución a posteriori de la mixtura

$$Q(d\vartheta) = \int_{\Gamma} P(d\vartheta;\gamma)\, dH(\gamma)$$

sería

$$Q_x(d\vartheta) = \int_{\Gamma} P_x(d\vartheta;\gamma)\, dH_x(\gamma)$$

donde las $P_x(d\vartheta;\gamma)$ se calculan aplicando la fórmula dé Bayes (2.1).

SEGUNDA PARTE.-

S-JUEGOS GENERALIZADOS CON INFORMACION PARCIAL.

Introducción y objetivos de esta segunda parte.

En esta parte tratamos los problemas relativos a la existencia de soluciones cuasi-bayesianas y su relación con los conceptos de admisibilidad y completitud a que dan lugar los preórdenes parciales asociados al concepto de información parcial tal como se ha considerado en la primera parte.

La información parcial ha sido caracterizada axiomáticamente por Girón y Ríos (1981). Sin embargo para los propósitos de esta parte es mejor considerar la información parcial, bien en términos de subconjuntos de medidas de probabilidad sobre el espacio paramétrico o bien en términos de preórdenes parciales en el espacio de las funciones de decisión que, a su vez, quedan caracterizados por conos convexos y cerrados en el espacio de las funciones de riesgo asociado al problema de decisión con incertidumbre parcial.

Ya se señaló en la primera parte que si el espacio paramétrico es demasiado general las dos formas de abordar los problemas de decisión con información parcial que acabamos de exponer, no son equivalentes, es decir, pueden existir en el espacio de las funciones de riesgo preórdenes parciales dados por conos convexos y cerrados que no quedan caracterizados por subconjuntos de medidas de probabilidad σ-aditivas, y por lo tanto no corresponderían a problemas de decisión con información parcial.

Nuestra teoría se podría extender al caso de que las probabilidades a priori consideradas fuesen finitamente aditivas siguiendo la línea de Heath and Sudderth (1972, 1978) y Sudderth (1980). Aunque recientemente esta línea está tomando un cierto desarrollo, hemos preferido, al igual que en la primera parte, desarrollar la teoría basada en probabilidades σ-aditivas. Y es por esta razón por la que en esta parte nos restringimos a los S-juegos. En Girón (1975, 1977) se dan razones por las que los S-juegos generalizados son el vehículo ideal para tratar una gran clase de problemas de decisión. Además las hipótesis sobre los S-juegos no pueden ser relajadas (en particular la compacidad del espacio paramétrico o la continuidad de la función de riesgo) si queremos que los teoremas del capítulo 3. sean válidos.

La explicación profunda del por qué estos resultados no pueden ser mejorados estriba en el hecho de que los espacios $\mathcal{C}(\Omega)$ y $\mathcal{C}^*(\Omega)$, dotado éste último de la topología débil $*$ están en dualidad, y de que además, $\mathcal{C}^*(\Omega)$ es isométricamente isomorfo al espacio de las medidas finitas, regulares y signadas sobre $(\Omega, \mathcal{B})$.

Una contribución importante de esta parte es el estudio de la <u>polaridad estricta</u> basada en un teorema de Klee (1955) y que permite investigar la estructura de los problemas de decisión con incertidumbre parcial, en relación con el concepto de soporte de una distribución. Para establecer estos resultados son también fundamentales ciertos resultados de la teoría de los baricentros de Choquet (1960), (véase también Choquet y Meyer, 1963), que como vimos en la primera parte sirve de soporte teórico para la construcción de los modelos jerárquicos.

Como se ha visto en la primera parte, la incorporación de la información mediante la regla de Bayes nos trasforma un problema de decisión con incertidumbre parcial en otro análogo, donde ahora la incertidumbre parcial queda caracterizada por el conjunto de distribuciones a posteriori que presenta las mismas características que el conjunto de distribuciones a priori, de modo que los resultados de esta parte son independientes de si el problema de decisión con incertidumbre parcial es con o sin experimentación.

CAPITULO 3.

DEFINICIONES Y CONCEPTOS FUNDAMENTALES.

Sea $(\Omega, D ; R)$ un problema de decisión en el que Ω es un espacio métrico compacto y $R(\theta, d)$ una función contínua de θ para todo $d \in D$. Si consideramos la aplicación:

$$\rho : D \longrightarrow \mathcal{C}(\Omega)$$

definida por $\rho(d) = R(\theta, d)$, la imagen de D por ρ es un subconjunto S de $\mathcal{C}(\Omega)$.

Al par (Ω, S) se le denomina representación canónica del problema $(\Omega, D; R)$ o simplemente el S-juego asociado.

Según esto un problema de decisión con incertidumbre parcial, será, a partir de ahora, una terna $(\Omega, S; K^*)$, donde K^* es un subconjunto de Ω^*.

Ω^* es como en la primera parte, el conjunto de todas las medidas de probabilidad σ-aditivas y regulares definidas sobre el espacio $(\Omega, \mathcal{B})$.

K^* engendra en $C(\Omega)$ -y por tanto en S- una relación de preórden parcial dada por

$$f \underset{K^*}{\gtrsim} g \text{ sí y solo sí } \int f \, dP \leq \int g \, dP \text{ para todo } P \in K^*.$$

Es fácil comprobar que la estructura de preórden sobre $C(\Omega)$ así definida es compatible con la estructura de espacio vectorial, es decir, satisface los dos axiomas siguientes (véase p.e. Bourbaki, (1966), p. 52).

EO1. $f \underset{K^*}{\gtrsim} g$ implica $f + h \underset{K^*}{\gtrsim} g + h$ cualquiera que sea $h \in C(\Omega)$.

E02. Si $f \underset{K^*}{\gtrsim} 0$ entonces $\lambda f \underset{K}{\gtrsim} 0$ para todo $\lambda \geq 0$

Con lo cual $C(\Omega)$ se convierte en un espacio vectorial preordenado y se tiene (véase Bourbaki, (1966), prop. 13).

Teorema 2.3.1. El conjunto de elementos K tales que son preferidos o indiferentes a 0, es decir,

$$K = \{ f \in C(\Omega);\ f \underset{K^*}{\gtrsim} 0 \} \qquad (3.1)$$

forman un cono convexo con vértice en el origen.

Recíprocamente si K es un cono convexo con vértice en el origen la relación

$$f \underset{K}{\gtrsim} g \text{ si y solo si } f - g \in K$$

es una relación de preórden en $C(\Omega)$ y es la única que es compatible con la estructura de espacio vectorial para la cual K sea el conjunto de los elementos preferidos o indiferentes a 0.

Así pues a cada K^* le corresponde un único cono convexo K tal que

$$\text{si } f \underset{K^*}{\gtrsim} g, \text{ entonces } f - g \in K$$

Una cuestión interesante es saber cuando el preórden generado por K^* es orden. Esto en términos del K asociado a K^* es fácil de saber. Más adelante, cuando demostremos que K a su vez caracteriza a K^* (de modo único si K^* se supone convexo y compacto en la topología débil, como se demuestra en el teorema 2.3.8.), daremos una caracterización de los órdenes en términos de K^*.

Definición 2.3.1. Si K es un cono convexo y con vértice en el origen, de denomina espacio de linealidad de K, y lo representaremos por L (K), a la máxima variedad lineal contenida en K.

Se demuestra (véase Bourbaki, (1966), pág. 50), que -- $L(K) = K \cap (-K)$, y que $L(K)$ representa todos los puntos de - $C(\Omega)$ indiferentes a 0 respecto del preórden dado por (3.1).

<u>Definición 2.3.2.</u> Un cono convexo K de $C(\Omega)$ se dice que es - puntiagudo si $K \cap (-K) = \{0\}$.

<u>Teorema 2.3.2.</u> Una condición necesaria y suficiente para que la relación $\underset{K^*}{\gtrsim}$ sea un orden parcial es que K definido por (3.1.) sea puntiagudo.

<u>Demostración.</u> Trivial.

Siguiendo nuestra investigación de las propiedades del cono K vamos a ver que la topología de $C(\Omega)$ es compatible con la estructura de espacio vectorial ordenado.

<u>Teorema 2.3.3.</u> K es cerrado respecto de la topología usual de - $C(\Omega)$ dada por la norma

$$\|f\| = \sup_{\vartheta \in \Omega} |f(\vartheta)|$$

<u>Demostración.</u> Si la sucesión $f_n \in K$, n = 1, 2, ... es tal que - $\{f_n\} \longrightarrow f$ se tiene que $\int f_n dP \leq 0$ para todo n y para todo - $P \in K^*$. Como f_n está acotada, por el teorema de la dominancia - acotada, se deduce que:

$$\int f \, d \, P = \int \lim_n f_n \, d \, P = \lim_n \int f_n \, d \, P \leq 0$$

y por lo tanto $f \in K$.

Otra demostración más interesante y que arroja más luz sobre la relación entre K y K^* se basa en el hecho de que

$$K = \bigcap_{P \in K^*} H_P^-$$

donde $H_P^- = \{ f \in C(\Omega);\ \int f dP \leq 0 \}$ son semiespacios cerrados, con lo cual K es cerrado y además por ser intersección de semiespacios que pasan por el origen es un cono convexo.

La última propiedad interesante de los conos K definidos por (3.1) refleja el que $K^* \subset \Omega^*$.

<u>Definición 2.3.3.</u> Se denomina el <u>ortante fundamental</u> de $C(\Omega)$, y lo representaremos por K_0 al conjunto definido por

$$K_0 = \{ f \in C(\Omega);\ f(\theta) \leq 0 \text{ para todo } \theta \in \Omega \} .$$

<u>Teorema 2.3.4.</u> Si $f \in K_0$ y $P \in \Omega^*$ siempre se tiene que $\int f dP \leq 0$ y por lo tanto $f \underset{K^*}{\approx} 0$ cualquiera que sea $K^* \subset \Omega^*$.

Más adelante veremos que este teorema es una consecuencia trivial de la dualidad entre K y K^*.

Señalamos por último que la restricción de que $K^* \subset \Omega^*$ implica que el preórden $\underset{K^*}{\gtrsim}$ definido en $C(\Omega)$ es no trivial, que correspondería al caso de que todos los elementos de $C(\Omega)$ fuesen indiferentes, o, lo que es equivalente $K = C(\Omega)$. Este resultado se puede expresar diciendo que el preórden $\underset{K^*}{\gtrsim}$ es compatible con el axioma de dominancia fuerte.

<u>Teorema 2.3.5.</u> Si $f(\theta) > g(\theta)$ para todo $\theta \in \Omega$, y K está definido por (3.1), entonces $f \underset{K^*}{>} g$.

<u>Demostración.</u>Trivial.

Si denominamos $\mathcal{C}$ la clase de todos los conos con vértice en el origen, convexos, cerrados, que contienen a K_0 y no son todo el espacio $C(\Omega)$ hemos visto que podemos establecer una aplicación de $\mathcal{P}(\Omega^*)$ en $\mathcal{C}$ definida por

$$\varphi : K^* \longrightarrow K$$

donde $K = \{ f \in C(\Omega);\ \int f dP \leq 0 \text{ para todo } P \in K^* \}$
y tal que $f \succsim_{K^*} g$ si y solo si $f-g \in K$.

Nuestro propósito ahora es estudiar las propiedades de esta aplicación. La idea básica es que la aplicación no es inyectiva como vamos a ver a continuación, pero -y este es uno de los resultados fundamentales del capítulo- si a K^* le exigimos el que sea convexo y cerrado en la topología débil la aplicación anteriormente definida es biunívoca con lo cual tendremos caracterizada la clase de los preórdenes parciales asociada a los problemas de decisión en términos de la calse $\mathcal{P}$ de conos de $C(\Omega)$, más manejable a la hora de tratar los problemas de admisibilidad y completitud.

Antes de proseguir con este tema merece la pena señalar que la clase $\mathcal{P}$ queda unívocamente caracterizada en términos de axiomas de racionalidad (véase el Teorema 3.1. de Girón y Ríos, 1981).

La clase de los <u>órdenes</u> queda además caracterizada en términos de los conos puntiagudos de $\mathcal{C}$ como afirma el teorema 2.3.2. y los preórdenes completos, que corresponderían al caso de que $K^* = \{P\}$, quedan caracterizados por los semiespacios de $\mathcal{P}$.

<u>Teorema 2.3.6.</u> Una condición suficiente para que $\succsim_{K^*}$ sea un preórden completo es que $K^* = \{P\}$ ó, equivalentemente, que $K \in \mathcal{P}$ sea un semiespacio.

<u>Demostración.</u> Trivial. (Más adelante en el teorema 2.3.10. se establece que esta condición es también necesaria, basándonos en resultados de Girón, (1977)).

A continuación y como preámbulo al teorema 2.3.8. vamos a ver como las operaciones de convexificación y de clausura (respecto de la topología débil) de K^* dejan invariante el preórden.

<u>Teorema 2.3.7.</u> Si $K^* \subset \Omega^*$, co (K^*) y $\overline{K^*}$ generan en C (Ω) el mismo preórden, es decir,

$$\varphi(K^*) = \varphi(\text{co } (K^*)) = \varphi(\overline{K^*}) = K$$

donde $K = \left\{ f \in C\ (\Omega);\ \int f dP \leq 0 \text{ para todo } P \in K^* \right\}$.

<u>Demostración.</u> Como $K^* \subset$ co (K^*), claramente si $f \underset{co(K^*)}{\gtrsim} g$ entonces $f \underset{K^*}{\gtrsim} g$. Recíprocamente, si $f \underset{K^*}{\gtrsim} g$ se tiene que $\int f dP \leq \int g dP$ para todo $P \in K^*$. Sea $Q \in$ co (K^*), es decir, $Q = \sum_{i=1}^{n} \alpha_i P_i$ con $\alpha_i \geq 0$, $\sum_{i=1}^{n} \alpha_i = 1$, $P_i \in K^*$. De aquí se tiene que

$$\sum_{i=1}^{n} \alpha_i \int f dP_i \leq \sum_{i=1}^{n} \alpha_i \int g dP_i, \text{ o sea,}$$

$$\int f d\left(\sum_{i=1}^{n} \alpha_i P_i\right) \leq \int g d\left(\sum_{i=1}^{n} \alpha_i P_i\right) \text{ y de aquí que } \int f dQ \leq \int g dQ$$

cualquiera que sea $Q \in$ co (K^*).

Como $K^* \subset \overline{K^*}$ también se tiene que si $f \underset{\overline{K^*}}{\gtrsim} g$ entonces $f \underset{K^*}{\gtrsim} g$. Recíprocamente, sea $f \underset{K^*}{\gtrsim} g$ y sea $Q \in \overline{K^*}$. Entonces existe una red $\{P_\alpha\} \xrightarrow{w} Q$ con $P_\alpha \in K^*$. Para los P_α se tiene que $\int f dP_\alpha \leq \int g dP_\alpha$ y por la convergencia débil que $\int f dQ \leq \int g dQ$, es decir, $f \underset{\overline{K^*}}{\gtrsim} g$.

Como consecuencia inmediata de este teorema se tiene el siguiente

<u>Corolario 2.3.1.</u> K^* y $\overline{\text{co}}$ (K^*) generan el mismo preórden en C (Ω), es decir,

$$\varphi(K^*) = \varphi(\overline{\text{co}}\ (K^*))$$

Como caso particular de este corolario se tiene que, si se toma $K^* = \Omega$, entonces por el teorema 1.2.3., $\overline{\text{co}}$ $(\Omega) = \Omega^*$, con lo cual se tendría que

$$\varphi(\Omega) = \varphi(\Omega^*) = K_0$$

como es fácil de comprobar. Dicho en palabras, el orden natural de C (Ω), dado por K_0, viene engendrado por Ω .

La idea básica del teorema 2.3.8. es establecer una correspondencia biunívoca entre los subconjuntos de Ω^* y los conos de la clase $\mathcal{P}$. A la vista del corolario 2.3.1. parece que nos deberíamos restringir a la clase de los subconjuntos de Ω^* convexos y cerrados. Esta idea es fecunda y permite encontrar la equivalencia buscada. Dicho de forma imprecisa: que la información parcial dada por K^* y K'^* es la misma siempre que $\overline{co}(K^*) = \overline{co}(K'^*)$ y que dos informaciones parciales representadas por conjuntos convexos y cerrados, son idénticas si y sólo si, estos conjuntos son iguales.

A la clase de todos los subconjuntos convexos y cerrados de Ω^* la representaremos por $\mathcal{P}^*$.

La demostración del teorema 2.3.8. se podría simplicar caso de que el espacio C (Ω) fuese reflexivo, pero, como es bien sabido, a no ser que C (Ω) sea finito-dimensional C (Ω) no es reflexivo. Esta dificultad se puede superar considerando en $C^*(\Omega)$, precisamente la topología débil * y mediante la siguiente definición que enlaza con la (3.1).

Aunque la siguiente definición se da para cualquier subconjunto de $C^*(\Omega)$ nuestro interés se va a centrar únicamente en los subconjuntos de Ω^* (que se puede considerar como un subconjunto de la esfera unidad de $C^*(\Omega)$).

<u>Definición 2.3.4.</u> Se llama <u>polar restringido</u> de un subconjunto K^* de $C^*(\Omega)$ al conjunto

$$\mathrm{pr}\,(K^*) = K = \{ f \in C(\Omega);\ x^*(f) \leq 0 \text{ para todo } x^* \in K^* \} \quad (3.2\ )$$

Obsérvese que si $K^* \subset \Omega^*$, entonces esta definición, junto con el teorema de representación de Riesz, coincide con la (3.1).

Es trivial comprobar que K es un cono convexo y cerrado de C (Ω), y, por los teoremas 2.3.1. a 2.3.5., si $K^* \subset \Omega^*$ entonces el cono K definido por (3.2.) pertenece a $\mathcal{C}$.

<u>Definición 2.3.5.</u> Se denomina <u>polar</u> de un subconjunto A de C (Ω) al conjunto

$$p\ (A) = A^* = \{ x^* \in C^*(\Omega);\ x^*(f) \leq 0 \quad \text{para todo} \quad f \in A \} \quad (3.3)$$

El conjunto A^* es un cono convexo y cerrado, (respecto de la topología débil * de $C^*(\Omega)$).

De hecho para nuestras aplicaciones, vamos a considerar la intersección de A^* con la esfera unidad de $C^*(\Omega)$, esto es el conjunto de los $x^* \in C^*(\Omega)$ tales que $\| x^* \| = 1$, y por lo tanto a partir de ahora A^* será el subconjunto de la esfera unidad definido por (3.3).

De las definiciones dadas se deduce inmediatamente el siguiente

<u>Lema 2.3.1.</u> Si $A \subset B$ entonces $p\ (A) \supset p\ (B)$. Si $K^* \subset K'^*$ entonces $pr\ (K^*) \supset pr\ (K'^*)$. Además

$$pr\ [p\ (A)] \supset A \qquad y \qquad p\ [pr\ (K^*)] \supset K^*]$$

<u>Demostración.</u> Trivial.

<u>Lema 2.3.2.</u> Sean K_1^* y K_2^* dos subconjuntos convexos, compactos y disjuntos de Ω^* . Entonces existe una función continua $f_o \in C(\Omega)$ tal que

$$\int f_o\ dP > 0 \qquad \text{para todo} \quad P \in K_1^*$$

$$\int f_o\ dQ < 0 \qquad \text{para todo} \quad Q \in K_2^*$$

<u>Demostración.</u> Sean $C(K_1^*)$ y $C(K_2^*)$ los conos con vértice en el origen engendrados por K_1^* y K_2^* respectivamente. Estos conos son cerrados y localmente compactos (véase p.e. Bourbaki, (1966), prop. 3, pág. 113) de $C^*(\Omega)$ con la topología débil $*$. Entonces aplicando primero el teorema (2.5) de Klee (1955) y posteriormente el teorema (3.14) de Rudin (1979) se tiene el resultado deseado, ya que Ω^* no contiene al origen.

<u>Teorema 2.3.8.</u> Si K^* es un subconjunto convexo y cerrado de la esfera unidad de $C^*(\Omega)$, entonces $P\left[pr\ (K^*)\right] = K^*$. Análogamente, si A es un cono convexo y cerrado de $C(\Omega)$, entonces $pr\left[p(A)\right] = A$.

<u>Demostración.</u> A la vista del lema basta demostrar que en las condiciones del teorema se verifican

a) $p\left[pr\ (K^*)\right] \subset K^*$ y b) $pr\left[p\ (A)\right] \subset A$

Las dos demostraciones son análogas pero las damos completas.

a) Sea $x_0^* \in p\left[pr\ (K^*)\right]$ y supongamos que $x_0^* \notin K^*$. Entonces los subconjuntos $\{x_0^*\}$ y K^* son convexos y cerrados en la topología débil $*$ y además $\{x_0^*\}$ es compacto. Existe por lo tanto un funcional lineal continuo que los separa, es decir existe una función continua $g_0 \in C(\Omega)$ y una constante C, que se puede tomar $C = 0$ por el lema 2.3.2., tales que

$$x_0^*(g_0) > 0 > x^*(g_0) \quad \text{para todo} \quad x^* \in K^*$$

De aquí se deduce que $g_0 \in pr\ (K^*)$, y, como por otro lado $x_0^* \in p\left[pr\ (K^*)\right]$ se tendría que $x_0^*(g_0) \leq 0$, contradicción que prueba la parte a).

b) Sea $f_0 \in pr\left[p\ (A)\right]$ y supongamos que $f_0 \notin A$. Entonces los conos $\{\lambda f_0;\ \lambda \geq 0\}$ y A son conos convexos cerrados tales que $\{\lambda f_o\} \cap A = \{0\}$ y $\{\lambda f_o\}$ es localmente compacto. Por un teorema de Klee (1955) existe un funcional lineal continuo γ_o^* tal que $\gamma_o^*(f) \leq 0$ para todo $f \in A$ e $\gamma_o^*(f_o) > 0$. De la primera desigualdad se tiene que $\gamma_o^* \in p(A)$ y como $f_o \in pr[p(A)]$ se tendría entonces que $\gamma_o^*(f_o) \leq 0$, contradicción que demuestra la parte b).

<u>Corolario 2.3.2.</u> La aplicación $\varphi = \mathcal{K}^* \longrightarrow \mathcal{K}$ es biunívoca y es tal que $\varphi(K^*) = pr(K^*)$ y $\varphi^{-1}(K) = p(K)$.

Tenemos así caracterizada la clase de los problemas de decisión con incertidumbre parcial en términos de la clase $\mathcal{K}$. La subclase de los preórdenes completos queda caracterizada en $\mathcal{K}^*$ por todos los subconjuntos formados por un solo elemento y en términos de $\mathcal{K}$ por los semiespacios. Por último, la clase de los órdenes queda caracterizada en $\mathcal{K}$ por los conos puntiagudos, mientras que su caracterización en términos de la clase $\mathcal{K}^*$ requiere la introducción de nuevos conceptos.

<u>Definición 2.3.6.</u> Un subconjunto C^* de Ω^* se dice que es <u>total o completo</u> si de ser $\int f dP = 0$ para todo $P \in C^*$ se deduce que $f \equiv 0$ siempre que $f \in C(\Omega)$.

El nombre de subconjunto total está tomado del análisis funcional, mientras que la idea de completitud de C^* es análoga a la de estadístico completo para una cierta familia de distribuciones.

<u>Teorema 2.3.9.</u> Una condición necesaria y suficiente para que el preórden engendrado por K^* sea orden es que K^* sea completo.

<u>Demostración.</u> Por el teorema 2.3.2. será necesario y suficiente que $pr(K^*)$ sea puntiagudo.

Sea K^* completo. Si $f \in K \cap -K$ se tiene que $\int f dP \leq 0$ y $\int (-f)\, dP \leq 0$ para todo $P \in K^*$. De aquí se deduce que $\int f dP = 0$

para todo $P \in K^*$, que implica que $f \equiv 0$ y por lo tanto $K \cap -K =$ $= \{0\}$ y K es puntiagudo.

Recíprocamente si K es puntiagudo y K^* fuese tal que -- pr $(K^*) = K$, entonces K^* es completo. Si no lo fuese, existiría un $f_0 \neq 0$ tal que $\int f_0 \, dP = 0$ para todo $P \in K^*$ y esto implicaría que $f_0 \in L(K) = K \cap -K = \{0\}$. Contradicción que demuestra la parte suficiente.

<u>Corolario 2.3.3.</u> Ω considerado como subconjunto de Ω^* es un -- subconjunto completo.

<u>Demostración.</u> Ω considerado como subconjunto de Ω^* tiene como polar restringido pr $(\Omega) = K_0$ y este es puntiagudo.

Como ejemplo de subconjunto completo de medidas de probabilidad cuando $\Omega = [0, 1]$ tenemos la clase de todas las medidas de probabilidad que se distribuyen según $\mathcal{B}e\ (\alpha, \beta)$ con $\alpha > 0$, $\beta > 0$. Se puede demostrar facilmente que si

$$K^* = \{ P \in \Omega^*; \ P \in \mathcal{B}e\ (\alpha, \beta) \text{ con } \alpha > 0, \ \beta > 0 \}$$

entonces $\overline{co}\ (K^*) = \Omega^*$.

Por último los preórdenes totales o completos, es decir, los preórdenes bayesianos de la clase $\mathcal{P}^*$ quedan caracterizados por el siguiente teorema debido a Giron (1977).

<u>Teorema 2.3.10.</u> Una condición necesaria y suficiente para que - el preórden engendrado por K^* sea completo es que K^* esté formado por una única distribución de probabilidad.

Siguiendo con nuestra investigación de la estructura de los preórdenes de la clase $\mathcal{P}^*$ que será básica para establecer la relación entre decisiones no dominadas y decisiones Bayes, - vamos a introducir un nuevo concepto que es el de polar estricto.

Daremos la definición en general pues puede ser útil a la hora de estudiar estructuras de dominancia en ciertos espacios funcionales sin que correspondan a problemas de decisión con información parcial. Nuestra definición es análoga a la definición 2.5. de Criado, Girón y Ríos (1981), pero adaptada a la nota que sigue a la definición 2.3.5.

<u>Definición 2.3.7.</u> Sea A un cono convexo y cerrado con vértice en el origen. Se define el <u>polar estricto</u> de A, y lo representaremos por A^{*+} al conjunto

$$A^{*+} = \{ x^* \in C^*(\Omega);\ \|x^*\| = 1;\quad x^*(f) < 0 \text{ para todo } f \in A \setminus L(A) \} \tag{3.4}$$

Claramente se tiene que A^{*+}, si existe, es un subconjunto convexo de $A^* = p\ (A)$.

La primera cuestión importante es la de la existencia del polar estricto, la cual no siempre se da. El teorema siguiente demuestra que si $K \in \mathcal{K}$ entonces K^{*+} es no vacío, aunque como puede verse en Criado, Girón y Ríos (1981) es suficiente que A sea puntiagudo sin necesidad de ser $A \in \mathcal{K}^*$, para que A^{*+} sea no vacío.

Para los $K \in \mathcal{K}$ la definición anterior adopta la forma.-

$$K^{*+} = \{ P \in \Omega^* \ ;\ \int f dP < 0 \quad \text{para todo } f \in K \setminus L(K) \} \tag{3.4a}$$

<u>Teorema 2.3.11.</u> Si $K \in \mathcal{K}$, entonces $K^{*+} \neq \emptyset$.

<u>Demostración.</u> En primer lugar de la definición de $\mathcal{K}$ se deduce que para todo $K \in \mathcal{K}$, $K \setminus L\ (K) \neq \emptyset$, pues $K \neq C\ (\Omega)$ y además $K \supset K_0$ que contiene a un abierto de $C\ (\Omega)$. Sea pues $f_0 \in - K \setminus L(K)$. Entonces los conos $\{ \lambda f_0;\ \lambda \geqslant 0 \}$ y K son convexos, cerrados y tales que $\{ \lambda f_0;\ \lambda \geqslant 0\} \cap K = \{0\}$. Además $\{ \lambda f_0;\ \lambda \geqslant 0 \}$ es localmente compacto. Entonces por el teorema (2.7) de Klee (1955) existe un funcional lineal continuo $x^* \in C^*(\Omega)$ tal que

$x^*(f) < 0$ para todo $f \in K \setminus L(K)$ y $x^*(f) = 0$ para todo $f \in L(K)$, de modo que si tomamos $\|x^*\| = 1$ entonces $x^* \in K^* \subset \Omega^*$ y por el teorema de representación de Riesz existe una única $P \in \Omega^*$, tal que $x^*(f) = \int f dP$, con lo cual $P \in K^{*+}$.

Vamos ahora a estudiar la relación entre los conjuntos K^* y K^{*+}. De momento sabemos que $K^{*+} \subset K$. Si consideramos el caso de que $K_Q^* = \{Q\}$, es decir los preórdenes bayesianos, sabemos que $K_Q = pr(K^*) = \{f \in C(\Omega);\ \int f dQ \leq 0\}$, con lo cual $L(K_Q) = K_Q \cap (-K_Q) = \{f \in C(\Omega);\ \int f dQ = 0\}$ y por lo tanto $K_Q^{*+} = \{Q\}$ pues por un lado $Q \in K^{*+}$ ya que $\int f dQ < 0$ para todo $f \in K \setminus L(K)$ y por el otro lado $K_Q^{*+} \subset K_Q^* = \{Q\}$, con lo cual $K_Q^{*+} = K_Q^* = \{Q\}$. Como veremos más adelante solamente para los preórdenes bayesianos se tiene que $K^{*+} = K^*$.

Consideramos ahora el otro caso extremo, es decir, el caso de incertidumbre total cuando $K^* = \Omega^*$ ó bien $K = K_o$. La idea básica es la de relacionar el concepto de polar estricto con la idea de soporte de una distribución. Este resultado es un caso particular de un teorema más general que veremos más adelante (teorema 2.3.16b), pero por su importancia lo damos ahora, ya que su demostración no requiere, a diferencia del teorema general, la utilización de la teoría de los baricentros de Choquet.

<u>Lema 2.3.3.</u> Para toda $f \in C(\Omega)$ y toda $P \in \Omega^*$ siempre se tiene que

$$\int_{\Omega} f dP = \int_{sop(\Omega)} f dP$$

<u>Demostración</u>. Como $\int_{\Omega} f dP = \int_{sop(P)} f dp + \int_{sop(P)^c} f dP$. Basta probar que $\int_{sop(P)^c} f dP = 0$, que es consecuencia directa de las desigualdades siguientes,

$$\left| \int_{sop(P)^c} f dP \right| \leq \int_{sop(P)^c} \|f\|\, dP = \|f\|\ P(sop(P)^c) = 0$$

<u>Teorema 2.3.12.</u> El polar estricto del conjunto K_o es precisamente el conjunto $\Omega^{*+} = \{P \in \Omega^*,\ sop(P) = \Omega\}$

Demostración.

a) $\Omega^{*+} \subset K_o^{+}$

Por ser K_o puntiagudo si sop $(P) = \Omega$ es suficiente probar que $\int f dP < 0$ para toda $f \in K_o$, $f \neq 0$. Por ser $f \in K_o \setminus \{0\}$ y por ser contínua existe un $\theta_o \in \Omega$ y un abierto $U(\theta_o)$ tal que $f(\theta) < 0$ para $\theta \in U(\theta_o)$. Entonces

$$\int_\Omega f dP = \int_{U(\theta_o)} f dP + \int_{\Omega \setminus U(\theta_o)} f dP \leq \int_{U(\theta_o)} f dP + 0 < 0.$$

b) $K_o^{*+} \subset \Omega^{*+}$

Supongamos que $Q \in K_o^{*+}$ y que sop $(Q) \neq \Omega$. Entonces existiría al menos un $\theta_o \in \Omega$ tal que $\theta_o \notin$ sop (Q). Por ser un espacio normal, existe un entorno $U(\theta_o)$ disjunto de sop(Q) y tal que $\overline{U(\theta_o)} \cap$ sop $(Q) \neq \emptyset$. Por el teorema de Urysohn existe una función continua $f_o \in C(\Omega)$ tal que

$$f_o(\theta) = \begin{cases} -1 & \text{si } \theta \in \overline{U(\theta_o)} \\ 0 & \text{si } \theta \in \text{sop } (Q) \end{cases}$$

Por lo tanto $f_o \in K_o \setminus \{0\}$ y

$$\int_\Omega f_o \, d\,Q = \int_{\text{sop}(Q)} f_o \, d\,Q = 0$$

es decir, $Q \notin K_o^{*+}$; contradicción que demuestra el teorema.

Tenemos aquí caracterizada la relación entre el polar y el polar estricto de los preórdenes bayesianos para los que ambos conjuntos coinciden y el orden natural en el que el polar - estricto está formado por aquellas distribuciones de probabilidad cuyo soporte es todo el espacio paramétrico y su polar es - el conjunto de todas las distribuciones de probabilidad. Por el teorema 1.2.4. sabemos que $\overline{\Omega^{*+}} = \Omega^{*}$ y para el caso bayesiano $\overline{K_p^{*+}} = \{P\} = K_p^{*}$. Esta relación es válida para cualquier --

problema de decisión con incertidumbre parcial como demuestra el teorema siguiente,

<u>Teorema 2.3.13.</u> $\overline{K^{*+}} = K^{*}$ si $K^{*} \in \mathcal{F}$

<u>Demostración.</u> Como K^{*} es compacto y $K^{*+} \subset K^{*}$, trivialmente es $\overline{K^{*+}} \subset K^{*}$.

Para demostrar la inclusión contraria sea $P_0 \in K^{*}$ y supongamos que $P_0 \notin \overline{K^{*+}}$. Los conjuntos $\{P_0\}$ y $\overline{K^{*+}}$ son convexos, compactos y disjuntos y por lo tanto, en virtud del Lema 2.3.2., existe una función continua $f_0 \in C(\Omega)$ tal que $\int f_0 dP_0 > 0 > \int f_0 dQ$ para todo $Q \in \overline{K^{*+}}$.

Esto implica que $f_0 \notin K = pr(K^{*})$. Si aplicamos el teorema de Klee a los conos $\{\lambda f_0;\ \lambda \geq 0\}$ y K, existe un $Q_0 \in K^{*+}$ tal que $\int f_0 dQ_0 > 0$. Pero esto contradice el que $\int f_0 dQ < 0$ para todo $Q \in \overline{K^{*+}}$.

Siguiendo con la investigación de la estructura de los conjuntos K^{*} y K^{*+} tiene interés el siguiente teorema que nos da información acerca de la estructura de los órdenes parciales.

<u>Teorema 2.3.14.</u> Si $K^{*} \in \mathcal{F}$ es completo o total, entonces $K^{*+} \subset \Omega^{*+}$. Es decir, si K es puntiagudo, su polar estricto está formado exclusivamente por distribuciones cuyo soporte es todo el espacio paramétrico.

<u>Demostración.</u> Sea $P_0 \in K^{*+}$ y supongamos que $P_0 \notin \Omega^{*+}$. Si denominamos $\Omega_0 = sop(P_0)$, existirá un $\vartheta_0 \in \Omega \setminus \Omega_0$. Por ser $P_0 \in K^{*+}$ y K^{*} completo, $K = pr(K^{*})$ es puntiagudo por los teoremas 2.3.2. y 2.3.9. y por lo tanto

$$\int f\, d P_0 < 0 \quad \text{para todo} \quad f \in K \setminus \{0\} .$$

Vamos ahora a construir una función $f_0 \in K \setminus \{0\}$ para la cual $\int f_0\, d P_0 = 0$, lo cual contradiría el que $sop(P_0) \neq \Omega$.

En efecto, por el teorema de Urysohn, existe una función contínua f_o definida por

$$f_o(\vartheta) = \begin{cases} 0 & \text{si } \vartheta \in \Omega_o \\ -1 & \text{si } \vartheta = \vartheta_o \end{cases}$$

$f_o \in K_o \subset K$ y $f_o \neq 0$. Además por el lema 2.3.3.

$$\int_\Omega f_o \, d\, P_o = \int_{\Omega_o} f_o \, d\, P_o = 0$$

Obsérvese que el teorema no es cierto si K no es puntiagudo pues en este caso f_o podría pertenecer a L (K). El recíproco del teorema, no es, en general, cierto como se ve facilmente con un contraejemplo: Basta tomar $K^* = \{P\}$ con $sop(P) = \Omega$. Nótese, también, que del teorema se deduce que si el soporte de alguna distribución del polar estricto de K^* no es todo el espacio paramétrico, K^* no puede ser completo.

Hasta ahora tenemos caracterizados en términos del polar estricto los preórdenes bayesianos, aquellos para los cuales -- $K^{*+} = K^* = \{P\}$ con $P \in \Omega^*$, y la ignorancia total por $\Omega^{*+} = \{P \in \Omega^* ; sop(P) = \Omega\}$.

La última caracterización de K^{*+} va a depender del conjunto de los puntos extremos de K^*, por un lado, y por otro, - en la extensión de la idea de soporte de una medida de probabilidad sugerida por el resultado del teorema 2.3.12. La idea básica consiste, al igual que en los modelos jerárquicos tratados en la primera parte, en considerar el conjunto de todas las distribuciones de probabilidad sobre el conjunto K_e^* de los puntos extremos de K^* y mediante la aplicación del teorema de Choquet (1960) identificarlas con su baricentro y demostrar después que aquellas distribuciones cuyo soporte sea K_e^* pueden identificarse precisamente con un subconjunto de K^{*+}. Desgraciadamente, a no ser que el conjunto K^* sea un simplex (véase p.e., Choquet et Meyer, (1963), o Jacobs, (1978)), no se puede asegurar la unicidad de la representación integral, aunque esto no afecta a la veracidad del teorema 2.3.15.

Sea $K^* \in \mathcal{K}^*$. Representemos por K_e^* el conjunto de los puntos extremos de K^*. Es bien sabido que K_e^* es un G_δ y que $K^* = \overline{co}\ (K_e^*)$. También se tiene que por ser Ω métrico compacto Ω^* también lo es, (véase Varadarajan, (1963); Billinsley, (1968); Parthasaraty, (1967) ó Dalal and Gaineford, (1980)), y por consiguiente K_e^* es un espacio polaco. Si designamos por $\mathcal{M}\ (K_e^*, \mathcal{B}(K_e^*))$ al conjunto de todas las medidas de probabilidad sobre K_e^*, donde $\mathcal{B}\ (K_e^*)$ representa la σ-álgebra de Borel engendrada por los abiertos de la topología débil de Ω^* restringidos a K_e^*, éste será a su vez un espacio polaco dotado de su correspondiente métrica de Prohorov.

Por $\mathcal{M}^+\ (K_e^*, \mathcal{B}(K_e^*))$ representaremos el conjunto de todas las medidas de probabilidad cuyo soporte sea K_e^*, es decir,

$$\mathcal{M}^+((K_e^*), \mathcal{B}\ (K_e^*)) = \{\mu \in \mathcal{M}(K_e^*, \mathcal{B}\ (K_e^*));\ \text{sop}\ (\mu) = K_e^*\}$$

Vamos a establecer ahora una correspondencia entre $\mathcal{M}(K_e^*, \mathcal{B}(K_e^*))$ y K^* de modo que a cada $\mu \in \mathcal{M}\ (K_e^*, \mathcal{B}(K_e^*))$ le hacemos corresponder su <u>baricentro</u> definido por

$$Q\ (A) = \int_{K_e^*} P\ (A)\ \mu\ (dP) \quad \text{para todo } A \in \mathcal{B}(\Omega) \qquad (3.5\ )$$

donde $P \in K_e^*$ es una medida de probabilidad aleatoria según la nomenclatura de Jagers (1974). Además (véase p.e., Dalal and Gaineford, (1980), proposición 2), para cualquier $A \in \mathcal{B}\ (\Omega)$, $P\ (A)$ es una función medible.

A la aplicación así definida la representaremos por

$$\beta: \mathcal{M}\ (K_e^*, \mathcal{B}\ (K_e^*)) \longrightarrow K^*$$

$$\beta\ (\mu) = \int_{K_e^*} P\ \mu\ (dP).$$

El teorema de Choquet (1960) nos garantiza que esta aplicación es sobreyectiva. Si además K^* es un simplex, entonces (teorema 11. de Choquet y Meyer, (1963)), β es biyectiva.

La aplicación β (baricentro) induce otra aplicación $\overline{\beta}$, entre los espacios $C(\Omega)$ y $C(K^*_e)$, definida del modo siguiente:

$$\overline{\beta}(f) = \int_\Omega f(\vartheta)\, P(d\vartheta) = g(P) \qquad (3.6)$$

de modo que $$\int_\Omega f(\vartheta)\, Q(d\vartheta) = \int_{K^*_e} g(P)\, \mu(dP) \qquad (3.7)$$

donde $g(P) = \int_\Omega f(\vartheta)\, P(d\vartheta)$ y μ es tal que su baricentro es Q, como se ve facilmente aplicando el teorema de Fubini y las definiciones (3.5) y (3.6).

Las propiedades de las aplicaciones β y $\overline{\beta}$ las resumimos en el lema siguiente:

Lema 2.3.4. Las aplicaciones β y $\overline{\beta}$ son lineales y contínuas respecto de las topologías usuales de los espacios entre los que están definidas. Además $\overline{\beta}(K) \subset K_0$ donde K_0 es el ortante fundamental de $C(K^*_e)$ y $\overline{\beta}(L(K)) = 0$.

Demostración. La linealidad es trivial por definición, así como la continuidad de $\overline{\beta}$. Para establecer la continuidad de β sea $\{\mu_\alpha\}$ una red que converge a μ. Si $Q_\alpha = \beta(\mu_\alpha)$ y $Q = \beta(\mu)$, tenemos que demostrar que $Q_\alpha \xrightarrow{w} Q$. Sea $f \in C(\Omega)$. Basta pues demostrar que $\int_\Omega f\, dQ_\alpha \longrightarrow \int_\Omega f\, dQ$.

Por (3.7) se tiene que:

$$\int_\Omega f\, dQ_\alpha = \int_{K^*_e} g(P)\, \mu_\alpha(dP) \quad \text{y} \quad \int_\Omega f\, dQ = \int_{K^*_e} g(P)\, \mu(dP)$$

y como $g(P)$ es una función continua de P, (respecto de la topología débil inducida por K^*_e), y por otro lado, por hipótesis $\{\mu_\alpha\} \longrightarrow \mu$, queda demostrado que $\int f\, dQ_\alpha \longrightarrow \int f\, dQ$. "

Sea $f \in K$. Esto equivale a que $\int f\, dP \leq 0$ para todo $P \in K^*$, y en particular $\int_\Omega f(\vartheta)\, P(d\vartheta) = g(P) \leq 0$ para todo $P \in K^*_e$, es decir, $\overline{\beta}(f) \in K_0$. Del mismo modo si $f \in L(K)$, $\overline{\beta}(f) = 0$.

<u>Teorema 2.3.15.</u> $\beta(\mathcal{M}^{+}(K_e^*, \mathcal{B}(K_e^*))) \subset K^{*+}$ y además $\overline{\beta(\mathcal{M}^{+}(K_e^*, \mathcal{B}(K_e^*)))} = K^*$.

<u>Demostración.</u> Sea $\mu \in \mathcal{M}^{+}(K_e^*, \mathcal{B}(K_e^*))$ y sea $Q = \beta(\mu)$. Hay que demostrar que $\int_\Omega f(\vartheta)\, Q(d\vartheta) < 0$ para toda $f \in K \setminus L(K)$. Ahora bien por el Lema anterior $\overline{\beta}(f) = g(P) \in K_0 \setminus \{0\}$ y por (3.7).

$$\int_\Omega f(\vartheta)\, Q(d\vartheta) = \int_{K_e^*} g(P)\, \mu(dP).$$

Como $g(P) \neq 0$ existirá un entorno O tal que $g(P) < 0$ para todo $P \in O$. Como por hipótesis sop $(\mu) = K_e^*$, se tendrá

$$\int_\Omega f\, dQ = \int_{K_e^*} g(P)\, \mu(dP) = \int_O g(P)\, \mu(dP) + \int_{K_e^* - O} g(P)\mu(dP) < 0$$

Para demostrar la segunda parte del teorema tenemos, por un lado que por ser β contínua, es

$$\beta(\overline{\mathcal{M}^{+}(K_e^*, \mathcal{B}(K_e^*))}) \subset \overline{\beta(\mathcal{M}^{+}(K_e^*, \mathcal{B}(K_e^*)))}$$

Por otra parte sabemos, por el teorema 1.2.4., que por ser K_e^* separable $\overline{\mathcal{M}^{+}(K_e^*, \mathcal{B}(K_e^*))} = \mathcal{M}(K_e^*, \mathcal{B}(K_e^*))$. Por último, el teorema de Choquet nos dice que $\beta(\mathcal{M}(K_e^*, \mathcal{B}(K_e^*))) = K^*$, y, por lo tanto

$$K^* \subset \overline{\beta(\mathcal{M}^{+}(K_e^*, \mathcal{B}(K_e^*)))}$$

Ahora, como siempre es $\beta(\mathcal{M}^{+}(K_e^*, \mathcal{B}(K_e^*))) \subset K^*$, se tiene finalmente por ser K^* cerrado

$$K^* = \overline{\beta(\mathcal{M}^{+}(K_e^*, \mathcal{B}(K_e^*)))}$$

Obsérvese que como corolario de este teorema se deduce de nuevo el teorema 2.3.13., pues por la primera parte se tiene que

$$\beta(\mathcal{M}^{+}(K_e^*, \mathcal{B}(K_e^*))) \subset K^{*+} \subset K \quad \text{y tomando clausuras}$$

$$\overline{\beta(\mathcal{M}^{+}(K_e^*, \mathcal{B}(K_e^*)))} \subset \overline{K^{*+}} \subset K^*.$$

Y por la segunda parte se tiene que $\overline{K^{*+}} = K^*$

Hay que reseñar que el teorema <u>no afirma</u> que -- - $\beta(\mathcal{M}^+(K_e^*, \mathcal{B}(K_e^*))) = K^{*+}$ aunque esta igualdad parece - ser cierta en muchos casos. El saber si la igualdad anterior es cierta en general, es un problema abierto, del cual solamente - hemos encontrado soluciones particulares. Una de ellas se da a continuación en el teorema siguiente.

<u>Teorema 2.3.16.</u> Si $\overline{\beta}(K) = K_o$, entonces $\beta(\mathcal{M}^+(K_e^*, \mathcal{B}(K_e^*))) = K^{*+}$.

<u>Demostración.</u> Por el teorema anterior, basta demostrar que -- $K^{*+} \subset \beta(\mathcal{M}^+(K_e^*, \mathcal{B}(K_e^*)))$.

Sea $Q \in K^{*+}$. Por el teorema de Choquet existe al menos una $\mu \in \mathcal{M}(K_e^*, \mathcal{B}(K_e^*))$ tal que $Q = \int_{K_e^*} P\mu(d\,P)$. Si - ahora demostramos que sop $(\mu) = K_e^*$ habremos concluído. Supongamos que fuese Sop $(\mu) \neq K_e^*$. Existiría un entorno 0 tal que $\overline{0} \cap$ sop $(\mu) = \emptyset$. Por el teorema de Urysohn existirá una función contínua $g_\mu(P)$, de modo que:

$$g_\mu(P) = \begin{cases} 0 & \text{si } P \in \text{sop } (\mu) \\ -1 & \text{si } P \in 0 \end{cases}$$

y, por consiguiente, $g_\mu \in K_o \setminus \{0\}$. Además

$$\int_{K_e^*} g\,\mu(P)\,\mu(d\,P) = 0$$

<u>Por</u> ser $\overline{\beta}(K) = K_o$ existirá una $f_\mu \in K \setminus L(K)$ tal que $\overline{\beta}(f_\mu) = g_\mu(P)$. Por (3.7) se tendría que

$$\int_\Omega f_\mu\, d\,Q = \int_{K_e^*} g\,\mu(P)\,\mu\,(d\,P) = 0,$$

que contradiría el que $Q \in K^{*+}$.

Este teorema, aunque no cubre todos los casos de interés resulta a veces fácil de aplicar en los casos siguientes:

a) Si card $(K_e^*) < \aleph_0$ y card $(K_e^*) \leq$ card (Ω) entonces $\overline{\beta}(K) = K_0$. Es decir, si el conjunto de los puntos extremos es finito y su número es menor o igual que la dimensión del espacio $C(\Omega)$, el teorema 2.3.16. siempre es cierto. Obsérvese que el caso a) contiene como caso particular el caso bayesiano, en que $K^* = \{P\} = K^{*+} = K_e^*$.

b) Si $K^* = \Omega^*$, caso de ignorancia total, el teorema es aplicable, pues, en este caso, es fácil comprobar que Ω_e^* puede identificarse con Ω (identificando los estados de la Naturaleza con las distribuciones de probabilidad degeneradas). Así pues la correspondència $\overline{\beta}$ entre $C(\Omega)$ y $C(\Omega_e^*)$ se reduce a la identidad y como $K = K_0$ resulta que $\overline{\beta}(K_0) = K_0$. Obsérvese que también la aplicación β se reduce a la identidad.

CAPITULO 4.

RESULTADOS FUNDAMENTALES.

En este capítulo estudiamos las cuestiones referentes a la existencia de soluciones admisibles y cuasi-bayesianas de los problemas de decisión con información parcial, la relación entre ambas y ciertas cuestiones relacionadas con estos problemas.

Sea (Ω, S , K) un problema de decisión con información parcial, donde, sin pérdida de generalidad podemos suponer que $K^* \subset \mathcal{E}^*$. En el capítulo 3. se ha estudiado ampliamente la caracterización de los K^* en términos de la clase $\mathcal{E}$ y también en términos de los puntos extremos de K^*, y la relación entre ambos.

Como veremos a continuación, según la índole del problema a tratar será más conveniente trabajar a veces con el polar restringido de K^*, es decir, K.

Respecto del conjunto de riesgo S, sabemos que si el decisor considera estrategias mixtas, lo podemos suponer convexo. Sin embargo, muchas veces es interesante el no considerar estrategias aleatorizadas, en particular, cuando se consideran problemas de estimación estadística, pero interesa p.e., saber si cierta clase de decisiones no aleatorizadas es completa, etc.

Así pues a lo largo de este capítulo no supondremos, a no ser que se diga explícitamente, que el conjunto de riesgo sea convexo. Esta condición será sustituída por la de cono-convexidad debida a Yu (1973), y que suele ser más fácil de comprobar, pues como es sabido existen muchos problemas de decisión para los cuales el conjunto de riesgo no es convexo pero sí es cono-convexo, con lo cual esta teoría permite, en algunos problemas, prescindir de la aleatorización.

Si K^* es el conjunto de incertidumbre de un problema de decisión (Ω, S), representaremos por K a su polar restringido. Ambos originan en $C(\Omega)$ un preórden parcial dado por la relación.

$$f \underset{K}{\succsim} g \quad \text{si y solo si} \quad f - g \in K \qquad (4.1)$$

A partir de ahora emplearemos la notación $K(f) = f + K$, $I(f) = f + L(K)$ y $M(f) = K(f) \setminus I(f)$, donde $K(f)$, $I(f)$ y $M(f)$ representan respectivamente los puntos mejores que f, indiferentes a f, y estrictamente mejores que f, respecto del preórden definido por (4.1), es decir

$$K(f) = \{ g \in C(\Omega);\ g \underset{K}{\succsim} f \} = \{ g \in C(\Omega);\ g - f \in K \}$$

$$I(f) = \{ g \in C(\Omega);\ g \underset{K}{\sim} f \} = \{ g \in C(\Omega);\ g - f \in L(K) \}$$

$$M(f) = \{ g \in C(\Omega);\ g \underset{K}{\succ} f \} = \{ g \in C(\Omega);\ g - f \in K \setminus L(K) \}$$

<u>Definición 2.4.1.</u> $f \in S$ es <u>K - admisible</u> para el problema $(\Omega, S; K^*)$ si no existe ningún $g \in S$ tal que $g \underset{K}{\succ} f$. De modo equivalente, si $M(f) \cap S = \emptyset$.

<u>Definición 2.4.2.</u> Un subconjunto $C \subset S$ es una <u>clase completa</u> para el problema $(\Omega, S; K^*)$ si para todo $g \in S - C$ existe un $f \in C$ tal que $f \underset{K}{\succ} g$.

<u>Definición 2.4.3.</u> Un subconjunto $C \subset S$ es una <u>clase esencialmente completa</u> para el problema $(\Omega; S; K^*)$ si para todo $g \in S - C$ existe un $f \in C$ tal que $f \underset{K}{\succsim} g$.

Estas definiciones son análogas a las usuales en teoría de la decisión (véase p.e., Ferguson, (1967), pp. 55-56), con la diferencia de que el orden no es el orden natural sino el dado por (4.1) y no hacen referencia al problema de decisión sino a su forma canónica o S-juego asociado.

Los conceptos de clase completa y esencialmente completa aquí dados coinciden caso de que el preórden parcial $\underset{K}{\gtrsim}$ sea un orden y por lo tanto tenemos el siguiente resultado

Lema 2.4.1. Si K es puntiagudo toda clase es completa, si y solo si, es esencialmente completa.

Otros resultados referentes a estas definiciones sencillos de demostrar son los siguientes

Lema 2.4.2. Si $C(S;K)$ es una clase completa para el problema $(\Omega, S; K)$ y $\mathcal{A}(S;K)$ representa el conjunto de las estrategias admisibles, entonces $\mathcal{A}(S;K) \subset C(S;K)$.

Lema 2.4.3. Si $C(S;K)$ es una clase esencialmente completa para el problema $(\Omega, S; K)$ y $f \in \mathcal{A}(S;K)$ es tal que $f \notin C(S;K)$, entonces existe $g \in C(S;K)$ tal que $g \underset{K}{\sim} f$.

Lema 2.4.4. Si $C(S;K)$ es completa para el problema $(\Omega, S; K')$ entonces $C(S;K)$ es esencialmente completa.

El recíproco de este lema no es cierto, en general, aunque dada una clase esencialmente completa, se puede generar una clase completa incluyendo simplemente todos los elementos de S que sean indiferentes a los de la clase esencialmente completa y, recíprocamente, de toda clase completa se puede extraer una clase esencialmente completa recurriendo al axioma o principio de elección, con lo cual tenemos el siguiente

Lema 2.4.5. Si $C(S;K)$ es una clase completa para el problema $(\Omega, S; K)$, existe una subclase $C'(S;K) \subset C(S;K)$ que es esencialmente completa.

Para el problema $(\Omega, S; K)$ siempre existen clases completas y por el lema 2.4.4. esencialmente completas; basta tomar como $C(S;K) = S$. Pero de lo que se trata es de encontrar una clase completa que sea lo más pequeña posible, lo cual sugiere la siguiente

Definición 2.4.4. Una clase C (S; K) es (esencialmente) minimal completa para el problema (Ω , S; K) si es (esencialmente) completa y ninguna subclase propia de C (S; K) es (esencialmente) completa.

De hecho una clase minimal completa sería la intersección de todas las clases completas para dicho problema. Ahora bien, esta intersección puede ser vacía y, por lo tanto, en general, no existirán clases minimales completas. De este problema trataremos a continuación. De momento podemos asegurar lo siguiente (su demostración es idéntica a la del teorema 1. de Ferguson, 1967, p. 56, ya que sirve para cualquier preórden parcial).

Teorema 2.4.1. Si existe una clase minimal completa C (S; K) para el problema (Ω , S; K), entonces C (S; K) = $\mathcal{A}$ (S; K).

El recíproco de este teorema no es en general cierto, pero, sin embargo se tiene el siguiente recíproco parcial.

Teorema 2.4.2. Si $\mathcal{A}$(S; K) existe y es completa, entonces $\mathcal{A}$ (S; K) es minimal completa.

Demostración. Si representamos por M C (S; K) a la clase minimal completa, por el Lema 2.4.2. se tiene que $\mathcal{A}$ (S; K) $\subset$ M C(S;K). Por otro lado como $\mathcal{A}$(S; K) es completa se tiene que M C (S; K) $\subset$ $\subset$ $\mathcal{A}$ (S; K), y, por lo tanto, $\mathcal{A}$ (S; K) = M C (S; K).

La existencia de la clase admisible y su relación con el concepto de frontera inferior, que vamos a definir seguidamente, depende de ciertas propiedades topológicas de S relacionadas con K y no dependen, en modo alguno, de otras propiedades - como pueden ser la convexidad o la cono-convexidad, etc. La definición siguiente es una generalización de la de Hartley (1978).

Definición 2.4.5. Se dice que S es K-compacto si para todo -- $f \in S$, el conjunto $K(f) \cap S$ es compacto.

La definición siguiente es una generalización de las - dadas por Ferguson (1967), Girón (1975) y Ríos (1976) y sirve - para cualquier cono de la clase $\mathcal{L}$ definida en el capítulo 3. Y, como veremos, contiene como caso particular a las dadas por los anteriores autores para el caso de los conos puntiagudos.

Definición 2.4.6. Se dice que $f \in \bar{S}$ pertenece a la K - frontera inferior de S si $K(f) \cap \bar{S} \subset I(f)$.

La K-frontera inferior de un conjunto S se representará por $\lambda(S; K)$ y su existencia y su relación con la clase -- $\mathcal{a}(S; K)$ se tratarán en los teoremas que siguen. De momento - señalemos que la definición anterior es equivalente a las siguientes como es fácil comprobar.

Lema 2.4.6. $f \in \lambda(S; K)$ es equivalente a las condiciones a), b) y c).

a) $f \in \bar{S}$ y $K(f) \cap \bar{S} = I(f) \cap \bar{S}$

b) $f \in \bar{S}$ y $M(f) \cap \bar{S} = \emptyset$

c) $(M(f) \cup \{f\}) \cap \bar{S} = \{f\}$

Además si K es puntiagudo $f \in \lambda(S; K)$ es equivalente a que $K(f) \cap \bar{S} = \{f\}$.

Definición 2.4.7. Un conjunto S se dice K-cerrado por abajo si $\lambda(S; K) \subset S$.

Los dos teoremas que sigen dan condiciones suficientes para la existencia de $\mathcal{a}(S; K)$ y de $\lambda(S; K)$, mientras que los teoremas 2.4.5. y 2.4.6. establecen la relación entre ambas.

Teorema 2.4.3. Si S es K-compacto entonces $\mathcal{a}(S; K) \neq \emptyset$.

Demostración. Sea $f \in S$ arbitario. Por hipótesis $K(f) \cap S$ es compacto. Vamos a demostrar en primer lugar, que si $g \in \mathcal{A}(K(f) \cap S;\, K)$ entonces $g \in \mathcal{A}(S;\, K)$, con lo cual el teorema se reduce a demostrar que si un conjunto es compacto, entonces posee estrategias admisibles.

Supongamos que $g \notin \mathcal{A}(S;\, K)$. Entonces existiría un $h \in S$ tal que $h \underset{K}{>} g$. Por otro lado $g \in K(f) \cap S$, luego $g \underset{K}{\geq} f$, y por transitividad, $h \underset{K}{>} f$, es decir $h \in M(f) \subset K(f)$ y de aquí que $h \in K(f) \cap S$, lo que contradiría el que g fuese admisible para el problema $(\Omega,\, K(f) \cap S;\, K)$.

Supongamos, pues, que S sea compacto. Por el teorema 2.3.11., sabemos que por ser $K \in \mathcal{C}$, existe al menos un $P \in K^{*+}$. Como S es compacto y el funcional $\int (\,.\,)\, d\,P$ es continuo, existirá al menos un $f_o \in S$ tal que:

$$\int f_o \, d\,P = \min_{f \in S} \int f \, d\,P$$

Vamos a ver ahora que $f_o \in \mathcal{A}(S;\, K)$. En efecto, si f_o no fuese admisible, existiría un $g \in S$ tal que $g \underset{K}{>} f_o$, es decir $g - f_o \in K \setminus L(K)$. Como $P \in K^{*+}$ se tendría que $\int (g - f_o) dP < o$ y de aquí que $\int g \, d\,P < \int f_o \, d\,P$, lo que contradiría la definición de f_o.

Teorema 2.4.4. Una condición suficiente para que S posea K-frontera inferior es que $\bar{S}$ sea K-compacto.

Demostración. Sea $f \in S$ cualquiera. Por hipótesis $K(f) \cap \bar{S}$ es compacto. Se demostrará primero, como en el teorema anterior, que si $g \in \lambda(K(f) \cap \bar{S};\, K)$, entonces $g \in \lambda(S;\, K)$, con lo cual sería suficiente demostrar que todo conjunto compacto posee K-frontera inferior.

Por hipótesis se tiene que $g \quad \overline{K(f) \cap \bar{S}} = K(f) \cap \bar{S}$ y, por lo tanto, $g \in \bar{S}$. Además por el lema 2.4.6.a),

$$K(g) \cap \overline{[K(f) \cap \bar{S}]} = I(g) \cap \overline{[K(f) \cap \bar{S}]},$$

que simplificada se convierte en

$K(g) \cap \bar{S} = I(g) \cap \bar{S}$, con lo cual, de nuevo por el lema 2.4.6.a), $g \in \mathcal{A}(S;\ K)$.

Supongamos pues que S sea compacto. Como en el teorema anterior sea $P \in K^{*+}$ cualquiera, y definamos f_o de la misma manera. Entonces $f_o \in \lambda(S;\ K)$.

Si no fuese así y por ser $f_o \in \bar{S} = S$, se tendrá, por el lema 2.4.6.b) que existirá un $g \in M(f_o) \cap \bar{S} = M(f_o) \cap S$, y por lo tanto $g - f_o \in K \setminus L(K)$. Como $P \in K^{*+}$ sería $\int (g - f_o)\, d\,P < 0$, o sea $\int g\, d\,P < \int f_o\, d\,P$, que contradiría la definición de f_o, toda vez que $g \in S$.

Vamos a estudiar ahora la relación entre las clases $\mathcal{A}(S;\ K)$ y el subconjunto $\lambda(S;\ K)$. Lo primero que hacemos notar es que mientras que $\lambda(S;\ K) = \lambda(\bar{S};\ K)$ como se deduce de su definición, no existe, en general relación alguna entre $\mathcal{A}(S;\ K)$ y $\mathcal{A}(\bar{S};\ K)$ a no ser que S sea cerrado; y como acabamos de ver en los dos teoremas precedentes las condiciones para la existencia de $\mathcal{A}(S;\ K)$ y de $\lambda(S;\ K)$ son distintas a no ser, de nuevo, que S fuese cerrado.

<u>Teorema 2.4.5.</u> Si S es K-cerrado por abajo entonces $\lambda(S;\ K) \subset \mathcal{A}(S;\ K)$.

<u>Demostración.</u> Sea $f \in \lambda(S;\ K) \subset S$. Si f no fuese admisible existiría un $g \in S$ tal que $g \underset{K}{\geq} f$. De aquí se tendría $g \in M(f) \cap S \subset M(f) \cap \bar{S}$ y por el lema 2.4.6.b), $f \notin \lambda(S;\ K)$. Contradicción que demuestra el teorema.

Obsérvese que el teorema no dice nada acerca de la existencia de $\lambda(S;\ K)$ y de $\mathcal{A}(S;\ K)$. Tampoco se puede prescindir de la hipótesis de que S sea K-cerrado por abajo como se ve facilmente con un contraejemplo.

Contraejemplo 2.4.1. Sea $\Omega = \{\theta_1, \theta_2\}$, $S = \{(x_1, x_2);$ $x_1 > 0, x_2 > 0\} \cup \{(x_1, x_2);\ x_1 = 0\ \ 0 < x_2 \leq 1\}$, y $K = K_o =$ $= \{(x_1, x_2),\ x_1 \leq 0;\ x_2 \leq 0$. Entonces se tiene que $\mathcal{A}(S;\ K) = \emptyset$ y, sin embargo $\lambda(S;\ K) = \{(0,0)\}$.

Un recíproco parcial de este teorema es el siguiente

Teorema 2.4.6. Si S es K-cerrado por abajo y $\bar{S}$ es K-compacto, entonces $\mathcal{A}(S;\ K) \subset \lambda(S;\ K)$.

Demostración. Sea $f \in \mathcal{A}(S;\ K)$ y supongamos que $f \notin \lambda(S;\ K)$. Definamos $S_1 = K(f) \cap \bar{S}$. Por hipótesis S_1 es compacto y por el teorema 2.4.4. posee K-frontera inferior $\lambda(S_1;\ K) \neq \emptyset$. En primer lugar $f \notin \lambda(S_1,\ K)$ pues caso contrario se tendría

$$K(f) \cap \bar{S}_1 = I(f) \cap \bar{S}_1$$

lo cual, sustituyendo S_1 por $K(f) \cap \bar{S}$, daría

$$K(f) \cap \bar{S} = I(f) \cap \bar{S}$$

y como $f \in S$, por el lema 2.4.6.a), se tendría que $f \in \lambda(S;\ K)$, en contra de la hipótesis inicial. Sea pues $g \in \lambda(S_1;\ K)$ cualquiera. Vamos a ver que $g \in S$ y que $g \underset{K}{\succ} f$, lo cual contradiría el que f fuese admisible, y por tanto, demostraría el teorema.

Por ser $g \in \lambda(S;\ K)$, en virtud del lema 2.4.6.b), se tendría que $g \in \bar{S}_1 = K(f) \cap \bar{S}$ y $M(g) \cap \bar{S}_1 = \emptyset$. De aquí se deduce que $g \in \bar{S}$ y $M(g) \cap \bar{S}_1 = M(g) \cap K(f) \cap \bar{S} = M(g) \cap \bar{S} =$ $= \emptyset$, y que de nuevo por el lema 2.4.6.b), $g \in \lambda(S;\ K)$. Como $\bar{S}$ es K-cerrado por abajo se tiene que $g \in S$.

Por otro lado, por ser $g \in K(f)$ sería $g \underset{K}{\succsim} f$, y de aquí $g \underset{K}{\succ} f$, pues si fuese $g \underset{K}{\sim} f$ se tendría que $M(g) = M(f)$ y como $M(g) \cap \bar{S} = \emptyset$, resultaría que $M(f) \cap \bar{S} = \emptyset$ y por el lema 2.4.6.b), que $f \in \lambda(S;\ K)$, en contra de la hipótesis.

<u>Corolario 2.4.1.</u> Si S es un conjunto K-cerrado por abajo y $\bar{S}$ - es K-compacto, entonces $\mathcal{A}(S;\ K) \neq \emptyset$ y además $\mathcal{A}(S,\ K) = \lambda(S;K)$.

<u>Corolario 2.4.2.</u> Si S es un conjunto K-cerrado por abajo y $\bar{S}$ es K-compacto, entonces $\lambda\ (S,\ K)$ es una clase minimal completa para el problema de decisión con información parcial $(\Omega\ ,\ S;\ K^{*})$, donde K^{*} es el polar de K.

Vamos a estudiar ahora el comportamiento de los preórdenes engendrados por los conos de la clase $\mathcal{C}$ respecto de la relación de inclusión y su relación con los subconjuntos $\mathcal{A}(S;\ K)$ y $\lambda\ (S;\ K)$.

<u>Definición 2.4.8.</u> Sean $\succsim$ y $\succsim^{*}$ dos preórdenes parciales definidos en $C\ (\Omega)$. Se dice que la relación $\succsim$ <u>está incluída en la</u> $\succsim^{*}$ si y solo si $f \succsim g$ implica que $f \succsim^{*} g$.

<u>Teorema 2.4.7.</u> Sean $K_1,\ K_2 \in \mathcal{C}$. Una condición necesaria y suficiente para que el preórden $\underset{K_1}{\succsim}$ esté incluído en el $\underset{K_2}{\succsim}$ es que $K_1 \subset K_2$.

<u>Demostración.</u> La suficiencia es clara pues si $K_1 \subset K_2$ se tiene que si $f - g \in K_1$ entonces $f - g \in K_2$, es decir, $f \underset{K_2}{\succsim} g$. Recíprocamente si $\underset{K_1}{\succsim}$ está incluído en $\underset{K_2}{\succsim}$ y no fuese $K_1 \subset K_2$, existiría al menos un $h \in K_1$ tal que $h \notin K_2$. Esto implicaría que p.e., $h \underset{K_1}{\succsim} 0$ y por hipótesis que $h \underset{K_2}{\succsim} 0$, es decir $h \in K_2$, contradicción que demuestra el teorema.

Como consecuencia de la dualidad se tiene el siguiente

<u>Corolario 2.4.3.</u> Si $K_1^{*},\ K_2^{*} \subset \Omega^{*}$ entonces $K_1^{*} \supset K_2^{*}$ implica que -- $\underset{K_1^{*}}{\succsim}$ está incluído en $\underset{K_2^{*}}{\succsim}$. Además si $K_1^{*},\ K_2^{*} \in \mathcal{C}$ entonces el recíproco es cierto.

<u>Lema 2.4.7.</u> Si S es K_1-compacto y $K_2 \subset K_1$, entonces S es K_2-compacto.

<u>Demostración.</u> Por hipótesis para todo $f \in S$, $K_1(f) \cap S$ es compacto. Como $K_2(f) \subset K_1(f)$ de aquí resulta que

$$K_2(f) \cap S = K_2(f) \cap \left[K_1(f) \cap S\right]$$

es cerrado y por ser subconjunto de un compacto es también compacto.

A partir de ahora al conjunto $K \setminus L(K) = M(o)$ lo representaremos, simplemente por M.

<u>Teorema 2.4.8.</u> Si $M_1 \subset M_2$ entonces $\mathcal{A}(S;\ K_1) \supset \mathcal{A}(S;\ K_2)$.

<u>Demostración.</u> Sea $f \in \mathcal{A}(S;\ K_2)$ y supongamos que $f \notin \mathcal{A}(S;\ K_1)$. Entonces existiría un $g \in S$ tal que $g \underset{K_1}{\succ} f$, es decir $g-f \in M_1 \subset M_2$, y por lo tanto $g \underset{K_2}{\succ} f$, en contra de la hipótesis.

Para que el teorema sea cierto no se puede sustituir la hipótesis por la más débil de que $K_1 \subset K_2$, como se puede ver - facilmente con un contraejemplo. Se tiene sin embargo el siguiente corolario.

<u>Corolario 2.4.4.</u> Si K_1, $K_2 \in \mathcal{G}$ son puntiagudos, o bien, si -- K_1^*, $K_2^* \in \mathcal{G}^*$ son completos, entonces $K_1 \subset K_2$ implica $\mathcal{A}(S;K_1) \supset \mathcal{A}(S;\ K_2)$.

Resultados análogos son válidos para $\lambda(S;\ K)$.

<u>Teorema 2.4.9.</u> Si $M_1 \subset M_2$ entonces $\lambda(S;\ K_1) \supset \lambda(S;\ K_2)$.

<u>Demostración.</u> Si $f \in \lambda(S;\ K_2)$, por el lema 2.4.6.b), $M_2(f) \cap \bar{S} = \emptyset$ y como $M_1(f) \subset M_2(f)$ de aquí que $M_1(f) \cap \bar{S} = \emptyset$ y como $f \in \bar{S}$ se tiene que $f \in \lambda(S;\ K_1)$.

<u>Corolario 2.4.5.</u> Si K_1, $K_2 \in \mathcal{G}$ son puntiagudos, entonces $K_1 \subset K_2$ implica $\lambda(S;\ K_1) \supset \lambda(S;\ K_2)$.

Tampoco en el teorema 2.4.9. se puede sustituir la condición $M_1 \subset M_2$ por la $K_1 \subset K_2$ como demuestra el siguiente con--

traejemplo.

Contraejemplo 2.4.2. Sea $\Omega = \{\theta_1, \theta_2\}$, $C(\Omega) = \mathbb{R}^2$. $S = \{(x_1, x_2);\ x_1 \geq 0\}$ y sea $K_1 = \{(x_1, x_2);\ x_1 \leq 0;\ x_2 \leq 0\}$ y $K_2 = \{(x_1, x_2);\ x_1 \leq 0\}$. Claramente $K_1 \subset K_2$ pero sin embargo $\lambda(S; K_1) = \emptyset$ y $\lambda(S; K_2) = \{(x_1, x_2);\ x_1 = 0\}$. Obviamente $M_1 = K_1 - \{(0,0)\}$ no está contenido en $M_2 = \{(x_1, x_2);\ x_1 < 0\}$.

Corolario 2.4.6. Si S está K_1-cerrado por abajo y $M_1 \subset M_2$, entonces S está K_2-cerrado por abajo. Si además K_1 y K_2 son puntiagudos, entonces $K_1 \subset K_2$ implica que si S está K_1-cerrado por abajo, entonces está K_2-cerrado por abajo.

Hasta ahora hemos estudiado el problema de la admisibilidad y su relación con la frontera inferior del conjunto de riesgo, pero no disponemos de medios efectivos de calcular bien $\mathcal{A}(S; K)$ o bien $\lambda(S; K)$. Vamos ahora a definir una nueva clase de estrategias que, de hecho, van a ser estrategias Bayes respecto de las posibles distribuciones a priori del conjunto de incertidumbre; estrategias fáciles de calcular y nuestro propósito será establecer relaciones entre esta clase y las hasta ahora consideradas.

Definición 2.4.9. Una estrategia $f \in S$ es K^*- Bayes para el problema $(\Omega, S; K^*)$ si existe al menos una $P \in K^*$ tal que f es Bayes respecto a P.

Al conjunto de todas las estrategias K^*- Bayes las representaremos por $\mathcal{B}(S; K^*)$ ó $\mathcal{B}(S; K)$ donde K es el polar restringido de K^* . Al conjunto de todas las estrategias de S que son Bayes frente a las distribuciones de K^{*+} lo representaremos por $\mathcal{B}^+(S; K^*) = \mathcal{B}^+(S; K^{*+})$, o simplemente por $\mathcal{B}^+(S; K)$.

El problema de dar condiciones necesarias y suficientes para la existencia de soluciones Bayes es un problema complicado y no resuelto, y menos, como en nuestro caso cuando se prescinde de la convexidad. En Girón ((1977), pp. 73-75) se dan algunas condiciones suficientes para la existencia de la frontera bayesiana, íntimamente relacionada con las clases $\mathcal{B}^+(S; K^*)$ y $\mathcal{B}(S; K)$. También en Ferguson ((1967), pp. 67-69), se dan condiciones suficientes para la existencia de reglas de Bayes, en el caso de que Ω sea finito.

La condición de K-compacidad que nos ha sido tan útil a la hora de asegurar la existencia de decisiones K-admisibles no es suficiente para garantizar la existencia de soluciones Bayes como muestra el siguiente contraejemplo:

Contraejemplo 2.4.3. Sea $\Omega = \{\theta_1, \theta_2\}$ y $S \subset S_1 \cup S_2$ donde $S_1 = \{(x_1, x_2);\ x_1 + 2x_2 \geq 0\}$; $S_2 = \{(x_1, x_2);\ 2x_1 + x_2 \geq 0\}$ y $K = \{(x_1, x_2);\ x_1 \leq 0\ \ x_2 \leq 0\} = K_o$. Claramente se tiene que S es K-compacto, de hecho,

$$\mathcal{A}(S, K) = \{(x_1, x_2);\ x_1 + 2x_2 = 0;\ x_1 \leq 0\} \cup \{(x_1, x_2);\ 2x_1 + x_2 = 0;\ x_1 \geq 0\}$$ y sin embargo $\mathcal{B}(S; K) = \emptyset$.

El teorema siguiente da una condición suficiente, que a veces es fácil de comprobar, que asegura la existencia de reglas Bayes respecto de una cierta distribución a priori.

Si $P \in \Omega^*$, representaremos por H_p el polar restringido de P, es decir, $H_p = \{f \in C(\Omega);\ \int f dP \leq 0\}$.

Teorema 2.4.10. Una condición suficiente para que el problema $(\Omega; S)$ posea solución Bayes respecto de la distribución a priori $P \in \Omega^*$, es que S sea H_p-compacto.

Demostración. Consideremos la familia de subconjuntos de S formada por $\{H_p(f) \cap S\}$ cuando $f \in S$. Por hipótesis, todos los miembros de la familia son subconjuntos compactos y que poseen la propiedad de la intersección finita; por lo tanto su in tersección es no vacía $\bigcap_{f \in S} \{ H_p(f) \cap S \}$ y es precisamente igual a $\mathcal{B}(S; H_p)$.

La condición no es, sin embargo, necesaria como muestra el siguiente

Contraejemplo 2.4.4. Sea $\Omega = \{ \vartheta_1, \vartheta_2 \}$; $S = \{ (x_1, x_2);\ x_1 + x_2 \geq 0 \}$ y sea $P_o(\vartheta_1) = P_o(\vartheta_2) = \frac{1}{2}$. El conjunto S no es H_p-compacto para ningún $P \in \Omega^*$, sin embargo:

$$\mathcal{B}^+(S, H_{P_o}) = \{ (x_1, x_2);\ x_1 + x_2 = 0 \}$$

Corolario 2.4.7. Una condición suficiente para que las clases $\mathcal{B}^+(S; K)$ y $\mathcal{B}(S, K)$ sean no vacías, es que exista al menos un $P \in K^{*+}$, respectivamente un $P \in K^*$, tal que S sea H_p-compacto.

Obsérvese que si S es H_p-compacto para todo $P \in K^*$, entonces S es K-compacto, donde $K = \mathrm{pr}(K^*)$, como se deduce facilmente de que

$$K = \bigcap_{P \in K^*} H_p$$

aunque el recíproco no sea, evidentemente, cierto. De hecho por el teorema de Choquet es suficiente que S sea H_p-compacto para todo $P \in K_e^*$ para que S sea K-compacto.

Una vez establecida la existencia de $\mathcal{B}^+(S; K)$ vamos a relacionar esta clase con la $\mathcal{A}(S; K)$.

Teorema 2.4.11. $\mathcal{B}^+(S; K) \subset \mathcal{A}(S; K)$

Demostración. Sea $f_o \in \mathcal{B}^+(S; K)$ y supongamos que f_o no fuese admisible. Existiría entonces un $g_o \in S$ tal que $g_o \in M(f_o)$, es decir $g_o - f_o \in M$. Como $f_o \in \mathcal{B}^+(S; K)$, existe al menos un $P_o \in K^{*+}$ tal que

$$\int f_o \, d P_o = \operatorname*{Min}_{g \in S} \int g \, d P_o$$

Por otro lado por ser $g - f_o \in M = K - L(K)$ y por ser $P_o \in K^{*+}$ se tiene que $\int (g_o - f_o)\, d\, P_o < 0$, es decir, $\int g_o\, d\, P_o < \int f_o\, d\, P_o$, que contradice el que f_o sea Bayes respecto de P_o.

Obsérvese que para la validez del teorema no se ha exigido condición alguna de convexidad o cono-convexidad al conjunto S.

Si comparamos este teorema con el teorema 2.4.3., vemos que la demostración de la segunda parte de este teorema es idéntica a la del 2.4.11., con lo cual las condiciones del teorema 2.4.3. pueden simplificarse y exigir simplemente la H_p-compacidad de S para algún $P \in K^{*+}$.

Para obtener más información acerca del conjunto $\mathcal{A}(S;K)$ será necesario, a partir de ahora, imponer más hipótesis en el conjunto S. Como hemos advertido anteriormente la hipótesis de convexidad de S es demasiado restrictiva y la sustituímos por la de cono-convexidad dada por Yu (1974).

<u>Definición 2.4.10.</u> Se dice que S es <u>K-convexo</u> si S - K es convexo, donde

$$S - K = \{ g \in C(\Omega);\ g = f - h \quad \text{con } f \in S;\ h \in K \}$$

<u>Lema 2.4.8.</u> Si S es K_1-convexo y $K_2 \supset K_1$, entonces S es K_2-convexo.

<u>Demostración.</u> Por hipótesis $S - K_1$ es convexo. Además $S - K_2 = (S - K_1) - K_2$ por ser $K_2 \supset K_1$. Y como la suma de dos conjuntos convexos es un conjunto convexo, de aquí resulta que $S - K_2$ es convexo y por lo tanto S es K_2 - convexo.

<u>Lema 2.4.9.</u> $\mathcal{A}(S;\ K) \subset \mathcal{A}(S - K;\ K)$. Además si K es puntiagudo, entonces $\mathcal{A}(S;\ K) = \mathcal{A}(S - K;\ K)$.

<u>Demostración.</u> Sea $f \in \mathcal{A}(S;\ K)$. Si $f \notin \mathcal{A}(S - K;\ K)$ existiría

un $g \in S - K$ tal que $g \succ_K f$. Por ser $g \in S - K$, existiría un $h \in S$ tal que $h \succ_K g$, y por transitividad se tendría que $h \succ_K f$, que contradiría el que f fuese admisible para el problema (Ω ; S).

Supongamos ahora que K fuese puntiagudo y que $f \in \mathcal{A}(S-K;K)$. Por ser $f \in S - K$ pueden ocurrir dos cosas: a) $f \in S$ en cuyo caso si $f \notin \mathcal{A}(S;K)$, existiría un $g \in S$ tal que $g \succ_K f$; y como $g \in S$ esto implica que $g \in S - K$ (por ser $0 \in K$) lo que contradiría la hipótesis; b) $f \notin S$, entonces $f = s - k$ con $s \in S$ y $k \in K$ con $k \neq 0$. Como K es puntiagudo, esto implica que $s \succ_K f$ y de nuevo, como $s \in S - K$, esto contradiría el que f fuese admisible para el problema (Ω ; S - K).

<u>Teorema 2.4.12.</u> Si Ses K-convexo, entonces $\mathcal{A}(S;K) \subset \mathcal{B}(S;K)$.

<u>Demostración.</u> Sea $f_o \in \mathcal{A}(S;K)$. Por la primera parte del lema anterior se tiene que los conjuntos $M(f_o)$ y S-K son convexos y disjuntos y además el interior de $M(f_o)$ es no vacío (de hecho el interior contiene al conjunto abierto $f_o + \mathring{K}_o$). Estamos entonces en condiciones de aplicar el teorema de separación débil (véase, p.e., Girón, (1977), p. 99). Existe por lo tanto un funcional lineal continuo no nulo $X^* \in C(\Omega)$ y una constante c tales que

$$X^*(f) \leq c \qquad \text{para todo} \quad f \in \overline{M(f_o)} = K(f_o)$$

y

$$X^*(f) \geq c \qquad \text{para todo} \quad f \in S - K$$

Ahora, como $f_o \in K(f_o)$ y $f_o \in S$, de las desigualdades anteriores se tiene que $X^*(f_o) = c$ y por lo tanto la primera de ellos se puede escribir

$$X^*(f) \leq 0 \qquad \text{para todo} \quad f \in K.$$

Si ahora normalizamos X^*, $\|X^*\| = 1$, y aplicamos el teorema de representación de Riesz, tenemos que

$$X^*(f) = \int f \, dP^*$$

La primera desigualdad se reduce a

$$\int f \, d\, P^{*} \leq 0 \qquad \text{para todo} \quad f \in K$$

es decir, $P^{*} \in K^{*}_{,}$ y la segunda desigualdad se reduce a

$$\int f \, d\, P^{*} \geq \int f_{o} \, d\, P^{*} \qquad \text{para todo } f \in S - K$$

que implica que $f_{o} \in \mathcal{B}(S;\ K)$.

<u>Corolario 2.4.8.</u> Si S es K-convexo, entonces $\mathcal{B}^{+}(S;\ K) \subset$ $\subset \mathcal{A}(S;\ K) \subset \mathcal{B}(S;K)$.

Como consecuencia de los teoremas 2.4.3. y 2.4.12. se - tiene el siguiente e importante

<u>Corolario 2.4.9.</u> Si S es K-convexo y K-compacto, entonces -- $\mathcal{B}(S;\ K) \neq \emptyset$.

Este último corolario es importante pues generaliza el teorema 2.4.10. cuando la condición de K-compacidad de S se complementa con la K-convexidad. Obsérvese, de paso que en el contraejemplo 2.4.3. el conjunto S no es K-convexo.

Otra consecuencia interesante del teorema se refiere al caso bayesiano, es decir, cuando $K^{*} = \{ P \}$. En este caso, como consecuencia directa del corolario 2.4.8. se tiene por ser - $K^{*+} = K^{*} = \{ P \}$ el siguiente

<u>Corolario 2.4.10.</u> Si $K^{*} = \{ P \}$ entonces

$$\mathcal{B}^{+}(S;\ P) = \mathcal{A}(S;\ P) = \mathcal{B}(S;\ P).$$

<u>Demostración.</u> Como siempre se tiene que $\mathcal{B}^{+}(S;\ P) \subset \mathcal{A}(S;\ P)$ es suficiente demostrar que, cualquiera que sea $S \subset C(\Omega)$ siempre es H_{p}-convexo. Ahora bien

$$S - H_{p} = \bigcup_{f \in S} H_{p}^{+}(f)$$

donde $H_p^+(f) = \{ g \in C(\Omega);\ \int f\, d P \leq \int g\, d P \}$, y por lo tanto $\bigcup_{f \in S} H_p^+(f)$ es convexo. Por lo tanto $\mathcal{A}(S;\ P) \subset \mathcal{B}(S,P)$ y como $\mathcal{B}^+(S;\ P) = \mathcal{B}(S;\ P)$, de aquí resulta el corolario.

Así pues en el caso bayesiano, los conceptos de admisibilidad y de ser Bayes coinciden.

<u>Corolario 2.4.11.</u> Si S es K-cerrado por abajo y K-convexo y $\bar{S}$ es K-compacto, entonces $\mathcal{B}(S;\ K)$ es una clase completa para el problema $(\Omega,\ S;\ K)$ y $\mathcal{A}(S;\ K)$ es una clase minimal completa.

Hasta ahora hemos visto que si S es K-convexo entonces la clase K-admisible está contenida en la K^*-Bayes, y que las K^{*+} Bayes son admisibles.

Los dos resultados siguientes permiten afinar más estos resultados bajo ciertas hipótesis que son una generalización de la noción de convexidad estricta de un conjunto.

Precisamente vamos a considerar el análogo de la K-frontera inferior de un conjunto, $\lambda(S;\ K)$ -que jugaba un papel importante en su relación con la clase admisible $\mathcal{A}(S;\ K)$- y que va a jugar un papel análogo respecto de la clase de las estrategias K-Bayes, $\mathcal{B}(S;\ K)$.

<u>Definición 2.4.11.</u> Llamaremos <u>frontera K-bayesiana de</u> S y la representaremos por $\mu(S;\ K)$, a la clase de las estrategias K-Bayes del problema $(\Omega,\ \bar{S};\ K)$, donde $K = p(K)$, es decir $\mu(S;\ K) = \mathcal{B}(\bar{S};\ K)$.

Analogamente se definirá $\mu^+(S;\ K) = \mathcal{B}(\bar{S};\ K^{*+})$.

Obsérvese que todos los teoremas dados para la clase $\mathcal{B}(S;\ K)$ son igualmente válidos para la K-frontera bayesiana sin más que sustituir las hipótesis de S por $\bar{S}$. En particular se tienen los siguientes

Teorema 2.4.13. Para todo S se tiene que $\mu^{+}(S;\ K) \subset \lambda(S;\ K)$. Además si $\bar{S}$ es K-convexo, entonces $\lambda(S;\ K) \subset \mu(S;\ K)$.

Teorema 2.4.14. Una condición suficiente para que $\mu(S;\ K) \neq \emptyset$ es que $\bar{S}$ sea K-convexo y K-compacto.

Definición 2.4.12. Un conjunto S se dice que está K-bayesiana-- mente cerrado por abajo si $\mu(S;\ K) \subset S$.

De los anteriores teoremas y de la última definición se deduce inmediatamente el siguiente

Corolario 2.4.12. Si $\bar{S}$ es K-convexo y está K-bayesianamente cerrado por abajo, entonces S está K-cerrado por abajo.

La relación entre $\mathcal{B}(S;K)$ y $\mu(S;K)$ es análoga a la existente entre $\mathcal{A}(S;K)$ y $\mu(S;K)$. En particular tenemos el equivalente del corolario 2.4.1. que constituye el

Teorema 2.4.15. Si S está K-bayesianamente cerrado por abajo y $\bar{S}$ es K-compacto, entonces $\mathcal{B}(S,K) \neq \emptyset$ y además $\mu(S;K) = \mathcal{B}(S;K)$.

Obsérvese que en estas circunstancias (las del teorema 2.4.15.) la clase $\mu(S;K)$ es una clase completa y $\lambda(S;K)$ es - una clase minimal completa del problema $(\Omega, S;\ K)$.

El análogo del lema 2.4.9. es el siguiente que es igualmente válido para $\mu(S;K)$ y $\mu^{+}(S;K)$.

Lema 2.4.10. $\mathcal{B}(S;K) \subset \mathcal{B}(S\text{-}K;K)$. Además si K es puntiagudo entonces $\mathcal{B}^{+}(S\text{-}K;K) = \mathcal{B}^{+}(S;K)$.

Demostración. Sea $f_0 \in \mathcal{B}(s;K)$. Por definición existe al menos - un $P \in K^{*}$ tal que

$$\int f_0 \, d\,P = \underset{f \in S}{\mathrm{Min}} \int f \, d\,P$$

Ahora bien para todo $k \in K$ se verifica que $\int k \, d\, P \leq 0$ y como $0 \in K$, de aquí se deduce que

$$\operatorname*{Min}_{f \in S} \int f \, d\, P = \operatorname*{Min}_{\substack{f \in S \\ k \in K}} \int (f - k) \, d\, P$$

y por lo tanto, $f_o \in \mathcal{B}(S - K;\, K)$.

Como no se ha utilizado en absoluto el hecho de que -- $P \in K^{*+}$, se tendrá analogamente que $\mathcal{B}^{+}(S;\, K) \subset \mathcal{B}^{+}(S - K;\, K)$.

Supongamos ahora que K fuese puntiagudo y que $P \in K^{*+}$. Sea $f_o \in \mathcal{B}(S - K;\, P)$. Como $\int k \, d\, P \leq 0$ para todo $k \in K$ se tie ne que

$$\operatorname*{Min}_{f \in S-K} \int f \, d\, P = \operatorname*{Min}_{g \in S} \int g \, d\, P$$

con lo cual

$$\int f_o \, d\, P = \operatorname*{Min}_{g \in S} \int g \, d\, P$$

Esto implica que $f_o \in S$ pues si fuese $f_o = g_o - K_o$ con $g_o \in S$ y $K_o \neq 0$ se tendría que, por una parte

$$\int f_o \, d\, P \leq \int g_o \, d\, P = \int (f_o + K_o) \, d\, P$$

que implicaría $\int K_o \, d\, P \geq 0$. Como $K_o \in K - \{0\}$ y $P \in K^{*+}$ se - tendría que $\int K_o \, d\, P < 0$ contradicción que demuestra la segunda parte del lema.

La segunda conclusión del lema no es válida para $\mathcal{B}(S;K)$ como muestra el siguiente

<u>Contraejemplo 2.4.5.</u> Sea $\Omega = \{\theta_1, \theta_2\}$; $S = \{(0,0)\}$ y $K = K_o = \{(x_1, x_2);\ x_1 \leq 0;\ x_2 \leq 0\}$. K es puntiagudo y $S - K = \{(x_1, x_2);\ x_1 \geq 0;\ x_2 \geq 0\}$. $\mathcal{B}(S,K) = S = \{(0,0)\}$, mien tras que $\mathcal{B}(S - K;K) = \{(x_1, x_2);\ x_1 = 0,\ x_2 \geq 0\} \cup \{(x_1,x_2);\ x_1 \geq 0;\ x_2 = 0\}$.

La definición siguiente generaliza el concepto de convexidad estricta y es útil a la hora de comprobar el que ciertas estrategias Bayes (no necesariamente respecto de K^{*+}) son admisibles.

Definición 2.4.12. Un conjunto S se dice que es estrictamente K-convexo si $\bar{S}$ es K-convexo y además para todo f, $g \in \mu(S;K)$ con $f \neq g$ y $\lambda \in (0,1)$, se tiene que $\lambda f + (1-\lambda) g \notin \mu(S;K)$.

Teorema 2.4.16. Si S es estrictamente K-convexo, entonces para todo $P \in K^*$, si $\mu(S;H_p) \neq \emptyset$ está constituído por una sola estrategia, digamos $\{f_o\}$, y además $f_o \in \lambda(S;K)$.

Demostración. Supongamos que $g_o \neq f_o$ fuese tal que $g_o \in \mu(S;H_p)$. Entonces se tendría que, por ser $\mu(S;H_p)$ convexo $\lambda f_o + (1-\lambda)g_o \in \mu(S;H_p) \subset \mu(S;K)$ para todo $\lambda \in [0,1]$, y esto contradiría el que S fuese estrictamente K-convexo.

Supongamos ahora que $f_o \notin \lambda(S;K)$. Como $f_o \in \bar{S}$ (por ser $f_o \in \mu(S;K)$) se tendrá por el lema 2.4.6.b) que $M(f_o) \cap \bar{S} \neq \emptyset$. Sea $g_o \in M(f_o) \cap \bar{S}$. Entonces $g_o - f_o \in K - L(K)$ con lo cual

$$\int g_o \, dP \leq \int f_o \, dP$$

y esto implicaría que $g_o \in \mu(S;K) = \{f_o\}$ por la primera parte del teorema. Contradicción que demuestra la segunda parte.

Corolario 2.4.13. Si S es estrictamente K-convexo y posee K-frontera inferior $\lambda(S;K)$, entonces $\lambda(S;K) = \mu(S;K)$.

Demostración. Por el teorema 2.4.13. se tiene que $\lambda(S;K) \subset \mu(S;K)$. Para demostrar el recíproco sea $f_o \in \mu(S;K)$. Esto implica que existe un $P \in K^*$ tal que $f_o \in \mu(S;H_p)$ y por el teorema 2.4.16., $f_o \in \lambda(S;K)$; con lo cual $\lambda(S;K) = \mu(S;K)$.

Los últimos resultados se refieren al comportamiento de los conjuntos $\lambda(S;K)$ y $\mu(S;K)$ respecto a variaciones en el conjunto K, o mejor dicho, respecto del conjunto K^*.

<u>Teorema 2.4.17.</u> Si la sucesión decreciente K_n^*, con $K_n^* \in \mathcal{K}^*$; converge hacia $K^* = \bigcap_{n=1}^{\infty} K_n^*$, entonces la sucesión $\mu(S;K_n)$ (donde $K_n = p\ r\ (K_n^*)$) es decreciente y converge hacia $\mu(S;K)$ (donde $K = p\ r\ (K^*)$), es decir

$$\mu(S;\ K) = \bigcap_{i=1}^{\infty} \mu(S;K_n)$$

<u>Demostración.</u> Claramente, por ser K_n^* decreciente, también lo es $\mu(S;\ K_n)$ y por ser $K^* \subset K_n^*$ para todo n , se tiene que - $\mu(S;\ K) \subset \mu(S;K_n)$ para todo n y por lo tanto

$$\mu(S;\ K) \subset \bigcap_{n=1}^{\infty} \mu(S;\ K_n)$$

Recíprocamente si $f_o \in \bigcap_{n=1}^{\infty} \mu(S;K)$, existe una sucesión $\{P_n\}$ con $P_n \in K_n^*$, tal que f_o es Bayes respecto de todos los P_n. Como $\{P_n\} \subset K_1^*$ y éste es compacto, existe una subsucesión $\{P_{n_k}\} \xrightarrow{w} P_o$. Vamos a demostrar en primer lugar que $P_o \in K^*$ con lo cual $f_o \in \mu(S;\ K)$.

Por construcción se tiene que

$$\int f_o\ d\ P_{n_k} \leq \int g\ d\ P_{n_k} \qquad \text{para todo } g \in \overline{S}$$

y de aquí, por ser $\{P_{n_k}\} \xrightarrow{w} P_o$, queda

$$\int f_o\ d\ P_o \leq \int g\ d\ P_o \qquad \text{para todo } g \in \overline{S}$$

es decir $f_o \in \mu(S;\ P_o)$.

Supongamos ahora que $P_o \notin K^* = \bigcap_{n=1}^{\infty} K_n^*$. Entonces existirá un cierto n_o tal que $P_o \notin K_n^*$ para $n \geq n_o$. Si ahora aplicamos el lema 2.3.3. a los conjuntos $\{P_o\}$ y $K_{n_o}^*$, existirá un $g_o \in C\ (\Omega)$ tal que

$$\int g_o\ d\ P_o > 0$$

y

$$\int g_o\ d\ P < 0 \qquad \text{para todo } P \in K_{n_o}^*$$

En particular, existirá un n_{k_o} tal que $\int g_o \, d\, P_{n_k} < 0$ para todo $n_k \geq n_o$, de donde, por la convergencia débil, se tendría que $\int g_o \, d\, P_o \leq 0$, contradicción que termina la demostración del teorema.

El análogo del teorema 2.4.17. para $\lambda\,(S;K)$ no es cierto en general. Sin embargo tenemos el siguiente resultado.

<u>Teorema 2.4.18.</u> Si la sucesión K_n^*, con $K_n^* \in \mathcal{E}$, es decreciente y los K_n^* son totales o completos para todo n, y $K^* = \lim_n K_n^*$ es tal que $K^* \subset K_n^*$ para todo n, entonces $\lambda\,(S;\,K_n)$ (donde $K_n = pr(K_n)$) es decreciente y converge hacia $\mu(S;\,K)$ (donde $K = p\,r\,(K^*)$), es decir

$$\mu\,(S;K) = \bigcap_{i=1}^{\infty} \lambda\,(S;\,K_n).$$

si S es K_1-convexo.

<u>Demostración.</u> Por ser S K_1-convexo, es K_n-convexo para todo n, por el lema 2.4.8. y por lo tanto $\lambda(S;K_n) \subset \mu\,(S;K_n)$. Por otro lado como $K^* \subset K_n^{*+}$ se tiene que $\mu(S;K) \subset \lambda\,(S;K_n)$ por el teorema 2.4.11. aplicado a $\bar{S}$. De aquí resulta que

$$\mu\,(S;\,K) \subset \lambda\,(S;K_n) \quad \text{para todo } n \text{, luego}$$

$$\mu(S;\,K) = \bigcap_{n=1}^{\infty} \lambda\,(S;\,K_n)$$

Además, por ser los K_n puntiagudos $\{\lambda(S;\,K_n)\}$ es decreciente.

Recíprocamente, si $f \in \bigcap_{n=1} \lambda\,(S;K_n)$ se tiene que, para todo n, $f \in \lambda\,(S;\,K_n) \subset \mu\,(S;\,K_n)$. Entonces existe una sucesión $\{P_n\}$ con $P_n \in K_n^*$ tal que $f \in \mu(S;\,P_n)$ y exactamente igual que en el teorema anterior, se concluye que $f \in \mu\,(S;\,K)$.

La hipótesis de que $K^* \subset K_n^{*+}$ no puede excluirse del teorema anterior, como muestra el contraejemplo siguiente.

<u>Contraejemplo 2.4.6.</u> Sea $\Omega = \{\vartheta_1, \vartheta_2\}$, $S = \{(x_1,\, x_2);$ --

$x_1 + x_2 - 1 \geq 0;\ x_1 \leq 1;\ x_2 \leq 1 \}$ que es convexo sea $P_o = (\frac{1}{2}, \frac{1}{2})$ y $P_n = (\frac{n-1}{2n}, \frac{n+1}{2n})$ y $K_n^* = \text{co}\ (P_o, P_n)$. Los K_n^* son totales y además $\lim K_n^* = P_o$. Sin embargo $P_o \notin K^{*+}$ y, se tiene que $\lambda(S; K_n^*) = \{(1,0)\}$ y $\lambda(S; P_o) = \{(x_1, x_2);\ x_1 + x_2 = 1;\ x_1 \geq 0;\ x_2 \geq 0\}$.

Tampoco se puede deducir del teorema anterior que $\lim \lambda(S; K_n) = \lambda(S; K)$ a no ser bajo ciertas hipótesis.

<u>Corolario 2.4.14.</u> Si la sucesión $K_n^* \downarrow \{P\}$, donde K_n^* son totales o completos; $P \in K_n^{*+}$ para todo n y S es K_1-convexo, entonces $\lambda(S; K) = \lim \lambda(S; K_n)$ donde $K_n = \text{pr}\ (K_n^*)$ y $K = H_p$.

<u>Demostración.</u> Por el teorema anterior se tiene que $\mu(S; H_p) = \bigcap_{n=1}^{\infty} \lambda(S; K_n) = \lim \lambda(S; K_n)$. Por el corolario 2.4.10. aplicado a $\overline{S}$, se tiene que $\lambda(S; H_p) = \mu(S; H_p)$; luego $\lambda(S; H_p) = \lim \lambda(S; K_n)$.

Otra instancia para la cual se cumple lo anterior viene dada por el siguiente

<u>Corolario 2.4.15.</u> Si la sucesión $K_n^* \downarrow K^*$ y S posee K_1-frontera inferior y es estrictamente K_1-convexo por abajo, entonces -- $\lambda(S, K) = \lim \lambda(S; K_n)$ donde $K = \text{pr}\ (K^*)$ y $K_n = \text{pr}\ (K_n^*)$.

<u>Demostración.</u> Por el teorema 2.4.17. resulta que $\mu(s, K) = \lim \mu(S; K_n)$. Por otro lado, S posee frontera K_n-inferior (y K-inferior) y es estrictamente K_n-convexo (y K-convexo) por abajo. Entonces, por el corolario 2.4.13. resulta que $\lambda(S; K_n) = \mu(S; K_n)$ para todo n y $\lambda(S; K) = \mu(S; K)$, de donde resulta el corolario.

Los análogos de los teoremas y corolarios anteriores no tienen equivalentes para sucesiones crecientes puesto que no podemos asegurar que si $K_n^* \uparrow K^*$ con $K_n^* \in \mathcal{E}^*$ su límite K^* pertenezca a $\mathcal{E}^*$.

Otra forma de enfocar el problema sería el considerar - el espacio topológico (Ω^*, w) dotado de su correspondiente métrica de Prohorv, ρ , y sobre esta métrica construir la métrica de Hausdorff para los elementos de la clase $\mathcal{K}^*$ y considerar la continuidad de las multiaplicaciones $\lambda(S; K^*)$ y $\mu(S; K^*)$ de $\mathcal{K}^*$ en $\mathcal{P}(S)$.

En vez de seguir este enfoque alternativo concluímos esta memoria dando un teorema referente al comportamiento de la frontera $\mu^+(S; K)$ respecto a las sucesiones decrecientes dentro de la clase $\mathcal{K}$. Como demostraremos con contraejemplos estos resultados no son válidos para $\lambda(S; K)$, ni para $\mu(S; K)$, ni para el caso de Ω arbitrario en general. La validez solo se puede demostrar en el caso de que Ω sea finito.

La explicación de por qué los resultados referentes a K^* y K son distintos, está, por un lado en que la polaridad o dualidad no conserva la continuidad, es decir, el ser $K_n^* \downarrow K^*$ no implica pr $(K_n^*) \uparrow$ pr (K^*) ni vice-versa, si $K_n \downarrow K$ no <u>necesariamente</u> es $K_n^* \uparrow K^*$. Sin embargo cuando Ω es finito, se verifica el teorema siguiente.

<u>Teorema 2.4.19.</u> Sea Ω finito. Entonces si $K_n \in \mathcal{K}$ para todo n, y son puntiagudos y si $K_n \downarrow K$, es decir $K = \bigcap_{n=1}^{\infty} K_n$, entonces $K \in \mathcal{K}$ y es puntiagudo y $K_n^{*+} \uparrow K^{*+}$, es decir $K^{*+} = \bigcup_{n=1}^{\infty} K_n^{*+}$.

<u>Demostración.</u> Es evidente que $K \in \mathcal{K}$ y además es puntiagudo. Por otro lado, para todo n, es $K_n^{*+} \subset K_{n+1}^{*+}$ pues si $P \in K_n^{*+}$ se tiene que $\int f \, d P < 0$ para todo $f \in K_n - \{0\}$, luego en particular para todo $f \in K_{n+1} - \{0\} \subset K_n - \{0\}$, y por lo tanto $P \in K_n^{*+}$. Por la misma razón $K_n^{*+} \subset K^{*+}$ y por lo tanto $K^{*+} \supset \bigcup_{n=1}^{\infty} K^{*+}$.

Recíprocamente, sea $P_o \in K^{*+}$ y supongamos que $P_o \in \bigcup_{n=1}^{\infty} K_n^{*+}$. Esto implicaría la existencia de una sucesión $f_n \in K_n - \{0\}$ tal que $\int f_n \, d P_o \geq 0$ para todo n. Sin pérdida de generalidad podemos escoger los f_n tales que pertenezcan a la esfera unidad de C (Ω), es decir $\| f_n \| = 1$. Como esta esfera es compacta (por ser Ω finito) existe una subsucesión $\{f_{n_k}\} \longrightarrow f_o$. Además $f_o \neq 0$ por ser $\| f_o \| = 1$. Además aplicando el teorema de la convergencia dominada a la sucesión $\{f_{n_k}\}$ y a f_o se tiene que $\int f_o \, d\, P_o \geq 0$. Demostremos ahora que $f_o \in K$, con lo cual se tendría una contradicción pues por ser $f_o \in K - \{0\}$ y $P_o \in K^{*+}$ sería $\int f_o \, d\, P_o < 0$, que contradice lo anterior.

Si no fuese $f_o \in K$, existiría un n_1 tal que $f_o \notin K_{n_1}$ y el teorema de Klee aplicado a los conos $\{\lambda f_o;\ \lambda \geq 0\}$ y K_{n_1} nos aseguraría la existencia de un $P_1 \in K_{n_1}^{*+}$ tal que $\int f_o d\, P_1 > 0$. Pero por otro lado $\int f_{n_k} \, d\, P_1 < 0$ a partir de un cierto índice, con lo cual $\int f_o \, d\, P_1 \leq 0$. Contradicción que demuestra el teorema.

Corolario 2.4.16. Si se verifican las condiciones del teorema 2.4.19., entonces

$$\mu^{+}(S;\ K) = \lim_{n} \mu(S;\ K_n) = \bigcup_{n=1}^{\infty} \mu(S;\ K_n).$$

Contraejemplo 2.4.7. Sea $\Omega = \{\vartheta_1, \vartheta_2\}$; $S = \{(x_1, x_2);\ (x_1-1)^2 + (x_2-1)^2 \leq 1\}$. $K_n = H_n^1 \cap H_n^2$, donde $H_n^1 = \{(x_1,x_2);\ nx_1 + x_2 \leq 0$ y $H_n^2 = \{(x_1,x_2);\ x_1 + nx_2 \leq 0$. Obviamente $K_n \downarrow K = K_o = \{(x_1, x_2);\ x_1 \leq 0;\ x_2 \leq 0\}$. Sin embargo,

$$\bigcup_{n=1}^{\infty} \lambda(S;\ K_n) = \{(x_1,x_2);\ (x_1-1)^2 + (x_2-1)^2=1;\ x_1 < 1;\ x_2 < 1 ,$$

mientras que $\lambda(S;\ K_n) = \bigcup_{n=1}^{\infty} \lambda(S;\ K_n) \cup \{,0\} \cup \{0,1\}$.

Obsérvese que este contraejemplo sirve para demostrar igualmente que $\lim \mu(S;\ K_n) \neq \mu(S;\ K)$ puesto que, por ser S estrictamente K-convexo por abajo, se verifica que $\lambda(S;\ K_n) = \mu(S;\ K_n)$ y $\lambda(S;\ K) = \mu(S;\ K)$.

La restricción impuesta a Ω - la de ser finito puede levantarse, siendo el teorema cierto si se impone a los conos K_n la condición que los K_n sean conos con soporte compacto (para su definición véase p.e., Bourbaki, (1966), p. 114). Basta que lo sean a partir de uno de ellos. Esta nueva formulación del teorema siguiente incluye como caso particular al 2.4.19., puesto que si Ω es finito cualquier cono puntiagudo perteneciente a la clase $\mathcal{E}$ tiene soporte compacto.

<u>Teorema 2.4.20.</u> Sea $K_n \in \mathcal{C}$ una sucesión decreciente de conos puntiagudos y de soporte compacto tal que $K_n \downarrow K \in \mathcal{C}$. Entonces $K_n^{*+} \uparrow K^{*+}$.

<u>Demostración.</u> La primera parte de la demostración es idéntica a la del teorema anterior y por lo tanto

$$K^{*+} \supset \bigcup_{n=1}^{\infty} K_n^{*+}$$

Supongamos ahora que $P_o \in K^{*+}$ y que $P_o \notin \bigcup_{n=1}^{\infty} K_n^{*+}$. Esto implicaría la existencia de una sucesión $f_n \in K_n - \{0\}$ tal que $\int f_n \, d\, P_o \geq 0$ para todo n . Si designamos por C_n y C los soportes de K_n y K respectivamente, estos los podemos elegir de tal modo que $C_n \downarrow C$. Además siempre podemos elegir los $f_n \in C_n$ puesto que $C_n \cap \{0\} = \emptyset$. Entonces existe una sucesión $\{f_{n_k}\} \longrightarrow f_o$. Ahora bien por ser $f_{n_k} \in C_{n_k}$ se tiene que $f_o \in \bigcap_{n=1}^{\infty} C_n = C$ y por lo tanto $f_o \in K - \{0\}$ con lo cual se tendría $\int f_o \, d\, P_o < 0$. Por otro lado $\int f_{n_k} \, d\, P_o \geq 0$ para todo K. La aplicación inmediata del teorema de la convergencia dominada lleva a que $\int f_o dP_o \leq 0$, contradicción que demuestra el teorema.

REFERENCIAS BIBLIOGRAFICAS

ARDANUY,R. (1979). "El Soporte de una Distribución a Priori. Algunos Resultados y Aplicaciones". Actas de la XIª Reunión de Estadística, I.O. e Informatica. Sevilla.

BILLINGSLEY,P. (1968). "Convergence of Probability Measures". John Wiley, New York.

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

BLUM, J.R. and J. ROSENBLATT, (1967). "On Partial a Priori Information in Statistical Inference". Ann. Math. Statist. 38, 1671-1678.

BLYTH, C.R. (1951). "On Minimax Statistical Decision Procedures and their Admissibility". Ann. Math. Statist. 22, 22-42.

BOURBAKI, N. (1966). "Espaces Vectoriels Topologiques". Fascicule XV, Chap. 1 et 2. Hermann, Paris.

CHOQUET, G. (1960). "Le Theorème de Representation Integrale dans les Ensembles Convexes Compacts". Ann. Inst. Fourier, 10, 333-344.

CHOQUET, G. et P.A. MEYER,(1963). "Existence et Unicite des Representations Integrales dans les Convexes Compacts quelconques". Ann. Inst. Fourier, 13, 139-154.

DALAL, S.R. and J.H. GAINEFORD, (1980). "On Approximating Parametric Bayes Models by Nonparametric Bayes Models". Ann. Math. Statist. 8, 664-672.

DE ROBERTIS, L. and J.A. HARTIGAN, (1979). "Ranges of Prior Measures". Tech. Rep., Yale University.

DICKEY, J.M. (1976). "Approximate Posterior Distributions". J. Amer. Statist. Assoc. 71, 680-689.

EDWARDS, W., LINDMAN, H. and SAVAGE, L.J. (1963). "Bayesian Statistical Inference for Psychological Research". Psychological Rev. 70, 193-242.

FARREL,R.H. (1966). "Weak Limits of Sequences of Bayes Procedures in Estimation Theory". Proc. Fifth Berkeley Symp. Math. Statist. Prob., 1, 83-111.

FERGUSON, T.S. (1967). "Mathematical Statistics, a Decision Theoretic Approach". Academic Press, New York.

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

GIRON, F.J. (1975). "S-juegos Generalizados". Real Acad. Ciencias, Madrid. Tomo LXIX, cuad. 1º, 49-97.

GIRON, F.J. (1977). "Caracterización Axiomática de la Regla de Bayes y la Probabilidad Subjetiva". Real Acad. Ciencias, Madrid. Tomo LXXI, cuad. 1º, 19-101.

GIRON, F.J. y S. RIOS, (1981). "Quasi- Bayesian Behaviour: a more Realistic Approach to Decision Making? ". Proceedings of 1st Meeting on Bayesian Statistics, Valencia.

HEATH, D.C. and W.D. SUDDERTH, (1972). "On a Theorem of de Finetti oddsmaking and Game Theory". Ann. Math. Statist., 43, 2072-2077.

HEATH, D.C. and W.D. SUDDERTH, (1978). " On Finitely Additive Priors, Coherence, and Extended Admissibility". Ann. Statist., 6, 333-345.

HODGES, J.L. and E.L. LEHMANN, (1952). "The Use of Previous

Experience in Reaching Statistical Decisions". Ann. Math. Statist., 23, 396-467.

JACOBS, K. (1978). " Measure and Integral". Academic Press, New York.

JAGERS, P. (1974). "Aspects of Random Measures and Point Processes. Advances in Probability and Related Topics. P. Ney and S. Port, eds.. Vol 3, 179-240.

JAYNES, E.T. (1980). "Marginalization and Prior Probabilities" pp. 43-78. In "Bayesian Analysis in Econometrics and Statistics". Essays in Honor of Harold Jeffreys. A. Zellner, ed. North Holland.

KEYNES, J.M. (1921)."A Treatise on Probability". MacMillan, London.

KLEE, V.L. (1955). "Separation Properties of Convex Cones". Proc. Amer. Soc., 6, 313-318.

KOOPMAN, B.O. (1940). "The Axioms and Algebra of Intuitive Probability". Ann. Math., 41, 269-278.

KUDO, H. (1967). "On Partial Prior Information and the Property of Parametric Sufficiency". Proc. Fifth Berkeley Symp. Math. Statist. Prob. Vol. 1, 251-265.

MENGES, G. (1966). "On the "Bayesification" of the Minimax Principle". Unternehmensforschnn g. 10, Heff 2.

MENGES, G. (1968). "On Some Open Questions in Statistical Decision Theory". Pp. 140-162. In "Risk and Uncertainty". Karl Borch and Jan Mossin, eds.

PARTHASARATHY, K.R. (1967). " Probability Measures on Metric Spaces". Academic Press, New York.

RAIFFA, H. and R. SCHLAIFER, (1961). "Applied Statistical Decision Theory". H.U.P.,Cambridge.

RIOS, S. (1976a). "Analisis de Decisiones". Ediciones I.C.E., Madrid.

RIOS, S. (1976b). "Nuevos Criterios de Ordenación de Reglas de Decisión". Real Acad. Ciencias, Madrid. Tomo LXX, cuad. 2º, 235-253.

RUDIN, W. (1979). "Análisis Funcional".Reverté, Barcelona.

SACKS, J. (1963). "Generalized Bayes Solutions in Estimation Problems". Ann. Math. Statist., 34, 751-768.

SCHEFFE, H. (1947). "A Useful Convergence Theorem for Probability Distributions". Ann. Math. Statist., Vol. 18, 434-438.

SCHNEEWEISS, H. (1964). "Eine Entscheidnngsregel für den Fall Partiell Bekannter Wahrs cheinlichkeiten". Unternehmensforschung,8, 86-95.

SMITH, C.A.B. (1961). "Consistency in Statistical Inference and Decision" (with discussion). J. R. Statist. Soc. , 23, 1-25.

SUDDERTH, W.D. (1980). "Finitely Additive Priors, Coherence and the Marginalization Paradox". J. R. Statist. Soc., 42, 339-341.

VARADARAJAN,V.S. (1965). "Measures on Topological Spaces". Providence: Amer. Math. Soc. Translations, Series 2, 48 , 161-228.

YU, P.L. (1973). " Introduction to Domination Estructures in Multicriteria Decision Problems". In "Multiple Criteria Decision Making", J.L. Cochrane and M. Zeleny, eds. University of South Carolina Press, Columbia, pp. 249-261.

YU, P.L. (1974). "Cone-convexity, Cone Extreme, Points and Non dominated Solutions in Decision Problems with Multiobjectives". JOTA, Vol. 14, 319-337.