

UNIVERSIDAD DE OVIEDO

TRABAJO FIN DE GRADO DE MATEMÁTICAS

**Análisis de los modelos de
distorsión inducidos por la
distancia Euclídea y por la
divergencia de Kullback-Leibler**

Autor:
Arturo COLLAR BRAÑA

Tutor:
Ignacio MONTES

9 de junio de 2023



Índice general

Resumen	2
1. Preliminares	4
1.1. Notación	4
1.2. Probabilidades imprecisas	6
1.2.1. Probabilidades inferiores	6
1.2.2. Previsiones inferiores	8
1.2.3. Relación entre las previsiones inferiores y las probabi- lidades inferiores	10
1.3. Modelos de distorsión	12
2. Modelo de distorsión generado por la distancia Euclídea	16
2.1. Definición del modelo	16
2.2. Propiedades	28
2.3. Implementación	33
3. Modelo de distorsión generado por la divergencia de Kullback- Leibler	37
3.1. Definición del modelo	37
3.2. Propiedades	45
3.3. Implementación	46
4. Comparación de modelos de distorsión	49
4.1. Propiedades de las funciones de distorsión	49
4.2. Propiedades de las probabilidades inferiores coherentes aso- ciadas	50
4.3. Complejidad	51
5. Conclusiones	53
Bibliografía	55

Resumen

Este trabajo está incluido dentro del marco de la Teoría de la Probabilidad Imprecisa. Esta teoría nace como una alternativa para generalizar la Teoría de la Probabilidad Clásica en situaciones en las que la falta de suficiente información provoca que la probabilidad no quede perfectamente determinada.

A lo largo de este trabajo, nos centramos en unos modelos particulares dentro de las Probabilidades Imprecisas llamados modelos de distorsión [6, 7]. Estos modelos aparecen de forma natural cuando la distribución de probabilidad que modela la incertidumbre asociada a un experimento no está perfectamente determinada debido por ejemplo a falta de información, datos con excesivo ruido o expertos poco fiables. Para evitar que la falta de información afecte a las conclusiones o inferencias que se puedan realizar una posibilidad es “robustecer” la probabilidad dada tomando un entorno a su alrededor con un determinado radio que representará la cantidad de imprecisión que se añade. Por tanto, para determinar un modelo de distorsión son necesarios tres elementos: una probabilidad, un parámetro de distorsión y una función de distorsión que permita comparar probabilidades.

El objetivo de este trabajo es realizar un estudio complementario al realizado en [6, 7] mediante el estudio de los modelos de distorsión que se obtienen al tomar como función de distorsión la distancia Euclídea y la divergencia de Kullback-Leibler.

Este trabajo está dividido en cuatro capítulos. En el primer capítulo se realizará una introducción a la Teoría de las Probabilidades Imprecisas repasando sus principales conceptos. Además, en este primer capítulo también se estudiará qué es un modelo de distorsión, sus propiedades y algunos ejemplos de modelos de distorsión ya estudiados en otros trabajos.

En el segundo capítulo vamos a realizar el estudio del modelo de distorsión inducido por la distancia Euclídea. Definiremos este modelo, estudiaremos sus principales propiedades y finalizaremos el capítulo con ejemplos prácticos del cálculo de previsiones inferiores para este modelo y la representación gráfica de su conjunto credal.

En el tercer capítulo se realiza un estudio del modelo de distorsión inducido por la divergencia de Kullback-Leibler. Vamos a definir este modelo, estudiar sus principales propiedades y veremos algunos ejemplos para el

ÍNDICE GENERAL

cálculo de previsiones inferiores y previsiones superiores, también se realizará la representación de su conjunto credal. En este capítulo se deja abierto el estudio de algunas propiedades para este modelo debido a la complejidad de cálculo que implica trabajar con la divergencia de Kullback-Leibler.

Por último, en el cuarto capítulo de este trabajo vamos a realizar una comparación entre los diferentes modelos de distorsión, los ya estudiados en [6, 7] y los desarrollados a lo largo de este trabajo. Esta comparación se realizará en términos de las propiedades que verifican las funciones de distorsión, de las propiedades que verifica cada modelo de distorsión y en términos de la complejidad de cada modelo.

Capítulo 1

Preliminares

En este capítulo vamos a repasar los conceptos básicos de la Teoría de la Probabilidad, su notación y también daremos las condiciones generales sobre las que vamos a estar trabajando a lo largo de todo este documento.

Además, vamos a introducir todos los conceptos principales de la Teoría de las Probabilidades Imprecisas y sus notaciones, veremos qué son las probabilidades inferiores y las previsiones inferiores y la relación que mantienen entre sí estos dos conceptos.

Por último, veremos qué es un modelo de distorsión y algunas de las propiedades que estos modelos pueden cumplir. Para finalizar este capítulo veremos algunos ejemplos de modelos de distorsión conocidos como el modelo pari-mutuel [5, 8, 11], el modelo linear-vacuous [3, 11] o el modelo de la variación total [4, 9].

1.1. Notación

En esta sección vamos a introducir algunos conceptos básicos de la Teoría de la Probabilidad y veremos alguna notación que resulta muy común en el campo de las probabilidades imprecisas y que usaremos a lo largo de todo este trabajo.

En general, la estadística es la ciencia que se encarga del estudio de los experimentos aleatorios. Estos experimentos vienen modelados por los espacios de probabilidad. Un espacio de probabilidad es una terna (\mathcal{X}, σ, P) donde \mathcal{X} es un conjunto no vacío llamado espacio muestral, σ es una σ -álgebra y P es una medida de probabilidad, es decir, una función $P : \sigma \rightarrow [0, 1]$ verificando los axiomas de Kolmogorov que nos determina la verosimilitud de ocurrencia entre los diferentes sucesos de la σ -álgebra.

A lo largo de todo este documento estaremos trabajando sobre un espacio muestral finito $\mathcal{X} = \{x_1, \dots, x_n\}$ de cardinal $|\mathcal{X}| = n$, además vamos a considerar como σ -álgebra de sucesos asociada al experimento a las partes de \mathcal{X} , $\mathcal{P}(\mathcal{X})$. Usaremos la notación $A \subseteq \mathcal{X}$ para denotar un suceso A del

1.1. NOTACIÓN

espacio muestral \mathcal{X} .

Al trabajar sobre un espacio muestral finito la probabilidad queda determinada por su función de masa de probabilidad. Esta función asocia a cada elemento del espacio muestral \mathcal{X} la probabilidad de que se cumpla. En particular, se tiene que para nuestro espacio muestral $\mathcal{X} = \{x_1, \dots, x_n\}$ la probabilidad P vendrá dada por:

$$P(\{x_i\}) = p_i \quad \forall i = 1, \dots, n$$

y debe verificar que $p_i \geq 0$ para todo $i = 1, \dots, n$ y que $\sum_{i=1}^n p_i = 1$. Su notación vendrá dada por $P = (p_1, \dots, p_n)$. De esta forma, dado un suceso $A \subseteq \mathcal{X}$ con $A \neq \emptyset$ su probabilidad viene dada por:

$$P(A) = \sum_{x_i \in A} P(\{x_i\}) = \sum_{x_i \in A} p_i.$$

Cuando consideremos la probabilidad inicial P_0 , usaremos la notación $P_0 = (p_1^0, \dots, p_n^0)$ para denotar a su función de masa de probabilidad.

A lo largo de este trabajo, denotaremos por $\mathbb{P}(\mathcal{X})$ al conjunto de todas las probabilidades definidas sobre $\mathcal{P}(\mathcal{X})$, mientras que $\mathbb{P}^*(\mathcal{X})$ denotará al subconjunto de $\mathbb{P}(\mathcal{X})$ formado por aquellas probabilidades que aseguran probabilidad estrictamente positiva a todos aquellos sucesos no vacíos.

Consideramos a continuación el concepto de variable aleatoria o, siguiendo la terminología del probabilista italiano Bruno de Finetti [2], apuesta. Una apuesta, o variable aleatoria, es una función $f : \mathcal{X} \rightarrow \mathbb{R}$ que representa la ganancia obtenida $f(x)$ si se obtiene como resultado del experimento $x \in \mathcal{X}$. El conjunto de todas las posibles apuestas lo denotaremos por $\mathcal{L}(\mathcal{X})$.

En general, toda apuesta f se puede expresar como:

$$f(x) = \begin{cases} a_1 & \text{si } x = x_1 \\ a_2 & \text{si } x = x_2 \\ \dots & \\ a_n & \text{si } x = x_n \end{cases} = \sum_{i=1}^n a_i I_{\{x_i\}}(x). \quad (1.1)$$

Un caso particular de apuesta es la función indicadora sobre un subconjunto A de \mathcal{X} . Denotamos a esta apuesta como $f = I_A$ y vendrá dada por:

$$f(x) = I_A(x) = \begin{cases} 1 & \text{si } x \in A. \\ 0 & \text{si } x \notin A. \end{cases}$$

Definimos la esperanza de una apuesta f respecto a la probabilidad P como:

$$E_P(f) = \sum_{x_i \in \mathcal{X}} f(x_i) P(\{x_i\}) = \sum_{x_i \in \mathcal{X}} f(x_i) p_i.$$

Si la apuesta f se expresa como en la Ecuación (1.1), $f = \sum_{i=1}^n a_i I_{\{x_i\}}(x)$, su previsión o esperanza viene dada por:

$$E_P(f) = \sum_{i=1}^n a_i P(\{x_i\}) = \sum_{i=1}^n a_i p_i. \quad (1.2)$$

1.2. PROBABILIDADES IMPRECISAS

En particular, para la apuesta $f = I_A$ su esperanza viene dada por:

$$E_P(I_A) = \sum_{x_i \in \mathcal{X}} I_A(x_i)P(\{x_i\}) = \sum_{x_i \in A} P(\{x_i\}) = P(A).$$

Es decir, la probabilidad de un suceso es equivalente a la previsión o esperanza de su función indicadora. Este hecho tiene una consecuencia importante puesto que la probabilidad P y su previsión u operador esperanza E_P asociada contienen la misma información probabilística: a partir de P se puede calcular la esperanza o previsión de cualquier apuesta (incluida la función indicadora) y recíprocamente a partir de la previsión u operador esperanza se puede obtener la probabilidad de cualquier suceso.

Por este motivo, en el campo de las probabilidades imprecisas es habitual utilizar la notación $P(f)$, denominada como previsión de la apuesta f , en lugar de la habitual $E_P(f)$, entendiéndose que $P(A)$ es equivalente a $P(I_A)$ para cualquier subconjunto $A \subseteq \mathcal{X}$.

1.2. Probabilidades imprecisas

En esta sección vamos a explicar todos los conceptos básicos de las probabilidades y previsiones inferiores y la relación que existe entre estos dos conceptos.

1.2.1. Probabilidades inferiores

Definimos una probabilidad inferior como una función que asigna a cada suceso $A \subseteq \mathcal{X}$ un valor en el intervalo $[0, 1]$, es decir, $\underline{P} : \mathcal{P}(\mathcal{X}) \rightarrow [0, 1]$ satisfaciendo las siguientes propiedades:

1. $\underline{P}(\emptyset) = 0$ y $\underline{P}(\mathcal{X}) = 1$ (normalización).
2. Si $A \subseteq B$ entonces $\underline{P}(A) \leq \underline{P}(B)$ (monotonía).

Por otra parte, la probabilidad superior $\overline{P} : \mathcal{P}(\mathcal{X}) \rightarrow [0, 1]$ se define como la conjugada de la probabilidad inferior y viene dada por la siguiente relación:

$$\overline{P}(A) = 1 - \underline{P}(A^c) \quad \forall A \subseteq \mathcal{X}.$$

A partir de la definición de probabilidad superior e inferior podemos introducir el concepto de conjunto credal. Definimos el conjunto credal asociado a la probabilidad inferior \underline{P} como el conjunto de probabilidades cerrado y convexo que dominan a la probabilidad inferior o que son dominadas por la probabilidad superior, es decir:

$$\begin{aligned} \mathcal{M}(\underline{P}) &= \{P \in \mathbb{P}(\mathcal{X}) \mid P(A) \geq \underline{P}(A) \quad \forall A \subseteq \mathcal{X}\} \\ &= \{P \in \mathbb{P}(\mathcal{X}) \mid \overline{P}(A) \geq P(A) \quad \forall A \subseteq \mathcal{X}\}. \end{aligned}$$

1.2. PROBABILIDADES IMPRECISAS

Para poder trabajar en el campo de las probabilidades imprecisas, donde las probabilidades no se conocen con precisión, es necesario imponer ciertas condiciones de racionalidad que nos permitan trabajar de forma consistente.

La primera condición de consistencia es la propiedad de evitar la pérdida segura. Decimos que una probabilidad inferior \underline{P} evita la pérdida segura si su conjunto credal es no vacío, es decir, $\mathcal{M}(\underline{P}) \neq \emptyset$. Esta propiedad nos garantiza que al menos existe una probabilidad que es compatible con la información proporcionada por la probabilidad inferior.

Una segunda condición de racionalidad, más fuerte que la anterior, es la condición de coherencia. Decimos que una probabilidad inferior es coherente si para todo $A \subseteq \mathcal{X}$ existe $P \in \mathcal{M}(\underline{P})$ tal que $P(A) = \underline{P}(A)$. Es decir, \underline{P} es coherente cuando \underline{P} es la envolvente inferior de $\mathcal{M}(\underline{P})$.

Dada \underline{P} una probabilidad inferior coherente y \overline{P} su probabilidad superior conjugada, entonces para todo $A, B \subseteq \mathcal{X}$ se verifican las siguientes propiedades [11]:

1. $\underline{P}(A) = \min\{P(A) \mid P \in \mathcal{M}(\underline{P})\}$.
2. $\overline{P}(A) = \max\{P(A) \mid P \in \mathcal{M}(\underline{P})\}$.
3. Coherencia \Rightarrow Evitar pérdida segura.
4. $\underline{P}(A) \leq \overline{P}(A)$.
5. Si $A \cap B = \emptyset$ entonces $\underline{P}(A \cup B) \geq \underline{P}(A) + \underline{P}(B)$.
6. $\overline{P}(A \cup B) \leq \overline{P}(A) + \overline{P}(B)$.

Las probabilidades inferiores pueden tener diferentes interpretaciones, una de ellas es la interpretación epistémica de la probabilidad. En esta interpretación tenemos un experimento donde la incertidumbre es medida por una probabilidad P_0 , el problema es que no disponemos de suficiente información para garantizar el valor exacto de todos los valores de la probabilidad P_0 .

Para resolver este problema tomamos la probabilidad inferior y superior como cotas de la probabilidad P_0 , es decir:

$$\underline{P}(A) \leq P_0(A) \leq \overline{P}(A) \quad \forall A \subseteq \mathcal{X}.$$

De esta forma, desconocemos cómo es exactamente la probabilidad P_0 pero podemos garantizar que pertenece al conjunto credal $\mathcal{M}(\underline{P})$.

Otro tipo de interpretación es la comportamental de Bruno de Finetti [2]. Esta interpretación puede verse como un sistema de apuestas clásico donde vamos a estar apostando dinero (comprando y vendiendo apuestas) sobre un suceso A .

La interpretación de de Finetti marca un precio justo para la apuesta, es decir, un precio donde apostar no nos proporcionará ni beneficios ni pérdidas,

1.2. PROBABILIDADES IMPRECISAS

y a partir de este precio vamos a definir unas cotas a partir de las cuáles nos interesará comprar o vender la apuesta sobre el suceso A .

En este contexto, definimos la probabilidad inferior $\underline{P}(A)$ como el supremo del precio de compra aceptable para el suceso A . Por otro lado, definimos la probabilidad superior $\overline{P}(A)$ como el ínfimo del precio de venta aceptable para el suceso A .

Esto quiere decir que por debajo de la probabilidad inferior deberíamos comprar esa apuesta y por encima de la probabilidad superior deberíamos vender esa apuesta, pero no disponemos de suficiente información para decidir cuando el precio de compra/venta de la apuesta está en el intervalo $[\underline{P}(A), \overline{P}(A)]$.

1.2.2. Previsiones inferiores

La idea de previsión inferior es similar a la de probabilidad inferior, pero considerando un dominio más amplio: el conjunto de todas las apuestas $\mathcal{L}(\mathcal{X})$ en lugar de $\mathcal{P}(\mathcal{X})$.

Una previsión inferior es una función $\underline{P} : \mathcal{L}(\mathcal{X}) \rightarrow \mathbb{R}$ que asigna a cada apuesta $f \in \mathcal{L}(\mathcal{X})$ un valor numérico en los números reales.

Por otra parte, la previsión superior \overline{P} la definimos como el opuesto de la previsión inferior de la apuesta $-f$, es decir:

$$\overline{P}(f) = -\underline{P}(-f) \quad \forall f \in \mathcal{L}(\mathcal{X}). \quad (1.3)$$

El conjunto credal asociado a la previsión inferior \underline{P} viene definido de igual forma que para probabilidades inferiores, es decir, como el conjunto de probabilidades que dominan a la previsión inferior para todas las apuestas:

$$\begin{aligned} \mathcal{M}(\underline{P}) &= \{P \in \mathbb{P}(\mathcal{X}) \mid P(f) \geq \underline{P}(f) \quad \forall f \in \mathcal{L}(\mathcal{X})\} \\ &= \{P \in \mathbb{P}(\mathcal{X}) \mid \overline{P}(f) \geq P(f) \quad \forall f \in \mathcal{L}(\mathcal{X})\}. \end{aligned} \quad (1.4)$$

De igual forma que cuando trabajamos con probabilidades superiores e inferiores, es necesario imponer ciertas condiciones de consistencia. Estas propiedades son las mismas que para probabilidades inferiores, la propiedad de evitar la pérdida segura y la de coherencia.

Diremos que \underline{P} evita la pérdida segura cuando $\mathcal{M}(\underline{P}) \neq \emptyset$ y que \underline{P} es coherente si para cada $f \in \mathcal{L}(\mathcal{X})$ existe un $P \in \mathcal{M}(\underline{P})$ tal que $P(f) = \underline{P}(f)$.

Para previsiones inferiores, la condición de coherencia es equivalente a comprobar que para todo $f, g \in \mathcal{M}(\underline{P})$ y $\lambda > 0$ se tiene que:

1. $\underline{P}(f) \geq \min_{x \in \mathcal{X}} f(x)$.
2. $\underline{P}(\lambda f) = \lambda \underline{P}(f)$.
3. $\underline{P}(f + g) \geq \underline{P}(f) + \underline{P}(g)$.

1.2. PROBABILIDADES IMPRECISAS

Al igual que para probabilidades inferiores hablaremos de dos tipos de interpretaciones para la previsión inferior: la interpretación epistémica y la comportamental.

La interpretación epistémica para previsiones inferiores se basa en que no conocemos con exactitud la incertidumbre del experimento que medimos mediante la probabilidad P_0 . Entonces, vamos a utilizar toda la información disponible para tomar como cotas la previsión superior e inferior no solamente de sucesos, sino de apuestas:

$$\underline{P}(f) \leq P_0(f) \leq \overline{P}(f) \quad \forall f \in \mathcal{L}(\mathcal{X}).$$

Al igual que con probabilidades inferiores, desconocemos quién es exactamente P_0 pero podemos garantizar que pertenece al conjunto credal.

Por otra parte, la interpretación comportamental es muy similar al caso con probabilidades. La previsión de una apuesta f se puede entender como el precio de compra (o de venta) justo para esa apuesta y $f(x)$ nos proporciona la ganancia (o pérdida) obtenida en la apuesta f si el resultado del experimento ha sido $x \in \mathcal{X}$.

El beneficio de cada apuesta es sencillo de calcular pues sólo es necesario considerar la ganancia obtenida a la que hay que restarle el precio que se ha pagado por esa apuesta.

Hablaremos de precio justo si el beneficio obtenido por comprar (o vender) esa apuesta f es cero, es decir, ni perdemos ni ganamos.

En este caso, $\underline{P}(f)$ representa el supremo del precio de compra para la apuesta f , y $\overline{P}(f)$ representa el ínfimo del precio de venta para esa misma apuesta. La interpretación en este caso es la misma que en el caso de probabilidades, por encima de $\overline{P}(f)$ vendemos la apuesta y por debajo de $\underline{P}(f)$ la compramos.

Anteriormente hemos definido el conjunto credal como un conjunto de probabilidades cerrado y convexo, por este motivo este conjunto viene caracterizado por sus puntos extremos.

Diremos que P es un punto extremo del conjunto credal $\mathcal{M}(\underline{P})$ si verifica que si $P = \alpha P_1 + (1 - \alpha)P_2$ para algún $P_1, P_2 \in \mathcal{M}(\underline{P})$ y $\alpha \in [0, 1]$ entonces resulta que $P = P_1 = P_2$. Denotaremos al conjunto de todos los puntos extremos asociados al conjunto credal $\mathcal{M}(\underline{P})$ como $ext(\mathcal{M}(\underline{P}))$.

Si $P \in ext(\mathcal{M}(\underline{P}))$ entonces existe una apuesta $f \in \mathcal{L}(\mathcal{X})$ de forma que P es la única probabilidad en $\mathcal{M}(\underline{P})$ verificando que $\underline{P}(f) = P(f)$.

Por tanto, si \underline{P} es una previsión inferior coherente, se cumple que para todo $f \in \mathcal{L}(\mathcal{X})$:

$$\begin{aligned} \underline{P}(f) &= \min \{P(f) \mid P \in \mathcal{M}(\underline{P})\} = \min \{P(f) \mid P \in ext(\mathcal{M}(\underline{P}))\}. \\ \overline{P}(f) &= \max \{P(f) \mid P \in \mathcal{M}(\underline{P})\} = \max \{P(f) \mid P \in ext(\mathcal{M}(\underline{P}))\}. \end{aligned} \quad (1.5)$$

1.2.3. Relación entre las previsiones inferiores y las probabilidades inferiores

Hemos explicado anteriormente la relación que existe entre previsiones y probabilidades, concretamente vimos que toda probabilidad de un suceso A puede ser vista como la previsión de una apuesta dada por la función indicadora sobre A , es decir, $P(I_A) = P(A)$, y que por tanto una probabilidad y su previsión asociada son equivalentes.

Por otra parte, los conjuntos credales y las previsiones inferiores coherentes resultan conceptos equivalentes. Si tenemos información sobre una previsión inferior coherente es sencillo obtener su conjunto credal que viene dado por:

$$\mathcal{M}(\underline{P}) = \{P \in \mathbb{P}(\mathcal{X}) \mid P(f) \geq \underline{P}(f) \quad \forall f \in \mathcal{L}(\mathcal{X})\}.$$

Por otra parte, si tenemos un conjunto credal definido podemos expresar la previsión inferior coherente asociada como:

$$\underline{P}(f) = \min\{P(f) \mid P \in \mathcal{M}(\underline{P})\}.$$

Por lo tanto se trata de la misma información probabilística expresada de forma diferente.

Pero no ocurre lo mismo con la probabilidad inferior coherente ya que diferentes conjuntos credales, y por tanto diferentes previsiones inferiores coherentes, pueden tener la misma restricción a sucesos. Veamos un ejemplo:

Vamos a tomar dos conjuntos credales diferentes de tal forma que su probabilidad inferior asociada sea la misma y veremos cómo la previsión inferior para una misma apuesta es diferente.

Sea el espacio muestral $\mathcal{X} = \{x_1, x_2, x_3\}$ y consideremos las previsiones inferiores coherentes \underline{P} y \underline{Q} cuyos conjuntos credales $\mathcal{M}(\underline{P})$ y $\mathcal{M}(\underline{Q})$ tienen como extremos:

$$\text{ext}(\mathcal{M}(\underline{P})) = \{P_1, P_2, P_3\} = \{(0.1, 0.3, 0.6), (0.7, 0.1, 0.2), (0.5, 0.4, 0.1)\}.$$

$$\text{ext}(\mathcal{M}(\underline{Q})) = \{Q_1, Q_2, Q_3\} = \{(0.7, 0.2, 0.1), (0.3, 0.1, 0.6), (0.1, 0.4, 0.5)\}.$$

Teniendo en cuenta que la probabilidad inferior sobre un suceso A viene dada por $\underline{P}(A) = \min\{P(A) \mid P \in \mathcal{M}(\underline{P})\}$ y por $\underline{Q}(A) = \min\{Q(A) \mid Q \in \mathcal{M}(\underline{Q})\}$ se obtiene la siguiente tabla con la probabilidad inferior para \underline{P} y \underline{Q} .

	$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$
\underline{P}	0.1	0.1	0.1	0.4	0.6	0.3
\underline{Q}	0.1	0.1	0.1	0.4	0.6	0.3

Por tanto, la restricción a sucesos de \underline{P} y \underline{Q} coinciden, pero \underline{P} y \underline{Q} son distintos. Por ejemplo, sea f una apuesta dada por $f = 10I_{\{x_1\}} + 20I_{\{x_2\}} + 30I_{\{x_3\}}$. Esta apuesta es la que recompensa al jugador con 10 si el resultado

1.2. PROBABILIDADES IMPRECISAS

del experimento es x_1 , con 20 si el resultado es x_2 y con 30 si el resultado es x_3 .

Usando la Ecuación (1.5) y la Ecuación (1.2) para calcular las previsiones inferiores se tiene que:

$$\begin{aligned}\underline{P}(f) &= \min \{P_1(f), P_2(f), P_3(f)\} = \min \{25, 15, 16\} = 15. \\ \underline{Q}(f) &= \min \{Q_1(f), Q_2(f), Q_3(f)\} = \min \{14, 23, 24\} = 14.\end{aligned}$$

Acabamos de ver, que aunque \underline{P} y \underline{Q} tengan la misma restricción a sucesos su previsión inferior sobre una misma apuesta es diferente, es decir, dos previsiones inferiores diferentes pueden tener las misma restricción a sucesos.

Por este motivo vamos a trabajar a lo largo de todo este trabajo con previsiones inferiores ya que resultan más informativas que las probabilidades inferiores.

A continuación, vamos a estudiar una propiedad muy importante que cumplen algunas previsiones inferiores, se trata de la 2-monotonía. Diremos que \underline{P} es 2-monótona si verifica que:

$$\underline{P}(f \vee g) + \underline{P}(f \wedge g) \geq \underline{P}(f) + \underline{P}(g) \quad \forall f, g \in \mathcal{L}(\mathcal{X})$$

donde \vee indica el mínimo y \wedge el máximo punto a punto.

Si consideramos su restricción a sucesos, entonces decimos que \underline{P} es 2-monótona si:

$$\underline{P}(A \cap B) + \underline{P}(A \cup B) \geq \underline{P}(A) + \underline{P}(B) \quad \forall A, B \subseteq \mathcal{X}.$$

La importancia de la 2-monotonía en esta teoría es que nos permite relacionar las probabilidades inferiores con las previsiones inferiores ya que una previsión inferior 2-monótona y su probabilidad inferior 2-monótona asociada resultan equivalentes, es decir, expresan la misma información probabilística. Además, si una previsión o una probabilidad inferior es 2-monótona entonces también se tiene que cumple la propiedad de coherencia.

Ya hemos visto anteriormente cómo se puede obtener la probabilidad de un suceso A a partir de una previsión inferior, basta considerar la apuesta $f = I_A$, pero si disponemos de la propiedad de 2-monotonía es posible calcular la previsión inferior a partir de la probabilidad inferior.

Si $\underline{P} : \mathcal{L}(\mathcal{X}) \rightarrow \mathbb{R}$ es una previsión inferior 2-monótona y $\underline{P}' : \mathcal{P}(\mathcal{X}) \rightarrow [0, 1]$ denota su probabilidad inferior coherente asociada obtenida como la restricción a sucesos:

$$\underline{P}'(A) = \underline{P}(I_A) \quad \forall A \subseteq \mathcal{X},$$

entonces $\underline{P}(f)$ se puede calcular usando \underline{P}' como:

$$\underline{P}(f) = (C) \int f d\underline{P}' \tag{1.6}$$

1.3. MODELOS DE DISTORSIÓN

donde $(C) \int$ representa un operador llamado integral de Choquet [1] de la apuesta f respecto de la probabilidad inferior \underline{P} y viene dado por:

$$\underline{P}(f) = (C) \int f d\underline{P} = \min_{x \in \mathcal{X}} f(x) + \int_{\min f}^{\max f} \underline{P}(\{x \mid f(x) \geq \omega\}) d\omega.$$

La 2-monotonía de la probabilidad inferior coherente nos garantiza que la probabilidad inferior tiene una única extensión a apuestas en forma de previsión inferior 2-monótona. Además, otra ventaja de trabajar con probabilidades inferiores 2-monótonas es que el cálculo de los puntos extremos del conjunto credal es mucho más simple. Estos puntos extremos vienen dados por $\{P_\sigma \mid \sigma \in S_n\}$ donde S_n representa el conjunto de permutaciones de $\{1, \dots, n\}$ y $x_{\sigma(i)}$ representa la i -ésima permutación de los elementos del espacio muestral. De esta forma los puntos extremos P_σ vienen dados por:

$$\begin{aligned} P_\sigma(\{x_{\sigma(1)}\}) &= \underline{P}(\{x_{\sigma(1)}\}). \\ P_\sigma(\{x_{\sigma(2)}\}) &= \underline{P}(\{x_{\sigma(1)}, x_{\sigma(2)}\}) - \underline{P}(\{x_{\sigma(1)}\}). \\ &\dots \\ P_\sigma(\{x_{\sigma(i)}\}) &= \underline{P}(\{x_{\sigma(1)}, \dots, x_{\sigma(i)}\}) - \underline{P}(\{x_{\sigma(1)}, \dots, x_{\sigma(i-1)}\}). \\ &\dots \\ P_\sigma(\{x_{\sigma(n)}\}) &= \underline{P}(\{x_{\sigma(1)}, \dots, x_{\sigma(n)}\}) - \underline{P}(\{x_{\sigma(1)}, \dots, x_{\sigma(n-1)}\}) \\ &= 1 - \underline{P}(\{x_{\sigma(1)}, \dots, x_{\sigma(n-1)}\}). \end{aligned}$$

Es por ello que si \underline{P} es 2-monótona entonces $\mathcal{M}(\underline{P})$ tiene como mucho $n!$ puntos extremos [1].

1.3. Modelos de distorsión

A lo largo de este trabajo vamos a estudiar un tipo particular de previsiones inferiores que son las que se obtienen mediante los modelos de distorsión.

Los modelos de distorsión se basan en la idea de tomar una probabilidad inicial $P_0 \in \mathbb{P}(\mathcal{X})$ sobre la que se va a definir un entorno mediante una función de distorsión $d : \mathbb{P}(\mathcal{X}) \times \mathbb{P}(\mathcal{X}) \rightarrow [0, \infty)$ que sirve para comparar probabilidades, y un parámetro de distorsión $\delta > 0$ que indica la cantidad de imprecisión que se añade al modelo.

De esta forma se define el modelo de distorsión dado por P_0 , d y δ como la bola cerrada dada por:

$$B_d^\delta(P_0) = \{P \in \mathbb{P}(\mathcal{X}) \mid d(P, P_0) \leq \delta\}. \quad (1.7)$$

Este conjunto contiene a todas las probabilidades que distan de la probabilidad inicial P_0 como mucho δ . Las funciones de distorsión pueden satisfacer las siguientes propiedades:

1.3. MODELOS DE DISTORSIÓN

- Axioma 1 (Definida positiva): $d(P_1, P_2)=0$ si y sólo si $P_1 = P_2$
 $\forall P_1, P_2 \in \mathbb{P}(\mathcal{X})$.
- Axioma 2 (Simetría): $d(P_1, P_2) = d(P_2, P_1) \quad \forall P_1, P_2 \in \mathbb{P}(\mathcal{X})$.
- Axioma 3 (Desigualdad triangular): $d(P_1, P_3) \leq d(P_1, P_2) + d(P_2, P_3)$
 $\forall P_1, P_2, P_3 \in \mathbb{P}(\mathcal{X})$.
- Axioma 4 (Convexidad):
 $d(\alpha P_1 + (1 - \alpha)P_2, P_3) \leq \max\{d(P_1, P_3), d(P_2, P_3)\} \quad \forall \alpha \in [0, 1]$ y
 $\forall P_1, P_2, P_3 \in \mathbb{P}(\mathcal{X})$.
- Axioma 5 (Continuidad): $\forall P, P_1, P_2 \in \mathbb{P}(\mathcal{X})$ y $\forall \epsilon > 0, \exists \delta > 0$ tal que
si $\|P_1 - P_2\| < \delta$ entonces se tiene que $|d(P_1, P) - d(P_2, P)| < \epsilon$.

La función de distorsión d es una distancia cuando satisface los axiomas 1, 2 y 3. Además, si verifica los axiomas 4 y 5, es decir, la función de distorsión es continua y convexa, entonces se tiene que la bola definida en la Ecuación (1.7) es cerrada y convexa, es decir, se trata de un conjunto credal y por tanto coincide con $\mathcal{M}(\underline{P}_d)$ [6, Prop.3.1], donde \underline{P}_d es la previsión inferior coherente definida como la envolvente inferior de $B_d^\delta(P_0)$ dada por:

$$\begin{aligned} \underline{P}_d(f) &= \inf \{P(f) \mid P \in \mathcal{M}(\underline{P}_d)\} = \inf \{P(f) \mid P \in B_d^\delta(P_0)\} \\ &= \inf \{P(f) \mid d(P, P_0) \leq \delta\} \quad \forall f \in \mathcal{L}(\mathcal{X}). \end{aligned}$$

Como ya hemos explicado, el objetivo de este trabajo es estudiar diferentes modelos de distorsión para complementar los ya estudiados en [6, 7] y poder hacer una comparación de todos ellos.

En [6, 7] se recoge el estudio de los modelos de distorsión desde un punto de vista general y teórico. Además se particulariza este estudio para algunos modelos concretos de especial interés por sus propiedades como son el modelo pari-mutuel (*PMM*), el modelo linear-vacuous (*LV*) y el modelo total-variation (*TV*).

A lo largo de estos dos estudios se trabaja bajo la hipótesis de que la probabilidad inicial P_0 pertenece a $\mathbb{P}^*(\mathcal{X})$ y que el parámetro de distorsión δ es lo suficientemente pequeño para garantizar que $B_d^\delta(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$. Para mantener una consistencia con estos resultados ya estudiados trabajaremos a lo largo de este trabajo en nuevos resultados bajo estas mismas hipótesis.

Consideremos primero el modelo linear-vacuous, también llamado modelo de ϵ -contaminación [3, 11].

La función de distorsión que genera este modelo es $d_{LV} : \mathbb{P}^*(\mathcal{X}) \times \mathbb{P}^*(\mathcal{X}) \rightarrow [0, \infty)$ y viene dada por:

$$d_{LV}(P, Q) = \max_{A \neq \emptyset} \frac{Q(A) - P(A)}{Q(A)}.$$

1.3. MODELOS DE DISTORSIÓN

Dada la función de distorsión d_{LV} , una probabilidad inicial P_0 y un parámetro de distorsión $\delta > 0$ podemos considerar el siguiente conjunto credal:

$$B_{d_{LV}}^\delta(P_0) = \{P \in \mathbb{P}(\mathcal{X}) \mid d_{LV}(P, P_0) \leq \delta\}.$$

La probabilidad inferior y superior asociada a este modelo vienen dadas por:

$$\begin{aligned} \underline{P}_{LV}(A) &= \begin{cases} (1 - \delta)P_0(A) & \text{si } A \neq \mathcal{X}. \\ 1 & \text{si } A = \mathcal{X}. \end{cases} \\ \overline{P}_{LV}(A) &= \begin{cases} (1 - \delta)P_0(A) + \delta & \text{si } A \neq \emptyset. \\ 0 & \text{si } A = \emptyset. \end{cases} \end{aligned}$$

El modelo linear-vacuous es 2-monótono. Por tanto, su extensión a apuestas es única y viene dada por la integral de Choquet (1.6).

Consideremos ahora el modelo pari-mutuel. Este modelo está generado por la función de distorsión $d_{PMM} : \mathbb{P}^*(\mathcal{X}) \times \mathbb{P}^*(\mathcal{X}) \rightarrow [0, \infty)$ dada por:

$$d_{PMM}(P, Q) = \max_{A \subseteq \mathcal{X}} \frac{Q(A) - P(A)}{1 - Q(A)}.$$

Dado un parámetro de distorsión $\delta > 0$ y una probabilidad inicial P_0 , el conjunto credal asociado a este modelo viene dado por:

$$B_{d_{PMM}}^\delta(P_0) = \{P \in \mathbb{P}(\mathcal{X}) \mid d_{PMM}(P, P_0) \leq \delta\}.$$

Las probabilidades coherentes inferiores y superiores asociadas a este modelo vienen dadas para cada $A \subseteq \mathcal{X}$ por:

$$\begin{aligned} \underline{P}_{PMM}(A) &= \max\{0, (1 + \delta)P_0(A) - \delta\}. \\ \overline{P}_{PMM}(A) &= \min\{0, (1 + \delta)P_0(A)\}. \end{aligned}$$

Este modelo nace como un sistema para realizar apuestas donde δ representa el factor de ganancia de la casa de apuestas. Para más detalle sobre este modelo nos referimos a [8, 11].

Se tiene que la probabilidad inferior asociada \underline{P}_{PMM} es 2-monótona, y por tanto su extensión a apuestas, que también es 2 monótona, es única y viene dada por la integral de Choquet (1.6).

Por último, vamos a ver el modelo basado en la distancia de variación total [4, 9]. Dadas dos probabilidades $P, Q \in \mathbb{P}(\mathcal{X})$ la distancia de la variación total viene dada por:

$$d_{TV}(P, Q) = \max_{A \subseteq \mathcal{X}} |P(A) - Q(A)|.$$

Sea un parámetro de distorsión $\delta > 0$ y una probabilidad inicial P_0 , el conjunto credal asociado a este modelo viene dado por:

$$B_{d_{TV}}^\delta(P_0) = \{P \in \mathbb{P}(\mathcal{X}) \mid d_{TV}(P, P_0) \leq \delta\}.$$

1.3. MODELOS DE DISTORSIÓN

Las probabilidades inferiores y superiores para este modelo vienen dadas por:

$$\begin{aligned} \underline{P}_{TV}(A) &= \begin{cases} \text{máx}\{0, P_0(A) - \delta\} & \text{si } A \neq \mathcal{X}. \\ 1 & \text{si } A = \mathcal{X}. \end{cases} \\ \overline{P}_{TV}(A) &= \begin{cases} \text{mín}\{1, P_0(A) + \delta\} & \text{si } A \neq \emptyset. \\ 0 & \text{si } A = \emptyset. \end{cases} \end{aligned}$$

El modelo de variación total es 2-monótono [7] y por tanto tiene una única extensión a apuestas y viene dada en términos de la integral de Choquet (1.6).

Capítulo 2

Modelo de distorsión generado por la distancia Euclídea

A lo largo de este capítulo vamos a definir y estudiar las diferentes propiedades del modelo de distorsión generado por la distancia Euclídea. Además, definiremos las previsiones y probabilidades inferiores y superiores para este modelo.

Para finalizar este capítulo veremos un ejemplo de cálculo de las previsiones inferiores para este modelo y la representación gráfica de su conjunto credal.

2.1. Definición del modelo

En esta sección vamos a definir el modelo de distorsión generado por la distancia Euclídea. Para dos probabilidades $P, Q \in \mathbb{P}(\mathcal{X})$ la distancia Euclídea viene dada por

$$d_E(P, Q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}.$$

En este estudio vamos a considerar una probabilidad $P_0 \in \mathbb{P}^*(\mathcal{X})$ y un parámetro de distorsión $\delta \in (0, 1)$. Con estos elementos definimos la bola cerrada centrada en P_0 con parámetro de distorsión δ respecto a la distancia Euclídea que viene dada por:

$$\begin{aligned} B_E^\delta(P_0) &= \{P \in \mathbb{P}(\mathcal{X}) \mid d_E(P, P_0) \leq \delta\} \\ &= \left\{ P \in \mathbb{P}(\mathcal{X}) \mid \sqrt{\sum_{i=1}^n (P(x_i) - P_0(x_i))^2} < \delta \right\}. \end{aligned}$$

2.1. DEFINICIÓN DEL MODELO

Como ya hemos introducido en la Sección 1.3, a lo largo de este estudio vamos a trabajar bajo la hipótesis de que el parámetro de distorsión delta elegido es suficientemente pequeño para garantizar que $B_d^\delta(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$, de esta forma, seguimos trabajando bajo las hipótesis generales que se usan en los estudios [6, 7].

En la Sección 1.3 se definen cinco axiomas que una función de distorsión puede cumplir, vamos a comprobar los que verifica la distancia Euclídea.

Proposición 2.1. *La distancia Euclídea d_E definida satisface los axiomas: 1, 2, 3, 4 y 5.*

Demostración. Como d_E es una distancia es simétrica, definida positiva y cumple la desigualdad triangular por lo que cumple los axiomas 1, 2 y 3. Veamos que verifica el axioma 4. Sean $P, Q, H \in \mathbb{P}(\mathcal{X})$ y $\alpha \in [0, 1]$ arbitrarios tales que $p_i = P(\{x_i\})$, $q_i = Q(\{x_i\})$ y $h_i = H(\{x_i\})$ para cada $i = 1, \dots, n$. Entonces:

$$\begin{aligned} d_E(\alpha P + (1 - \alpha)Q, H) &= \sqrt{\sum_{i=1}^n ((\alpha P(\{x_i\}) + (1 - \alpha)Q(\{x_i\})) - H(\{x_i\}))^2} \\ &= \sqrt{\sum_{i=1}^n ((\alpha p_i + (1 - \alpha)q_i) - h_i)^2}. \end{aligned}$$

Sumando y restando αh_i se tiene que:

$$\begin{aligned} d_E(\alpha P + (1 - \alpha)Q, H) &= \sqrt{\sum_{i=1}^n ((\alpha p_i + (1 - \alpha)q_i) - h_i)^2} \\ &= \sqrt{\sum_{i=1}^n ((\alpha p_i + (1 - \alpha)q_i) - h_i + \alpha h_i - \alpha h_i)^2} \\ &= \sqrt{\sum_{i=1}^n (\alpha p_i - \alpha h_i + (1 - \alpha)q_i - (1 - \alpha)h_i)^2} \\ &= \sqrt{\sum_{i=1}^n (\alpha(p_i - h_i) + (1 - \alpha)(q_i - h_i))^2}. \end{aligned}$$

Aplicando la desigualdad de Minkowsky que dice:

$$\sqrt{\sum_{i=1}^n (a_i + b_i)^2} \leq \sqrt{\sum_{i=1}^n a_i^2} + \sqrt{\sum_{i=1}^n b_i^2}$$

2.1. DEFINICIÓN DEL MODELO

se obtiene lo siguiente:

$$\begin{aligned}
 d_E(\alpha P + (1 - \alpha)Q, H) &= \sqrt{\sum_{i=1}^n (\alpha(p_i - h_i) + (1 - \alpha)(q_i - h_i))^2} \\
 &\leq \sqrt{\sum_{i=1}^n (\alpha(p_i - h_i))^2} + \sqrt{\sum_{i=1}^n ((1 - \alpha)(q_i - h_i))^2} \\
 &= \alpha d_E(P, H) + (1 - \alpha)d_E(Q, H) \\
 &\leq \max\{d(P, H), d(Q, H)\}.
 \end{aligned}$$

Por lo que queda probado que la distancia Euclídea verifica el axioma de convexidad.

El axioma de continuidad se cumple porque la distancia Euclídea es una composición de funciones continuas, y por tanto, se trata de una función continua. \square

De acuerdo con [6, Prop.3.1], si la función de distorsión $d : \mathbb{P}(\mathcal{X}) \times \mathbb{P}(\mathcal{X}) \rightarrow \mathbb{R}^+$ es continua y convexa, entonces la bola $B_E^\delta(P_0)$ es cerrada y convexa, es decir, se trata de un conjunto credal y por tanto coincide con $\mathcal{M}(\underline{P}_E)$, donde \underline{P}_E es la previsión inferior coherente definida como la envolvente inferior de $B_E^\delta(P_0)$:

$$\underline{P}_E(f) = \min\{P(f) \mid P \in B_E^\delta(P_0)\} = \min\{P(f) \mid P \in d_E(P, P_0) \leq \delta\}$$

para cada apuesta $f \in \mathcal{L}(\mathcal{X})$. De esta forma se tiene que el conjunto credal viene dado por:

$$\mathcal{M}(\underline{P}_E) = \{P \in \mathbb{P}(\mathcal{X}) \mid P(f) \geq \underline{P}_E(f) \quad \forall f \in \mathcal{L}(\mathcal{X})\} = B_E^\delta(P_0).$$

A continuación, vamos a ver que los puntos extremos del conjunto credal $\mathcal{M}(\underline{P}_E)$ pertenecen a la frontera de la bola $B_E^\delta(P_0)$.

Proposición 2.2. *Sea $\delta > 0$ y $P_0 \in \mathbb{P}^*(\mathcal{X})$ de forma que $B_E^\delta(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$, y sea f una apuesta no constante. Entonces, si $\underline{P}_E(f) = P(f)$ para algún $P \in B_E^\delta(P_0)$ se tiene que $d_E(P, P_0) = \delta$.*

Demostración. Por reducción al absurdo, supongamos que $d_E(P, P_0) < \delta$ y veamos que existe $P^* \in B_E^\delta(P_0)$ tal que $P^*(f) < P(f)$ lo que implicaría la siguiente contradicción:

$$P^*(f) < P(f) = \underline{P}_E(f) = \min\{P(f) \mid P \in B_E^\delta(P_0)\}.$$

Sea la probabilidad $P = (p_1, \dots, p_n)$, y consideremos la apuesta f dada por $f(x) = \sum_{i=1}^n a_i I_{x_i}(x)$, cuya previsión es $P(f) = \sum_{i=1}^n a_i p_i$.

Sea $a_j = \max_{1 \leq i \leq n} a_i$ y sea $a_k = \min_{1 \leq i \leq n} a_i$, de tal forma que podemos asegurar que $a_k < a_j$ ya que la apuesta f es no constante por hipótesis.

2.1. DEFINICIÓN DEL MODELO

Sea $\epsilon > 0$ verificando que:

$$\epsilon < \min \left\{ p_j, 1 - p_k, \frac{1}{6} (\delta^2 - d_E(P, P_0)^2) \right\}.$$

Se tiene que $p_j > 0$ y $1 - p_k > 0$ pues por hipótesis $B_E^\delta(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$. Además $\frac{1}{6} (\delta^2 - d_E(P, P_0)^2) > 0$ ya que por hipótesis $d_E(P, P_0) < \delta$.

Definimos la probabilidad $P^* = (p_1^*, \dots, p_n^*)$ donde p_i^* vienen dados por:

$$p_i^* = P^*({x_i}) = \begin{cases} p_j - \epsilon & \text{si } i = j. \\ p_k + \epsilon & \text{si } i = k. \\ p_i & \text{si } i \neq j, k. \end{cases}$$

Veamos que P^* es una probabilidad, es decir, que p_1^*, \dots, p_n^* son no-negativos (de hecho, por construcción son estrictamente positivos) y su suma es 1:

Está claro que para todo $i \neq j, k$ $p_i^* = p_i > 0$ y que $p_k + \epsilon < 1$, puesto que $\epsilon < 1 - p_k$. Además, $p_j - \epsilon > 0$ ya que $\epsilon < p_j$ por hipótesis.

Por último, veamos que su suma es 1.

$$\sum_{i=1}^n p_i^* = \sum_{i=1}^n p_i + \epsilon - \epsilon = \sum_{i=1}^n p_i = 1.$$

Con todo esto, se tiene que la previsión $P^*(f)$ cumple que:

$$\begin{aligned} P^*(f) &= \sum_{i=1}^n a_i p_i^* = \sum_{i \neq j, k} a_i p_i^* + a_j p_j^* + a_k p_k^* \\ &= \sum_{i \neq j, k} a_i p_i + a_j (p_j - \epsilon) + a_k (p_k + \epsilon) = \sum_{i=1}^n a_i p_i + (a_k - a_j) \epsilon. \end{aligned}$$

Pero $a_k - a_j < 0$ por ser $a_k = \min_{1 \leq i \leq n} a_i$ y $a_j = \max_{1 \leq i \leq n} a_i$ y $\epsilon > 0$, por lo que tenemos que:

$$P^*(f) = \sum_{i=1}^n a_i p_i + (a_k - a_j) \epsilon < \sum_{i=1}^n a_i p_i = P(f) = \underline{P}_E(f).$$

Veamos ahora que $P^* \in B_E^\delta(P_0)$, es decir, que $d_E(P^*, P_0) \leq \delta$, o equivalentemente que $d_E^2(P^*, P_0) \leq \delta^2$. Usando la notación $P_0(\{x_i\}) = p_i^0$ para cada

2.1. DEFINICIÓN DEL MODELO

$i = 1, \dots, n$, se tiene que:

$$\begin{aligned}
d_E^2(P^*, P_0) &= \sum_{i=1}^n (P^*(\{x_i\}) - P_0(\{x_i\}))^2 = \sum_{i=1}^n (p_i^* - p_i^0)^2 \\
&= \sum_{i \neq j, k} (p_i^* - p_i^0)^2 + (p_j^* - p_j^0)^2 + (p_k^* - p_k^0)^2 \\
&= \sum_{i \neq j, k} (p_i - p_i^0)^2 + (p_j - \epsilon - p_j^0)^2 + (p_k + \epsilon - p_k^0)^2 \\
&= \sum_{i \neq j, k} (p_i - p_i^0)^2 + ((p_j - p_j^0) - \epsilon)^2 + ((p_k - p_k^0) + \epsilon)^2 \\
&= \sum_{i \neq j, k} (p_i - p_i^0)^2 + ((p_j - p_j^0)^2 - 2\epsilon(p_j - p_j^0) + \epsilon^2) \\
&\quad + ((p_k - p_k^0)^2 + 2\epsilon(p_k - p_k^0) + \epsilon^2) \\
&= \sum_{i \neq j, k} (p_i - p_i^0)^2 + (p_j - p_j^0)^2 + (p_k - p_k^0)^2 \\
&\quad - 2\epsilon(p_j - p_j^0) + 2\epsilon(p_k - p_k^0) + 2\epsilon^2 \\
&= \sum_{i=1}^n (p_i - p_i^0)^2 - 2\epsilon(p_j - p_j^0) + 2\epsilon(p_k - p_k^0) + 2\epsilon^2 \\
&= d_E^2(P, P_0) - 2\epsilon(p_j - p_k + p_k^0 - p_j^0) + 2\epsilon^2.
\end{aligned}$$

Como $\epsilon < p_j < 1$ se tiene que $\epsilon^2 < \epsilon$ y además se tiene que:

$$|p_j - p_k + p_k^0 - p_j^0| < 2.$$

Usando estas dos propiedades se obtiene que:

$$\begin{aligned}
d_E^2(P^*, P_0) &= d_E^2(P, P_0) - 2\epsilon(p_j - p_k + p_k^0 - p_j^0) + 2\epsilon^2 \\
&< d_E^2(P, P_0) + 4\epsilon + 2\epsilon = d_E^2(P, P_0) + 6\epsilon \\
&< d_E^2(P, P_0) + 6 \frac{\delta^2 - d_E^2(P, P_0)}{6} = \delta^2.
\end{aligned}$$

Finalmente, hemos encontrado una probabilidad P^* en la bola $B_E^\delta(P_0)$ tal que $P^*(f) < \underline{P}_E(f)$. Llegamos a contradicción al suponer que $d_E(P, P_0) < \delta$, por tanto $d_E(P, P_0) = \delta$. \square

A continuación, vamos a estudiar la expresión que toma la previsión inferior $\underline{P}_E(f)$ para una apuesta cualquiera $f(x) = \sum_{i=1}^n a_i I_{x_i}(x)$ no constante asociada al conjunto credal $\mathcal{M}(\underline{P}_E)$.

Proposición 2.3. *Sea $P_0 \in \mathbb{P}^*(\mathcal{X})$, $\delta > 0$ de forma que $B_E^\delta(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$, y sea f una apuesta no constante. Entonces la previsión inferior $\underline{P}_E(f)$ y superior $\overline{P}_E(f)$ vienen dadas por:*

$$\underline{P}_E(f) = P_0(f) - \delta\sqrt{n}S_f \quad (2.1)$$

2.1. DEFINICIÓN DEL MODELO

$$\overline{P}_E(f) = P_0(f) + \delta\sqrt{n}S_f \quad (2.2)$$

donde $\overline{f} = \frac{1}{n} \sum_{i=1}^n a_i$ es el valor medio de los valores de f y donde $S_f = \sqrt{\frac{1}{n} \sum_{i=1}^n (a_i - \overline{f})^2}$ es la desviación típica de los valores de la apuesta f .

Demostración. La previsión inferior inducida por este modelo viene definida por:

$$\underline{P}_E(f) = \min\{P(f) \mid P \in B_E^\delta(P_0)\} = \min\{P(f) \mid d_E(P, P_0) \leq \delta\}.$$

Por este motivo nos proponemos calcular $\min\{P(f) \mid P \in B_E^\delta(P_0)\}$ donde $P(f)$ viene dada por $P(f) = \sum_{i=1}^n a_i P(\{x_i\})$.

Fijémonos que toda $P \in B_E^\delta(P_0)$ se puede expresar como $P(\{x_i\}) = P_0(\{x_i\}) + \alpha_i$ donde $\sum_{i=1}^n \alpha_i = 0$. Entonces se tiene que:

$$\begin{aligned} \min_{P \in B_E^\delta(P_0)} P(f) &= \min_{P \in B_E^\delta(P_0)} \sum_{i=1}^n P(\{x_i\})a_i = \min_{\alpha_1, \dots, \alpha_n} \sum_{i=1}^n (P_0(\{x_i\}) + \alpha_i)a_i \\ &= \min_{\alpha_1, \dots, \alpha_n} \left(\sum_{i=1}^n P_0(\{x_i\})a_i + \sum_{i=1}^n a_i \alpha_i \right) \\ &= \min_{\alpha_1, \dots, \alpha_n} \left(P_0(f) + \sum_{i=1}^n a_i \alpha_i \right) = P_0(f) + \min_{\alpha_1, \dots, \alpha_n} \sum_{i=1}^n a_i \alpha_i. \end{aligned}$$

Además, puesto que podemos expresar $p_i = p_i^0 + \alpha_i$ donde $P(\{x_i\}) = p_i$, $P_0(\{x_i\}) = p_i^0$ para cada $i = 1, \dots, n$ y $d_E(P, P_0) = \delta$, se tiene que:

$$\begin{aligned} p_i = p_i^0 + \alpha_i &\Rightarrow p_i - p_i^0 = \alpha_i \Rightarrow (p_i - p_i^0)^2 = \alpha_i^2 \\ &\Rightarrow \sum_{i=1}^n (p_i - p_i^0)^2 = \sum_{i=1}^n \alpha_i^2 \\ &\Rightarrow \sum_{i=1}^n \alpha_i^2 = \delta^2. \end{aligned}$$

Nos planteamos calcular $\min_{\alpha_1, \dots, \alpha_n} \sum_{i=1}^n a_i \alpha_i$ sujeto a las siguientes condiciones:

$$\sum_{i=1}^n \alpha_i = 0; \quad \sum_{i=1}^n \alpha_i^2 = \delta^2.$$

Usando el método de los multiplicadores de Lagrange obtenemos una función con multiplicadores λ y η dada por:

$$L(\alpha_1, \dots, \alpha_n, \lambda, \eta) = \sum_{i=1}^n \alpha_i a_i - \lambda \sum_{i=1}^n \alpha_i - \eta \left(\sum_{i=1}^n \alpha_i^2 - \delta^2 \right).$$

2.1. DEFINICIÓN DEL MODELO

La función L es de tipo polinómica, por lo que es continua y derivable en todos sus puntos y su gradiente ($\nabla(L)$) viene dado por:

$$\nabla(L) = \begin{pmatrix} \frac{\partial L}{\partial \alpha_1} \\ \frac{\partial L}{\partial \alpha_2} \\ \vdots \\ \frac{\partial L}{\partial \alpha_n} \\ \frac{\partial L}{\partial \lambda} \\ \frac{\partial L}{\partial \eta} \end{pmatrix} = \begin{pmatrix} a_1 - 2\eta\alpha_1 - \lambda \\ a_2 - 2\eta\alpha_2 - \lambda \\ \vdots \\ a_n - 2\eta\alpha_n - \lambda \\ -\sum_{i=1}^n \alpha_i \\ -(\sum_{i=1}^n \alpha_i^2 - \delta^2) \end{pmatrix}$$

Igualando a cero el gradiente encontraremos los puntos estacionarios de esta función:

$$\nabla(L) = \begin{pmatrix} a_1 - 2\eta\alpha_1 - \lambda \\ a_2 - 2\eta\alpha_2 - \lambda \\ \vdots \\ a_n - 2\eta\alpha_n - \lambda \\ -\sum_{i=1}^n \alpha_i \\ -(\sum_{i=1}^n \alpha_i^2 - \delta^2) \end{pmatrix} = 0.$$

Sumando las n primeras ecuaciones y usando que $\sum_{i=1}^n \alpha_i = 0$ se tiene que:

$$\sum_{i=1}^n a_i - 2\eta \sum_{i=1}^n \alpha_i - n\lambda = 0 \Rightarrow \lambda = \frac{\sum_{i=1}^n a_i}{n} = \bar{f}.$$

donde \bar{f} es el valor medio de los valores de la apuesta f .

En general, las n primeras ecuaciones vienen dadas por $-2\eta\alpha_i - \bar{f} + a_i = 0$ para cada $i = 1, \dots, n$.

Despejando α_i se tiene que $\alpha_i = \frac{\bar{f} - a_i}{-2\eta}$, elevando al cuadrado esta expresión y sumando las n ecuaciones tenemos que:

$$\begin{aligned} \delta^2 &= \sum_{i=1}^n \alpha_i^2 = \sum_{i=1}^n \frac{(\bar{f} - a_i)^2}{4\eta^2} \Rightarrow 4\eta^2 \delta^2 = \sum_{i=1}^n (\bar{f} - a_i)^2 \Rightarrow \\ \eta &= \sqrt{\sum_{i=1}^n \frac{(\bar{f} - a_i)^2}{4\delta^2}} = \frac{1}{2\delta} \sqrt{n} S_f. \end{aligned}$$

Hemos encontrado los puntos estacionarios de la función L en $(\alpha_1, \dots, \alpha_n, \lambda, \eta)$ donde:

$$\lambda = \bar{f}; \quad \eta = \frac{1}{2\delta} \sqrt{n} S_f; \quad \alpha_i = \frac{\bar{f} - a_i}{-2\eta} = \delta \frac{a_i - \bar{f}}{\sqrt{n} S_f}.$$

Veamos si estos puntos estacionarios son máximos o mínimos. Para ello debemos calcular los $2 + r$ menores principales, denotados B_r , de la matriz Hessiana para $r = 2 + 1, \dots, n$ y observar si $(-1)^r |B_r|$ es positivo o negativo.

2.1. DEFINICIÓN DEL MODELO

La matriz Hessiana limitada ($\mathcal{H}(L)$) viene dada por las derivadas parciales de la función de Lagrange:

$$\begin{aligned} \mathcal{H}(L) &= \begin{pmatrix} \frac{\partial^2 L}{\partial \lambda^2} & \frac{\partial^2 L}{\partial \lambda \partial \eta} & \frac{\partial^2 L}{\partial \lambda \partial \alpha_1} & \cdots & \frac{\partial^2 L}{\partial \lambda \partial \alpha_n} \\ \frac{\partial^2 L}{\partial \lambda \partial \eta} & \frac{\partial^2 L}{\partial \eta^2} & \frac{\partial^2 L}{\partial \eta \partial \alpha_1} & \cdots & \frac{\partial^2 L}{\partial \eta \partial \alpha_n} \\ \frac{\partial^2 L}{\partial \lambda \partial \alpha_1} & \frac{\partial^2 L}{\partial \eta \partial \alpha_1} & \frac{\partial^2 L}{\partial \alpha_1^2} & \cdots & \frac{\partial^2 L}{\partial \alpha_1 \partial \alpha_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 L}{\partial \lambda \partial \alpha_n} & \frac{\partial^2 L}{\partial \eta \partial \alpha_n} & \frac{\partial^2 L}{\partial \alpha_1 \partial \alpha_n} & \cdots & \frac{\partial^2 L}{\partial \alpha_n^2} \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & -1 & -1 & \cdots & -1 \\ 0 & 0 & -2\alpha_1 & -2\alpha_2 & \cdots & -2\alpha_n \\ -1 & -2\alpha_1 & -2\eta & 0 & \cdots & 0 \\ -1 & -2\alpha_2 & 0 & -2\eta & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -2\alpha_n & 0 & 0 & \cdots & -2\eta \end{pmatrix} \end{aligned}$$

Entonces, los menores principales B_r para $r = 2 + 1, \dots, n$ vienen dados por:

$$\begin{aligned} B_r &= \begin{pmatrix} \frac{\partial^2 L}{\partial \lambda^2} & \frac{\partial^2 L}{\partial \lambda \partial \eta} & \frac{\partial^2 L}{\partial \lambda \partial \alpha_1} & \cdots & \frac{\partial^2 L}{\partial \lambda \partial \alpha_r} \\ \frac{\partial^2 L}{\partial \lambda \partial \eta} & \frac{\partial^2 L}{\partial \eta^2} & \frac{\partial^2 L}{\partial \eta \partial \alpha_1} & \cdots & \frac{\partial^2 L}{\partial \eta \partial \alpha_r} \\ \frac{\partial^2 L}{\partial \lambda \partial \alpha_1} & \frac{\partial^2 L}{\partial \eta \partial \alpha_1} & \frac{\partial^2 L}{\partial \alpha_1^2} & \cdots & \frac{\partial^2 L}{\partial \alpha_1 \partial \alpha_r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 L}{\partial \lambda \partial \alpha_r} & \frac{\partial^2 L}{\partial \eta \partial \alpha_r} & \frac{\partial^2 L}{\partial \alpha_1 \partial \alpha_r} & \cdots & \frac{\partial^2 L}{\partial \alpha_r^2} \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & -1 & -1 & \cdots & -1 \\ 0 & 0 & -2\alpha_1 & -2\alpha_2 & \cdots & -2\alpha_r \\ -1 & -2\alpha_1 & -2\eta & 0 & \cdots & 0 \\ -1 & -2\alpha_2 & 0 & -2\eta & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -2\alpha_r & 0 & 0 & \cdots & -2\eta \end{pmatrix} \end{aligned}$$

Si dividimos la matriz B_r en cuatro submatrices tal que:

$$B_r = \begin{pmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \end{pmatrix}$$

entonces el determinante de B_r viene dado por:

$$|B_r| = |B_{2,2}| \cdot |B_{1,1} - B_{1,2} B_{2,2}^{-1} B_{2,1}|. \quad (2.3)$$

Sean los vectores $\vec{\alpha}_r^t = (\alpha_1, \dots, \alpha_r)$ y $\vec{1}_r^t = (1, \dots, 1)$, ambos de dimensión r , y sea I_r la matriz identidad de orden r , entonces podemos tomar la

2.1. DEFINICIÓN DEL MODELO

siguiente descomposición en submatrices de la matriz B_r :

$$\begin{aligned} B_{1,1} &= 0_{2 \times 2}; & B_{1,2} &= \begin{pmatrix} -\vec{1}_r^t \\ -2\vec{\alpha}_r^t \end{pmatrix}; \\ B_{2,1} &= (-\vec{1}_r, -2\vec{\alpha}_r); & B_{2,2} &= -2\eta I_r. \end{aligned}$$

Utilizando esta descomposición de B_r y la Ecuación (2.3) se puede calcular su determinante:

$$\begin{aligned} |B_r| &= |-2\eta I_r| \left| \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} -\vec{1}_r^t \\ -2\vec{\alpha}_r^t \end{pmatrix} + \begin{pmatrix} -1 \\ 2\eta \end{pmatrix} I_r \begin{pmatrix} -\vec{1}_r, -2\vec{\alpha}_r \end{pmatrix} \right| \\ &= |-2\eta I_r| \left| \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} - \frac{1}{2\eta} \begin{pmatrix} -r & -\sum_{i=1}^r 2\alpha_i \\ -\sum_{i=1}^r 2\alpha_i & -\sum_{i=1}^r 4\alpha_i^2 \end{pmatrix} \right| \\ &= |-2\eta I_r| \left| \frac{1}{-2\eta} \begin{pmatrix} -r & -\sum_{i=1}^r 2\alpha_i \\ -\sum_{i=1}^r 2\alpha_i & -\sum_{i=1}^r 4\alpha_i^2 \end{pmatrix} \right| \\ &= (-2\eta)^r \left(\frac{4r}{(2\eta)^2} \sum_{i=1}^r \alpha_i^2 - \frac{4}{(2\eta)^2} \left(\sum_{i=1}^r \alpha_i \right)^2 \right) \\ &= (-2\eta)^{r-2} 4 \left(r \sum_{i=1}^r \alpha_i^2 - \left(\sum_{i=1}^r \alpha_i \right)^2 \right). \end{aligned}$$

Finalmente teniendo en cuenta que $\eta = \frac{1}{2\delta} \sqrt{n} S_f > 0$ y que la varianza de los r primeros valores de α_i , es decir $\alpha_1, \dots, \alpha_r$, es siempre positiva, ya que $\alpha_i \in \mathbb{R}$ para todo $i = 1, \dots, n$, y viene dada por:

$$\begin{aligned} Var(\alpha_1, \dots, \alpha_r) &= \frac{1}{r} \sum_{i=1}^r \alpha_i^2 - \left(\frac{1}{r} \sum_{i=1}^r \alpha_i \right)^2 \\ &= \frac{1}{r^2} \left(r \sum_{i=1}^r \alpha_i^2 - \left(\sum_{i=1}^r \alpha_i \right)^2 \right) \end{aligned} \quad (2.4)$$

de donde se deduce que $(r \sum_{i=1}^r \alpha_i^2 - (\sum_{i=1}^r \alpha_i)^2) > 0$. Con todo esto resulta que:

$$(-1)^r |B_r| = (-1)^r (-2\eta)^{r-2} 4 \left(r \sum_{i=1}^r \alpha_i^2 - \left(\sum_{i=1}^r \alpha_i \right)^2 \right)$$

pero $(-1)^r (-2\eta)^{r-2} > 0$ pues ambos exponentes son pares o impares simultáneamente y por la Ecuación (2.4) se tiene que $(-1)^r |B_r| > 0$ para todo $r = 2 + 1, \dots, n$, y por tanto queda demostrado que los puntos estacionarios son máximos.

2.1. DEFINICIÓN DEL MODELO

Entonces la previsión superior viene dada por:

$$\begin{aligned}\bar{P}_E(f) &= \max_{P \in B_E^{\delta}(P_0)} P(f) = P_0(f) + \max_{\alpha_1, \dots, \alpha_n} \sum_{i=1}^n a_i \alpha_i \\ &= P_0(f) + \sum_{i=1}^n a_i \left(\delta \frac{a_i - \bar{f}}{\sqrt{n} S_f} \right).\end{aligned}$$

Teniendo en cuenta que $\sum_{i=1}^n a_i \bar{f} = n \bar{f}^2$, ya que:

$$\begin{aligned}\sum_{i=1}^n a_i \bar{f} &= \sum_{i=1}^n a_i \frac{\sum_{j=1}^n a_j}{n} = \frac{(a_1 + \dots + a_n)(a_1 + \dots + a_n)}{n} \\ &= \frac{(a_1 + \dots + a_n)^2}{n} = n \frac{(a_1 + \dots + a_n)^2}{n^2} = n \bar{f}^2\end{aligned}\quad (2.5)$$

se puede simplificar la expresión $\sum_{i=1}^n a_i \left(\delta \frac{a_i - \bar{f}}{\sqrt{n} S_f} \right)$. Multiplicando y dividiendo por \sqrt{n} y usando (2.5) se tiene que:

$$\begin{aligned}\sum_{i=1}^n a_i \left(\delta \frac{a_i - \bar{f}}{\sqrt{n} S_f} \right) &= \frac{\delta \sqrt{n}}{S_f} \frac{1}{n} \sum_{i=1}^n (a_i^2 - a_i \bar{f}^2) = \frac{\delta \sqrt{n}}{S_f} \frac{1}{n} \left(\sum_{i=1}^n a_i^2 - \sum_{i=1}^n a_i \bar{f}^2 \right) \\ &= \frac{\delta \sqrt{n}}{S_f} \frac{1}{n} \left(\sum_{i=1}^n a_i^2 - n \bar{f}^2 \right) = \frac{\delta \sqrt{n}}{S_f} \left(\frac{1}{n} \sum_{i=1}^n a_i^2 - \bar{f}^2 \right).\end{aligned}$$

Veamos que $\frac{1}{n} \sum_{i=1}^n a_i^2 - \bar{f}^2$ es exactamente S_f^2 :

$$\begin{aligned}S_f^2 &= \frac{1}{n} \sum_{i=1}^n (a_i - \bar{f})^2 = \frac{1}{n} \sum_{i=1}^n (a_i^2 - 2a_i \bar{f} + \bar{f}^2) \\ &= \frac{1}{n} \sum_{i=1}^n a_i^2 + \frac{1}{n} \sum_{i=1}^n (-2a_i \bar{f} + \bar{f}^2).\end{aligned}$$

Usando (2.5) se tiene que:

$$\begin{aligned}\sum_{i=1}^n (-2a_i \bar{f} + \bar{f}^2) &= \left(-2 \sum_{i=1}^n a_i \bar{f} + \sum_{i=1}^n \bar{f}^2 \right) = (-2n \bar{f}^2 + n \bar{f}^2) \\ &= -2n \bar{f}^2 + n \bar{f}^2 = -n \bar{f}^2.\end{aligned}$$

y por tanto:

$$\begin{aligned}S_f^2 &= \frac{1}{n} \sum_{i=1}^n a_i^2 + \frac{1}{n} \sum_{i=1}^n (-2a_i \bar{f} + \bar{f}^2) = \frac{1}{n} \sum_{i=1}^n a_i^2 + \frac{1}{n} (-n \bar{f}^2) \\ &= \frac{1}{n} \sum_{i=1}^n a_i^2 - \bar{f}^2.\end{aligned}$$

2.1. DEFINICIÓN DEL MODELO

Finalmente hemos obtenido que:

$$\sum_{i=1}^n a_i \left(\delta \frac{a_i - \bar{f}}{\sqrt{n} S_f} \right) = \frac{\delta \sqrt{n}}{S_f} S_f^2 = \delta \sqrt{n} S_f$$

por lo que la expresión de la previsión superior viene dada por:

$$\bar{P}_E(f) = P_0(f) + \delta \sqrt{n} S_f.$$

Para calcular la previsión inferior solo debemos tener en cuenta que

$$\underline{P}_E(f) = -\bar{P}_E(-f),$$

donde la apuesta $-f$ viene dada por $-f(x) = \sum_{i=1}^n (-a_i) I_{\{x_i\}}(x)$. Entonces

$$\bar{P}_E(-f) = P_0(-f) + \delta \sqrt{n} S_{-f},$$

donde $P_0(-f) = -P_0(f)$. Veamos que $S_{-f} = S_f$, para ello vamos a ver primero que $\overline{-f} = -\bar{f}$:

$$\overline{-f} = \sum_{i=1}^n \frac{-a_i}{n} = - \sum_{i=1}^n \frac{a_i}{n} = -\bar{f}.$$

Entonces, se tiene que:

$$\begin{aligned} S_{-f} &= \sqrt{\frac{1}{n} \sum_{i=1}^n ((-a_i) - (\overline{-f}))^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n ((-a_i) + \bar{f})^2} \\ &= \sqrt{\frac{1}{n} \sum_{i=1}^n (a_i - \bar{f})^2} = S_f. \end{aligned}$$

Por todo esto, la previsión inferior queda determinada por:

$$\underline{P}_E(f) = - (P_0(-f) + \delta \sqrt{n} S_{-f}) = P_0(f) - \delta \sqrt{n} S_f.$$

□

Ahora vamos a estudiar la probabilidad inferior asociada mediante la restricción de la previsión inferior $\underline{P}_E(f)$ a sucesos, es decir, la previsión inferior para la apuesta $f = I_A$, que denotaremos $\underline{P}_E(I_A) = \underline{P}_E(A)$.

Esto es sencillo puesto que si de forma general $f(x) = \sum_{i=1}^n a_i I_{x_i}(x)$, en el caso particular de que $f = I_A$ se tiene que $a_i = 0$ si $x_i \notin A$ y $a_i = 1$ si $x_i \in A$.

2.1. DEFINICIÓN DEL MODELO

Corolario 2.4. Sea $P_0 \in \mathbb{P}^*(\mathcal{X})$, $\delta > 0$ de forma que $B_E^\delta(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$, y sea $A \neq \emptyset, \mathcal{X}$. Entonces la probabilidad inferior y superior del suceso A vienen dadas por:

$$\underline{P}_E(A) = P_0(A) - \delta \sqrt{\frac{|A|(n - |A|)}{n}} \quad (2.6)$$

$$\overline{P}_E(A) = P_0(A) + \delta \sqrt{\frac{|A|(n - |A|)}{n}} \quad (2.7)$$

donde $|A|$ indica el cardinal de A .

Demostración. Al tomar como apuesta f la función I_A se tiene que la media \bar{f} y la desviación típica S_f de los valores de la apuesta f quedan determinados en función del cardinal de A de la siguiente forma:

$$\bar{f} = \frac{\sum_{i=1}^n a_i}{n} = \frac{1}{n} \sum_{x_i \in A} a_i = \frac{1}{n} \sum_{x_i \in A} 1 = \frac{|A|}{n}.$$

$$S_f = \sqrt{\frac{\sum_{i=1}^n (a_i - \bar{f})^2}{n}} = \frac{1}{\sqrt{n}} \sqrt{\sum_{i=1}^n (a_i - \bar{f})^2} = \frac{1}{\sqrt{n}} \sqrt{\sum_{i=1}^n (a_i^2 - 2a_i\bar{f} + \bar{f}^2)}.$$

Teniendo en cuenta que:

$$\sum_{i=1}^n a_i^2 = \sum_{x_i \in A} a_i^2 = \sum_{x_i \in A} 1^2 = |A|.$$

$$\sum_{i=1}^n (-2a_i\bar{f}) = -2\bar{f} \sum_{x_i \in A} a_i = -2\bar{f}|A| = -2|A| \frac{\sum_{x_i \in A} a_i}{n} = -2 \frac{|A|^2}{n}.$$

$$\sum_{i=1}^n \bar{f}^2 = \sum_{i=1}^n \left(\frac{\sum_{x_i \in A} a_i}{n} \right)^2 = n \frac{|A|^2}{n^2} = \frac{|A|^2}{n}$$

se tiene que:

$$\begin{aligned} S_f &= \frac{1}{\sqrt{n}} \sqrt{\sum_{i=1}^n (a_i^2 - 2a_i\bar{f} + \bar{f}^2)} = \frac{1}{\sqrt{n}} \sqrt{|A| - 2 \frac{|A|^2}{n} + \frac{|A|^2}{n}} \\ &= \frac{1}{\sqrt{n}} \sqrt{|A| - \frac{|A|^2}{n}} = \frac{1}{\sqrt{n}} \sqrt{\frac{n|A| - |A|^2}{n}} = \frac{1}{\sqrt{n}} \sqrt{\frac{|A|(n - |A|)}{n}} \\ &= \frac{1}{n} \sqrt{|A|(n - |A|)}. \end{aligned}$$

Entonces, sustituyendo estas nuevas expresiones para la desviación típica y la media en las ecuaciones (2.1) y (2.2) obtenemos que:

$$\underline{P}_E(A) = P_0(A) - \delta \sqrt{\frac{|A|(n - |A|)}{n}}.$$

$$\overline{P}_E(A) = P_0(A) + \delta \sqrt{\frac{|A|(n - |A|)}{n}}.$$

□

2.2. Propiedades

En la Proposición 2.2 hemos visto que si $\underline{P}_E(f) = P(f)$ para algún $P \in B_E^\delta(P_0)$ y una apuesta f cualquiera entonces se tiene que $d_E(P, P_0) = \delta$, es decir, $ext(\mathcal{M}(\underline{P}_E)) \subseteq fr(B_E^\delta(P_0))$.

A continuación vamos a estudiar que $ext(\mathcal{M}(\underline{P}_E)) \supseteq fr(B_E^\delta(P_0))$ para así demostrar que realmente se trata de una igualdad.

Proposición 2.5. *Sea $P_0 \in \mathbb{P}^*(\mathcal{X})$, $\delta > 0$ de forma que $B_E^\delta(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$. Entonces:*

$$ext(\mathcal{M}(\underline{P}_E)) = fr(B_E^\delta(P_0)).$$

Demostración. En la Proposición 2.2 hemos visto que si $P \in ext(\mathcal{M}(\underline{P}_E))$, entonces se tiene que $d_E(P, P_0) = \delta$, es decir $P \in fr(B_E^\delta(P_0))$. Por lo tanto $ext(\mathcal{M}(\underline{P}_E)) \subseteq fr(B_E^\delta(P_0))$.

Veamos ahora que $ext(\mathcal{M}(\underline{P}_E)) \supseteq fr(B_E^\delta(P_0))$, para ello veamos que si $P \in fr(B_E^\delta(P_0))$, es decir, que si $d_E(P, P_0) = \delta$ entonces existe una apuesta f tal que $\underline{P}_E(f) = P(f)$ y además, que para cualquier otra probabilidad Q verificando que $\overline{P}_E(f) = Q(f)$ y que $d_E(Q, P_0) = \delta$ se tiene que necesariamente $Q = P$. Es decir, vamos a ver que P es la única probabilidad en la bola que alcanza la previsión superior de la apuesta.

Sea $P \in fr(B_E^\delta(P_0))$, en la Proposición 2.3 fue demostrado que podemos expresar cualquier $P \in B_E^\delta(P_0)$ en función de P_0 como $p_i = p_i^0 + \alpha_i$ verificandose que $\sum_{i=1}^n \alpha_i = 0$ y que $\sum_{i=1}^n \alpha_i^2 = \delta^2$.

Vamos a fijar la apuesta f dada por $f(x) = \sum_{i=1}^n \alpha_i I_{\{x_i\}}(x)$. Para esta apuesta tenemos que:

$$P(f) = \sum_{i=1}^n \alpha_i p_i = \sum_{i=1}^n \alpha_i (p_i^0 + \alpha_i) = \sum_{i=1}^n \alpha_i p_i^0 + \sum_{i=1}^n \alpha_i^2 = P_0(f) + \sum_{i=1}^n \alpha_i^2.$$

Usando que por hipótesis $\sum_{i=1}^n \alpha_i^2 = \delta^2$ se tiene que $P(f) = P_0(f) + \delta^2$.

Por otro lado, la previsión superior hemos visto que viene dada por $\overline{P}_E(f) = P_0(f) + \delta\sqrt{n}S_f$ donde S_f es la desviación típica de los valores de la apuesta f . Para esta apuesta f resulta que:

$$\overline{f} = \frac{\sum_{i=1}^n \alpha_i}{n} = \frac{0}{n} = 0.$$

Y por tanto la desviación típica de f viene dada por:

$$S_f = \frac{1}{\sqrt{n}} \sqrt{\sum_{i=1}^n (\alpha_i - \overline{f})^2} = \frac{1}{\sqrt{n}} \sqrt{\sum_{i=1}^n \alpha_i^2} = \frac{\sqrt{\delta^2}}{\sqrt{n}} = \frac{\delta}{\sqrt{n}}.$$

2.2. PROPIEDADES

Entonces la previsión superior viene dada por:

$$\bar{P}_E(f) = P_0(f) + \delta\sqrt{n}S_f = P_0(f) + \delta\sqrt{n}\frac{\delta}{\sqrt{n}} = P_0(f) + \delta^2.$$

Hemos encontrado una apuesta f para la que se verifica que $\bar{P}_E(f) = P(f)$.

Ahora, por reducción al absurdo, tomemos una probabilidad cualquiera $Q \neq P$ verificando que $d_E(Q, P_0) = \delta$ y que $Q(f) = \bar{P}_E(f)$ para la apuesta que se ha fijado previamente.

Podemos expresar la probabilidad Q como $q_i = p_i + \beta_i = p_i^0 + \alpha_i + \beta_i$ para todo $i = 1, \dots, n$, donde los valores β_i han de cumplir que:

$$\sum_{i=1}^n \beta_i = 0.$$

De esta forma, $d_E(Q, P_0)$ viene dada por:

$$\begin{aligned} \delta^2 &= d_E(Q, P_0)^2 = \sum_{i=1}^n (q_i - p_i^0)^2 = \sum_{i=1}^n (p_i^0 + \beta_i + \alpha_i - p_i^0)^2 \\ &= \sum_{i=1}^n (\beta_i + \alpha_i)^2 = \sum_{i=1}^n (\alpha_i^2 + 2\beta_i\alpha_i + \beta_i^2) \\ &= \sum_{i=1}^n \alpha_i^2 + \sum_{i=1}^n (\beta_i^2 + 2\beta_i\alpha_i) = \delta^2 + \sum_{i=1}^n (\beta_i^2 + 2\beta_i\alpha_i) \\ &\Rightarrow \sum_{i=1}^n (\beta_i^2 + 2\beta_i\alpha_i) = 0 \Rightarrow \sum_{i=1}^n \beta_i^2 + 2\sum_{i=1}^n \beta_i\alpha_i = 0. \end{aligned}$$

Resulta que $\sum_{i=1}^n \beta_i^2 \geq 0$ ya que $\beta_i^2 \geq 0 \quad \forall i = 1, \dots, n$, y por tanto es necesario que $2\sum_{i=1}^n \beta_i\alpha_i \leq 0$.

Por otro lado, desarrollando la expresión de $Q(f)$ se tiene que:

$$\begin{aligned} Q(f) &= \sum_{i=1}^n \alpha_i q_i = \sum_{i=1}^n \alpha_i (p_i + \beta_i) = \sum_{i=1}^n \alpha_i (p_i^0 + \alpha_i + \beta_i) \\ &= \sum_{i=1}^n \alpha_i p_i^0 + \sum_{i=1}^n \alpha_i^2 + \sum_{i=1}^n \alpha_i \beta_i = P_0(f) + \delta^2 + \sum_{i=1}^n \alpha_i \beta_i \end{aligned}$$

pero por hipótesis, $Q(f) = \bar{P}_E(f) = P_0(f) + \delta^2$, por lo que $\sum_{i=1}^n \alpha_i \beta_i = 0$.

Entonces, se tiene que:

$$\begin{aligned} 0 &= \sum_{i=1}^n \beta_i^2 + 2\sum_{i=1}^n \beta_i\alpha_i = \sum_{i=1}^n \beta_i^2 \\ &\Rightarrow \sum_{i=1}^n \beta_i^2 = 0 \Rightarrow \beta_i = 0 \quad \forall i = 1, \dots, n. \end{aligned}$$

2.2. PROPIEDADES

Esto nos lleva a contradicción, ya que si $\beta_i = 0 \quad \forall i = 1, \dots, n$, entonces $Q = P$, ya que $q_i = p_i + \beta_i \quad \forall i = 1, \dots, n$.

Acabamos de demostrar que para la apuesta $f = \sum_{i=1}^n \alpha_i I_{\{x_i\}}$, P es la única probabilidad en la bola $B_E^\delta(P_0)$ que alcanza la probabilidad superior, y por tanto, P es un punto extremo del conjunto credal $\mathcal{M}(\underline{P}_E)$. \square

A continuación, vamos a estudiar si la previsión inferior \underline{P}_E cumple la propiedad de 2-monotonía.

Proposición 2.6. *Sea $P_0 \in \mathbb{P}^*(\mathcal{X})$, $\delta > 0$ de forma que $B_E^\delta(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$. Entonces la previsión inferior \underline{P}_E no es 2-monótona.*

Demostración. Podemos verlo como una consecuencia directa de la Proposición 2.5 y del hecho de que si \underline{P}_E fuese 2-monótona entonces $B_E^\delta(P_0)$ tendría a lo sumo $n!$ puntos extremos.

La Proposición 2.5 nos dice que el número de puntos extremos del conjunto $B_E^\delta(P_0)$ es igual al número de puntos en la frontera de la bola $B_E^\delta(P_0)$, pero sabemos que existen infinitos puntos en la frontera de la bola por lo que el número de puntos extremos va a ser mayor que $n!$, y por tanto \underline{P}_E no puede ser 2-monótona. \square

Veamos ahora que considerando la restricción a sucesos de la previsión inferior se tiene que \underline{P}_E es 2-monótona. Antes vamos a demostrar dos propiedades de funciones cóncavas que nos serán útiles en la demostración de la 2-monotonía sobre sucesos.

Primero vamos a definir cuando una función es cóncava, decimos que una función φ definida sobre un intervalo $[a, b]$ es cóncava en $[a, b]$ si verifica que:

$$\varphi(tx + (1-t)y) \geq t\varphi(x) + (1-t)\varphi(y) \quad \forall t \in [0, 1] \quad \forall x, y \in [a, b].$$

Lema 2.7. *Sea $\varphi : [a, b] \rightarrow [a, b]$ una función cóncava y creciente y sea $\phi : [a, b] \rightarrow [a, b]$ una función cóncava. Entonces la composición $\varphi \circ \phi$ es también una función cóncava.*

Demostración. Veamos que $\varphi \circ \phi$ verifica la definición de concavidad. Usando que por hipótesis la función ϕ es cóncava se tiene que:

$$\phi(tx + (1-t)y) \geq t\phi(x) + (1-t)\phi(y) \quad \forall t \in [0, 1] \quad \forall x, y \in [a, b]$$

y como además φ es creciente podemos asegurar que:

$$(\varphi \circ \phi)(tx + (1-t)y) = \varphi(\phi(tx + (1-t)y)) \geq \varphi(t\phi(x) + (1-t)\phi(y)).$$

Además, como φ es también una función cóncava verifica que:

$$\varphi(tx + (1-t)y) \geq t\varphi(x) + (1-t)\varphi(y) \quad \forall t \in [0, 1] \quad \forall x, y \in [a, b]$$

2.2. PROPIEDADES

por lo tanto se tiene que:

$$\begin{aligned}\varphi(t\phi(x) + (1-t)\phi(y)) &\geq t\varphi(\phi(x)) + (1-t)\varphi(\phi(y)) \\ &= t(\varphi \circ \phi)(x) + (1-t)(\varphi \circ \phi)(y).\end{aligned}$$

Acabamos de demostrar que $\varphi \circ \phi$ es una función cóncava. \square

Lema 2.8. *La función $\varphi(t) = \sqrt{t(n-t)}$ definida sobre el intervalo $[0, n]$ es cóncava.*

Demostración. La función $\varphi(t)$ podemos descomponerla como composición de dos funciones $\varphi_1, \varphi_2 : [0, \infty) \rightarrow [0, \infty)$ dadas por:

$$\varphi_1(t) = \sqrt{t}; \quad \varphi_2(t) = t(n-t)$$

de forma que $\varphi(t) = \varphi_1(\varphi_2(t)) = \sqrt{t(n-t)}$.

Vamos a ver que la función $\varphi_1(t)$ es una función creciente y cóncava ya que su primera derivada es positiva y su segunda derivada resulta siempre negativa:

$$\begin{aligned}\varphi_1'(t) &= \frac{1}{2\sqrt{t}} > 0. \\ \varphi_1''(t) &= \frac{-1}{4\sqrt{t^3}} < 0.\end{aligned}$$

Además $\varphi_2(t)$ es también cóncava puesto que su segunda derivada es negativa:

$$\begin{aligned}\varphi_2'(t) &= n - 2t. \\ \varphi_2''(t) &= -2 < 0.\end{aligned}$$

Entonces, aplicando el Lema 2.7 se tiene que $\varphi = \varphi_1 \circ \varphi_2$ es una función cóncava sobre $[0, \infty)$. \square

Una vez vistos estos resultados previos, pasamos a estudiar la propiedad de 2-monotonía para la restricción a sucesos de la previsión inferior.

Proposición 2.9. *Sea $P_0 \in \mathbb{P}^*(\mathcal{X})$, $\delta > 0$ de forma que $B_E^\delta(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$. Entonces la probabilidad inferior \underline{P}_E es 2-monótona.*

Demostración. Sean $A, B \subset \mathcal{X}$, $A, B \neq \emptyset$ verificando que $A \cap B \neq \emptyset$ y que $A \cup B \neq \mathcal{X}$. La probabilidad inferior de estos sucesos viene dada por la

2.2. PROPIEDADES

Ecuacion (2.6), de tal forma que:

$$\begin{aligned}\underline{P}_E(A) &= P_0(A) - \delta \sqrt{\frac{|A|(n-|A|)}{n}}. \\ \underline{P}_E(B) &= P_0(B) - \delta \sqrt{\frac{|B|(n-|B|)}{n}}. \\ \underline{P}_E(A \cap B) &= P_0(A \cap B) - \delta \sqrt{\frac{|A \cap B|(n-|A \cap B|)}{n}}. \\ \underline{P}_E(A \cup B) &= P_0(A \cup B) - \delta \sqrt{\frac{|A \cup B|(n-|A \cup B|)}{n}}.\end{aligned}$$

Usando la fórmula de inclusión-exclusión de las probabilidades dada por $P_0(A \cup B) = P_0(A) + P_0(B) - P_0(A \cap B)$ se tiene que:

$$\begin{aligned}\underline{P}_E(A \cap B) + \underline{P}_E(A \cup B) &= P_0(A \cap B) - \delta \sqrt{\frac{|A \cap B|(n-|A \cap B|)}{n}} \\ &+ P_0(A \cup B) - \delta \sqrt{\frac{|A \cup B|(n-|A \cup B|)}{n}} \\ &= P_0(A \cap B) - \delta \sqrt{\frac{|A \cap B|(n-|A \cap B|)}{n}} + P_0(A) + P_0(B) \\ &- P_0(A \cap B) - \delta \sqrt{\frac{|A \cup B|(n-|A \cup B|)}{n}} = P_0(A) + P_0(B) \\ &- \frac{\delta}{\sqrt{n}} \left(\sqrt{|A \cap B|(n-|A \cap B|)} + \sqrt{|A \cup B|(n-|A \cup B|)} \right).\end{aligned}$$

Por otra parte, se tiene que:

$$\begin{aligned}\underline{P}_E(A) + \underline{P}_E(B) &= P_0(A) - \delta \sqrt{\frac{|A|(n-|A|)}{n}} + P_0(B) - \delta \sqrt{\frac{|B|(n-|B|)}{n}} \\ &= P_0(A) + P_0(B) - \frac{\delta}{\sqrt{n}} \left(\sqrt{|A|(n-|A|)} + \sqrt{|B|(n-|B|)} \right).\end{aligned}$$

Acabamos de comprobar que demostrar que $\underline{P}_E(A \cap B) + \underline{P}_E(A \cup B) \geq \underline{P}_E(A) + \underline{P}_E(B)$ es equivalente a demostrar que:

$$\begin{aligned}\sqrt{|A \cap B|(n-|A \cap B|)} + \sqrt{|A \cup B|(n-|A \cup B|)} &\leq \\ \sqrt{|A|(n-|A|)} + \sqrt{|B|(n-|B|)}.\end{aligned}\tag{2.8}$$

Vamos a considerar 4 números dados por $k_1 = |A \cap B|$, $k_2 = |A|$, $k_3 = |B|$, y $k_4 = |A \cup B|$. Sin pérdida de generalidad podemos suponer que $k_2 \leq k_3$. Así definidos está claro que verifican que $k_1 \leq k_2 \leq k_3 \leq k_4$ y que $k_1 + k_4 = k_2 + k_3$. Entonces se tiene que la Ecuación (2.8) es equivalente a:

$$\sqrt{k_1(n-k_1)} + \sqrt{k_4(n-k_4)} \leq \sqrt{k_2(n-k_2)} + \sqrt{k_3(n-k_3)}.\tag{2.9}$$

2.3. IMPLEMENTACIÓN

Finalmente, vamos a ver que se cumple la Ecuación (2.9). Vamos a considerar la función cóncava $\varphi(t) = \sqrt{t(n-t)}$ dada en el Lema 2.8.

Sean k_1, k_2, k_3, k_4 tales que $k_1 \leq k_2 \leq k_3 \leq k_4$ y que $k_1 + k_4 = k_2 + k_3$. En particular, para $k_1 \leq k_2 \leq k_4$ existe un $\alpha \in [0, 1]$ tal que se verifica que $k_2 = \alpha k_1 + (1 - \alpha)k_4$. Además, como $k_2 = k_1 + k_4 - k_3$ se tiene que:

$$\begin{aligned} k_1 + k_4 - k_3 &= \alpha k_1 + (1 - \alpha)k_4 \Rightarrow \\ k_1 - \alpha k_1 + k_4 - k_4 + \alpha k_4 &= k_3 \Rightarrow \\ (1 - \alpha)k_1 + \alpha k_4 &= k_3. \end{aligned}$$

Usando estas definiciones para k_2 y k_3 sobre la función cóncava φ , se obtiene que:

$$\begin{aligned} \varphi(k_2) &= \varphi(\alpha k_1 + (1 - \alpha)k_4) \geq \alpha \varphi(k_1) + (1 - \alpha)\varphi(k_4). \\ \varphi(k_3) &= \varphi(\alpha k_4 + (1 - \alpha)k_1) \geq (1 - \alpha)\varphi(k_1) + \alpha \varphi(k_4). \end{aligned}$$

Sumando estas expresiones se obtiene que:

$$\begin{aligned} \varphi(k_2) + \varphi(k_3) &\geq \alpha \varphi(k_1) + (1 - \alpha)\varphi(k_4) + (1 - \alpha)\varphi(k_1) + \alpha \varphi(k_4) \\ &= \varphi(k_1) + \varphi(k_4) \end{aligned}$$

Acabamos de demostrar que:

$$\sqrt{k_2(n - k_2)} + \sqrt{k_3(n - k_3)} \geq \sqrt{k_1(n - k_1)} + \sqrt{k_4(n - k_4)}.$$

Y por tanto queda demostrado que \underline{P}_E es 2-monótona en sucesos. \square

Acabamos de demostrar que en el caso del modelo de distorsión inducido por la distancia Euclídea la previsiones inferiores no verifican la propiedad de 2-monotonía, sin embargo, al tomar la restricción a sucesos de estas previsiones si que se verifica la propiedad de 2-monotonía.

2.3. Implementación

En esta sección vamos a realizar varios programas para realizar el cálculo de probabilidades inferiores y previsiones inferiores para el modelo Euclídeo y también para la representación de conjuntos credales. Para estos cálculos se han desarrollado varios programas en el software R y el entorno RStudio.

A partir de una probabilidad inicial P_0 , un parámetro de distorsión $\delta > 0$ y una apuesta $f \in \mathcal{L}(\mathcal{X})$ vamos a realizar los cálculos de la previsión inferior y superior de la apuesta f .

Trabajaremos sobre el espacio muestral $\mathcal{X} = \{x_1, \dots, x_n\}$, para el correcto uso de nuestro programa es necesario introducir P_0 como un vector que contiene las probabilidades de cada valor x_i en orden, es decir:

$$P_0 = (P_0(\{x_1\}), \dots, P_0(\{x_n\})).$$

2.3. IMPLEMENTACIÓN

Nuestro primer programa será para el cálculo de la previsión inferior de una apuesta $f \in \mathcal{L}(\mathcal{X})$ dada en general por $f(x) = \sum_{i=1}^n a_i I_{x_i}(x)$. Para este cálculo usaremos la Ecuación (2.1) en el siguiente programa:

```
1  lpr_E <- function(f,Po,delta){
2    #f es el vector fila que contiene los elementos a_i de la
      apuesta f.
3    #Po es el vector fila que contiene la probabilidad Po de cada
      {x_i}.
4    #delta es el parametro de distorsion.
5    #Calculamos el tamaño del espacio muestral.
6    n <- length(Po)
7    #Calculamos el valor de Po(f)
8    Pof <- sum(f*Po)
9    #Calculamos el valor medio de los valores de la apuesta f.
10   Mediaf <- mean(f)
11   #Calculamos la desviacion tipica de los valores de la apuesta
      f.
12   Sf <- sqrt((1/n)*sum((f-Mediaf)^2))
13   #Calculamos la prevision inferior mediante la Ecuacion 2.1.
14   lpr_E <- Pof - delta*sqrt(n)*Sf
15   return(lpr_E)
16 }
```

De forma muy similar se realiza el cálculo de la previsión superior con la Ecuación (2.2).

```
1  upr_E <- function(f,Po,delta){
2    #f es el vector fila que contiene los elementos a_i de la
      apuesta f.
3    #Po es el vector fila que contiene la probabilidad Po de cada
      {x_i}.
4    #delta es el parametro de distorsion.
5    #Calculamos el tamaño del espacio muestral.
6    n <- length(Po)
7    #Calculamos el valor de Po(f)
8    Pof <- sum(f*Po)
9    #Calculamos el valor medio de los valores de la apuesta f.
10   Mediaf <- mean(f)
11   #Calculamos la desviacion tipica de los valores de la apuesta
      f.
12   Sf <- sqrt((1/n)*sum((f-Mediaf)^2))
13   #Calculamos la prevision inferior mediante la Ecuacion 2.2.
14   upr_E <- Pof + delta*sqrt(n)*Sf
15   return(upr_E)
16 }
```

Otra forma de realizar el cálculo de $\overline{P}_E(f)$ es a partir de la relación de conjugación entre \underline{P}_E y \overline{P}_E dada en (1.3), sólo sería necesario realizar el siguiente cálculo:

```
1  #Cálculo de la previsión superior.
2  upr_E <- -lpr_E(-f,Po,delta)
```

A continuación vamos a usar estos programas para el cálculo de la previsión superior e inferior en un ejemplo. Para simplificar el ejemplo vamos a trabajar sobre el espacio muestral $\mathcal{X} = \{x_1, x_2, x_3\}$ de cardinal tres, que además permitirá realizar una representación gráfica. Vamos a tomar como ejemplo el parámetro de distorsión $\delta = 0.1$, la probabilidad $P_0 = (0.3, 0.2, 0.5)$ y la

2.3. IMPLEMENTACIÓN

apuesta dada por $f = 2I_{\{x_1\}} + 6I_{\{x_2\}} + 17I_{\{x_3\}}$ que introduciremos en el programa en forma de vector $(2, 6, 17)$. Utilizando las siguientes instrucciones:

```

1  Po<-c(0.3,0.2,0.5) # Probabilidad P_0
2  delta<-0.1         # Parámetro de distorsión
3  f<-c(2,6,17)      # Apuesta f
4  lpr_E(f,Po,delta) # Previsión inferior
5  upr_E(f,Po,delta) # Previsión superior

```

se obtienen los siguientes valores para la previsión inferior y superior:

$$\underline{P}(f) = 9.201; \quad \overline{P}(f) = 11.398.$$

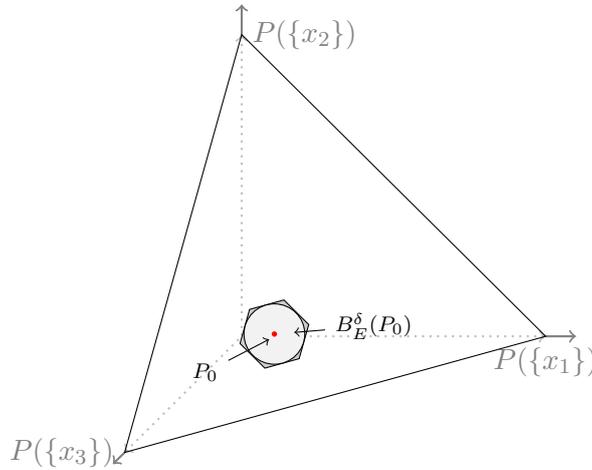


Figura 2.1: Representación gráfica de la bola $B_E^\delta(P_0)$ junto con el conjunto credal determinado por la restricción a sucesos.

En la Figura 2.1 se puede ver la representación gráfica de la bola $B_E^\delta(P_0)$ en el 3-símplex, es decir, en el espacio $\{(x, y, z) \mid x+y+z = 1, x, y, z \in [0, 1]\}$ que contiene todas las distribuciones de probabilidad en un espacio de 3-elementos. En este gráfico se puede ver que, como era de esperar, la bola es justamente un círculo centrado en P_0 . Además, su restricción a sucesos determina el conjunto credal:

$$\{P \in \mathbb{P}(\mathcal{X}) \mid P(A) \geq \underline{P}_E(A) \quad \forall A \subseteq \mathcal{X}\}.$$

Este, al ser 2-monótono, no coincide con la bola $B_E^\delta(P_0)$, si no que es una aproximación exterior.

Por último, parece claro que a medida que aumente el parámetro de distorsión, mayor será el radio de la bola. Esto puede verse en la Figura 2.2, donde se ha representado la bola $B_E^\delta(P_0)$ para distintos valores de δ , partiendo del inicial $\delta = 0.1$ y considerando los valores $\delta = 0.14$ (rojo), $\delta = 0.18$ (azul) y $\delta = 0.22$ (naranja).

2.3. IMPLEMENTACIÓN

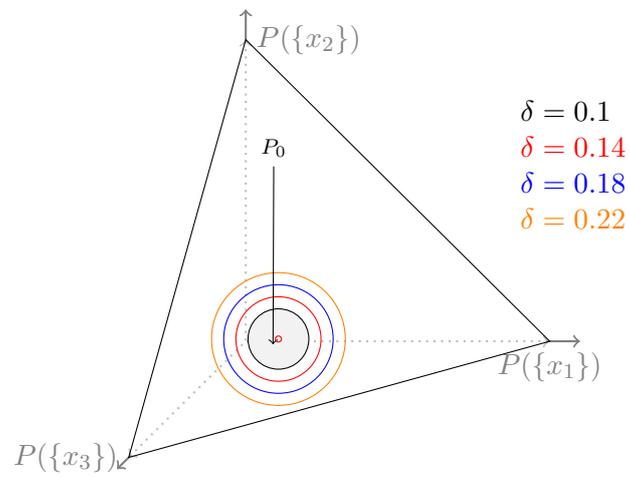


Figura 2.2: Representación gráfica de la bola $B_E^\delta(P_0)$ para distintos valores del parámetro de distorsión.

Capítulo 3

Modelo de distorsión generado por la divergencia de Kullback-Leibler

En este capítulo vamos a construir y estudiar el modelo de distorsión asociado a la divergencia de Kullback-Leibler. El estudio de este modelo no es tan extenso como el visto en el anterior capítulo para la distancia Euclídea ya que, como veremos a continuación, el cálculo de fórmulas explícitas para este modelo resulta muy complejo.

3.1. Definición del modelo

En esta sección vamos a definir el modelo de distorsión generado por la divergencia de Kullback-Leibler (KL) y vamos a estudiar algunas de sus propiedades.

La divergencia Kullback-Leibler para dos probabilidades $P, Q \in \mathbb{P}(\mathcal{X})$ viene dada por:

$$D_{KL}(P, Q) = \sum_{i=1}^n p_i (\log(p_i) - \log(q_i)) = \sum_{i=1}^n p_i \log\left(\frac{p_i}{q_i}\right). \quad (3.1)$$

Se trata de una medida de similitud entre dos probabilidades y viene dada por el promedio ponderado de la diferencia logarítmica entre las probabilidades P y Q . La divergencia de Kullback-Leibler también es conocida como la divergencia de la información por su importancia en la teoría de la información. Esta medida nos permite conocer la diferencia o similitud entre dos funciones de distribución.

Vamos a considerar una probabilidad inicial $P_0 \in \mathbb{P}^*(\mathcal{X})$ y un parámetro de distorsión $\delta \in (0, 1)$. Definimos la bola cerrada centrada en P_0 con

3.1. DEFINICIÓN DEL MODELO

parámetro de distorsión δ respecto a la divergencia (KL) como:

$$\begin{aligned} B_{KL}^\delta(P_0) &= \{P \in \mathbb{P}(\mathcal{X}) \mid D_{KL}(P, P_0) \leq \delta\} \\ &= \left\{ P \in \mathbb{P}(\mathcal{X}) \mid \sum_{i=1}^n p_i \log \left(\frac{p_i}{p_i^0} \right) \leq \delta \right\}. \end{aligned} \quad (3.2)$$

Usaremos a lo largo de este capítulo que para $k \in (0, 1)$ se tiene que $\log(kx) < \log(x)$ ya que logaritmo es una función creciente y $kx < x$ para todo $k \in (0, 1)$. También usaremos la desigualdad de la suma de logaritmos que viene dada por:

$$\left(\sum_{i=1}^n a_i \right) \log \left(\frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i} \right) \leq \sum_{i=1}^n a_i \log \left(\frac{a_i}{b_i} \right).$$

A continuación, vamos a estudiar qué axiomas de los introducidos en la Sección 1.3 son verificados por la divergencia de Kullback-Leibler.

Proposición 3.1. *La divergencia de Kullback-Leibler satisface los axiomas: 1, 4 y 5.*

Demostración. Veamos que D_{KL} es no negativa, es decir, $D_{KL}(P, Q) \geq 0$ para todo $P, Q \in \mathbb{P}(\mathcal{X})$. Para demostrar esta propiedad haremos uso de la desigualdad de Gibbs que nos dice que:

$$-\sum_{i=1}^n p_i \log(p_i) \leq -\sum_{i=1}^n p_i \log(q_i)$$

o equivalentemente:

$$\sum_{i=1}^n p_i \log(q_i) \leq \sum_{i=1}^n p_i \log(p_i). \quad (3.3)$$

De esta forma, usando la Ecuación (3.3), se tiene que:

$$\begin{aligned} D_{KL}(P, Q) &= \sum_{i=1}^n p_i \log \left(\frac{p_i}{q_i} \right) = \sum_{i=1}^n p_i (\log(p_i) - \log(q_i)) \\ &= \sum_{i=1}^n p_i \log(p_i) - \sum_{i=1}^n p_i \log(q_i) \\ &\geq \sum_{i=1}^n p_i \log(p_i) - \sum_{i=1}^n p_i \log(p_i) = 0. \end{aligned}$$

Veamos ahora que verifica el axioma 4 de convexidad, es decir, veamos que la divergencia de Kullback-Leibler verifica que:

$$D_{KL}(\alpha P + (1 - \alpha)Q, H) \leq \alpha D_{KL}(P, H) + (1 - \alpha)D_{KL}(Q, H)$$

3.1. DEFINICIÓN DEL MODELO

para todo $\alpha \in (0, 1)$.

Para estudiar la convexidad, consideremos la función $g_i : (0, 1) \rightarrow \mathbb{R}$, que viene dada para cada $i = 1, \dots, n$ por:

$$g_i(p_i) = p_i \log \left(\frac{p_i}{p_i^0} \right)$$

Veamos que la función g_i es convexa mediante el signo de su segunda derivada:

$$\begin{aligned} g_i'(p_i) &= \log \left(\frac{p_i}{p_i^0} \right) + p_i \frac{p_i^0}{p_i} = \log \left(\frac{p_i}{p_i^0} \right) + p_i^0. \\ g_i''(p_i) &= \frac{p_i^0}{p_i} > 0. \end{aligned}$$

Por tanto, la función g_i es convexa. Usando la convexidad de esta función, se tiene que:

$$\begin{aligned} D_{KL}(\alpha P + (1 - \alpha)Q, H) &= \sum_{i=1}^n (\alpha p_i + (1 - \alpha)q_i) \log \left(\frac{\alpha p_i + (1 - \alpha)q_i}{h_i} \right) \\ &\leq \sum_{i=1}^n \alpha p_i \log \left(\frac{p_i}{h_i} \right) + \sum_{i=1}^n (1 - \alpha)q_i \log \left(\frac{p_i}{h_i} \right) \\ &= \alpha D_{KL}(P, H) + (1 - \alpha)D_{KL}(Q, H) \\ &\leq \max\{D_{KL}(P, H), D_{KL}(Q, H)\} \end{aligned}$$

y por tanto, la divergencia de Kullback-Leibler cumple el axioma 4 de convexidad. Veamos ahora que también es continua, es decir, que verifica el axioma 5.

La función g_i ya definida es una función continua en $(0, 1)$, además es conocido que la función $\log(x)$ es continua en todo su dominio de definición $(0, +\infty)$, por lo tanto, como $p_i, p_i^0 \in (0, 1)$ se tiene que $\frac{p_i}{p_i^0} \in (0, +\infty)$. Por todo esto se tiene que la función $f(p_i) = \log \left(\frac{p_i}{p_i^0} \right)$ es continua para todo $p_i \in (0, 1)$.

Por otro lado, definimos la función $h(x) = p_i x$ que es continua en \mathbb{R} , en particular es continua en $(0, 1)$. Tomando la composición de las funciones continuas f y h obtenemos una función continua que verifica:

$$h \circ f(p_i) = p_i \log \left(\frac{p_i}{p_i^0} \right) = g_i(p_i).$$

Por todo esto, la función g_i resulta ser continua en $(0, 1)$ para todo $i = 1, \dots, n$, y por tanto, la suma de todas estas funciones, que coincide con la divergencia de Kullback-Leibler, también es una función continua. \square

3.1. DEFINICIÓN DEL MODELO

Acabamos de demostrar que la divergencia de Kullback-Leibler es continua y convexa, de acuerdo con [6, Prop.3.1] resulta que la bola $B_{KL}^\delta(P_0)$ es cerrada y convexa, es decir, se trata de un conjunto credal y por tanto coincide con $\mathcal{M}(\underline{P}_{KL})$, donde \underline{P}_{KL} es la previsión inferior coherente definida como la envolvente inferior de $B_{KL}^\delta(P_0)$:

$$\underline{P}_{KL}(f) = \min\{P(f) \mid P \in B_{KL}^\delta(P_0)\} = \min\{P(f) \mid P \in D_{KL}(P, P_0) \leq \delta\}$$

para cada apuesta $f \in \mathcal{L}(\mathcal{X})$.

Veamos ahora que la divergencia de Kullback-Leibler no verifica las propiedades de simetría y de desigualdad triangular, para ellos vamos a ver a continuación dos contraejemplos.

La divergencia de Kullback-Leibler no es simétrica, consideremos dos probabilidades P y Q dadas por $P = (0.5, 0.5)$ y $Q = (0.2, 0.8)$ y veamos que $D_{KL}(P, Q) \neq D_{KL}(Q, P)$:

$$D_{KL}(P, Q) = \sum_{i=1}^n p_i \log \left(\frac{p_i}{q_i} \right) = 0.5 \log \left(\frac{0.5}{0.2} \right) + 0.5 \log \left(\frac{0.5}{0.8} \right) = 0.0968$$

$$D_{KL}(Q, P) = \sum_{i=1}^n q_i \log \left(\frac{q_i}{p_i} \right) = 0.2 \log \left(\frac{0.2}{0.5} \right) + 0.8 \log \left(\frac{0.8}{0.5} \right) = 0.0837$$

por lo tanto, la divergencia de Kullback-Leibler no es simétrica.

Veamos un contraejemplo para demostrar que en general la divergencia de Kullback-Leibler no verifica la desigualdad triangular, es decir, no verifica que:

$$D_{KL}(P, Q) \leq D_{KL}(P, R) + D_{KL}(R, Q).$$

Sea $\mathcal{X} = \{x_1, x_2\}$ tal que tenemos tres probabilidades P , Q y R sobre este espacio dadas por:

A	$P(A)$	$Q(A)$	$R(A)$
$\{x_1\}$	0.5	0.2	0.3
$\{x_2\}$	0.5	0.8	0.7

Con estas probabilidades se tiene usando la Ecuación (3.1) que:

$$D_{KL}(P, Q) = 0.5 \log \left(\frac{0.5}{0.2} \right) + 0.5 \log \left(\frac{0.5}{0.8} \right) \approx 0.09691$$

$$D_{KL}(P, R) = 0.5 \log \left(\frac{0.5}{0.3} \right) + 0.5 \log \left(\frac{0.5}{0.7} \right) \approx 0.03786$$

$$D_{KL}(R, Q) = 0.3 \log \left(\frac{0.3}{0.2} \right) + 0.7 \log \left(\frac{0.7}{0.8} \right) \approx 0.01224$$

3.1. DEFINICIÓN DEL MODELO

por lo que resulta evidente que se verifica:

$$D_{KL}(P, Q) > D_{KL}(P, R) + D_{KL}(R, Q).$$

Por lo tanto, la divergencia de Kullback-Leibler no es una distancia.

Veamos que los puntos extremos del conjunto credal $\mathcal{M}(\underline{P}_{KL})$ pertenecen a la frontera de la bola $B_{KL}^\delta(P_0)$.

Proposición 3.2. *Sea $\delta > 0$ y $P_0 \in \mathbb{P}^*(\mathcal{X})$ de forma que $B_{KL}^\delta(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$, y sea f una apuesta no constante. Entonces, si para algún $P \in B_{KL}^\delta(P_0)$ se cumple que $\underline{P}_{KL}(f) = P(f)$ para una apuesta cualquiera $f \in \mathcal{L}(\mathcal{X})$, entonces se tiene que $d_{KL}(P, P_0) = \delta$.*

Demostración. Por reducción al absurdo vamos a suponer que $D_{KL}(P, P_0) < \delta$ para algún $P \in B_{KL}^\delta(P_0)$.

Veamos que existe una probabilidad $P^* \in B_{KL}^\delta(P_0)$ de tal forma que $P^*(f) < \underline{P}_{KL}(f)$ para una apuesta cualquiera $f \in \mathcal{L}(\mathcal{X})$, esto nos llevará a una contradicción ya que por definición de previsión inferior ha de ser la mínima de las previsiones.

Sean la probabilidad $P = (p_1, \dots, p_n)$ y una apuesta cualquiera dada por $f(x) = \sum_{i=1}^n a_i I_{\{x_i\}}(x)$ que sin pérdida de generalidad podemos suponer que $a_1 \leq a_2 \leq \dots \leq a_n$. La previsión de esta apuesta f viene dada por $P(f) = \sum_{i=1}^n a_i p_i$.

Sea $\mu > 0$ tal que:

$$\mu < \min\{1 - p_1, p_n\}.$$

Como $P \in B_{KL}^\delta(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$, se tiene que $1 - p_1 > 0$ y que $p_n > 0$.

Definimos la probabilidad P^* como:

$$P^*(\{x_i\}) = p_i^* = \begin{cases} p_1 + \mu & \text{si } i = 1. \\ p_i & \text{si } i \neq 1, 2. \\ p_n - \mu & \text{si } i = n. \end{cases}$$

Veamos que P^* es una probabilidad, es decir, que p_1^*, \dots, p_n^* son no negativos y que su suma es 1. Está claro que, por la definición de μ , $p_i^* > 0$ para todo $i = 1, \dots, n$ y además también se tiene que $p_1 + \mu < 1$ puesto que $\mu < 1 - p_1$.

Veamos ahora que su suma es 1:

$$\begin{aligned} \sum_{i=1}^n p_i^* &= p_1^* + \sum_{i=2}^{n-1} p_i^* + p_n^* = p_1 + \mu + \sum_{i=2}^{n-1} p_i + p_n - \mu \\ &= \sum_{i=1}^n p_i + \mu - \mu = \sum_{i=1}^n p_i = 1. \end{aligned}$$

3.1. DEFINICIÓN DEL MODELO

A continuación vamos a comprobar que P^* verifica $P^*(f) < \underline{P}_{KL}(f)$:

$$\begin{aligned} P^*(f) &= \sum_{i=1}^n a_i p_i^* = a_1 p_1^* + \sum_{i=2}^{n-1} a_i p_i^* + a_n p_n^* \\ &= a_1(p_1 + \mu) + \sum_{i=2}^{n-1} a_i p_i + a_n(p_n - \mu) = \sum_{i=1}^n a_i p_i + \mu(a_1 - a_n) \end{aligned}$$

pero resulta que $\mu > 0$ y $a_1 < a_n$ por hipótesis, y por lo tanto $\mu(a_1 - a_n) < 0$, entonces:

$$P^*(f) = \sum_{i=1}^n a_i p_i + \mu(a_1 - a_n) < \sum_{i=1}^n a_i p_i = P(f) = \underline{P}_{KL}(f).$$

Vamos a demostrar que $P^* \in B_{KL}^\delta(P_0)$, es decir, que $D_{KL}(P^*, P_0) \leq \delta$. Para ello vamos a usar una propiedad ya demostrada en la Proposición 3.1 y es que en general la función dada por:

$$g_i(x) = x \log \left(\frac{x}{p_i^0} \right)$$

es continua en el intervalo $(0,1)$.

Por definición de continuidad de la función g_1 se tiene que para todo $\epsilon > 0$, existe $\delta_1 > 0$ tal que si $|p_1^* - p_1| < \delta_1$ entonces $|g(p_1^*) - g(p_1)| < \epsilon$, es decir, si $|p_1 + \mu - p_1| = \mu < \delta_1$ entonces $|g(p_1 + \mu) - g(p_1)| < \epsilon$. Por lo tanto se tiene que:

$$g_1(p_1 + \mu) < \epsilon + g_1(p_1). \quad (3.4)$$

Lo mismo ocurre sobre la función g_n , se tiene que para todo $\epsilon > 0$, existe $\delta_n > 0$ tal que si $|p_n^* - p_n| < \delta_n$ entonces $|g(p_n^*) - g(p_n)| < \epsilon$, es decir, si $|p_n - \mu - p_n| = \mu < \delta_n$ entonces $|g(p_n - \mu) - g(p_n)| < \epsilon$. Por lo tanto se tiene que:

$$g_n(p_n - \mu) < \epsilon + g_n(p_n). \quad (3.5)$$

Tomemos $\epsilon = \frac{\delta - D_{KL}(P, P_0)}{2}$ y los δ_1, δ_n definidos. Con todo esto, si tomamos μ suficientemente pequeño, es decir, si μ cumple no sólo $\mu < \min\{1 - p_1, p_n\}$ sino que también cumple que $\mu < \min\{1 - p_1, p_n, \delta_1, \delta_n\}$ vamos a comprobar

3.1. DEFINICIÓN DEL MODELO

que $D_{KL}(P^*, P_0) \leq \delta$:

$$\begin{aligned}
D_{KL}(P^*, P_0) &= \sum_{i=1}^n p_i^* \log \left(\frac{p_i^*}{p_i^0} \right) = \sum_{i=1}^n g_i(p_i^*) \\
&= g_1(p_1^*) + \sum_{i=2}^{n-1} g_i(p_i^*) + g_n(p_n^*) \\
&= g_1(p_1 + \mu) + \sum_{i=2}^{n-1} g_i(p_i) + g_n(p_n - \mu) \\
&< g_1(p_1) + \epsilon + \sum_{i=2}^{n-1} g_i(p_i) + g_n(p_n) + \epsilon = \sum_{i=1}^n g_i(p_i) + 2\epsilon \\
&= D_{KL}(P, P_0) + 2 \frac{\delta - D_{KL}(P, P_0)}{2} = \delta.
\end{aligned}$$

Hemos encontrado una probabilidad P^* en la bola $B_{KL}^\delta(P_0)$ tal que $P^*(f) < \underline{P}_{KL}(f)$. Llegamos a contradicción al suponer que $D_{KL}(P, P_0) < \delta$ y por tanto $D_{KL}(P, P_0) = \delta$. \square

La proposición que acabamos de demostrar nos garantiza que si para algún $P \in B_{KL}^\delta(P_0)$ se tiene que $P(f) = \underline{P}_{KL}(f)$ para alguna apuesta $f \in \mathcal{L}(\mathcal{X})$ entonces se tiene que $D_{KL}(P, P_0) = \delta$, es decir, $P \in fr(B_{KL}^\delta(P_0))$.

A continuación, vamos a plantear el cálculo de $\underline{P}_{KL}(f)$ y $\overline{P}_{KL}(f)$ como un problema de optimización, y demostraremos por el Teorema de Weierstrass que el problema tiene solución.

Proposición 3.3. *Sea $P_0 \in \mathbb{P}^*(\mathcal{X})$ y $\delta > 0$ de forma que $B_{KL}^\delta(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$, y sea f una apuesta no constante dada por $f(x) = \sum_{i=1}^n a_i I_{\{x_i\}}(x)$. Entonces, la previsión $\underline{P}_{KL}(f)$ viene dada por la solución al siguiente problema de optimización:*

$$\begin{aligned}
\underline{P}_{KL}(f) &= \min_{\alpha_1, \dots, \alpha_n} \sum_{i=1}^n a_i p_i^0 \alpha_i \\
\text{s.a.} &: \sum_{i=1}^n p_i^0 \alpha_i = 1. \\
&\sum_{i=1}^n p_i^0 \alpha_i \log(\alpha_i) = \delta.
\end{aligned}$$

Demostración. En general, la previsión inferior \underline{P}_{KL} viene dada por:

$$\underline{P}_{KL}(f) = \min\{P(f) \mid P \in B_{KL}^\delta(P_0)\}$$

donde f es una apuesta cualquiera dada en general por la expresión $f(x) = \sum_{i=1}^n a_i I_{\{x_i\}}(x)$. Su previsión viene dada por $P(f) = \sum_{i=1}^n a_i p_i$.

3.1. DEFINICIÓN DEL MODELO

Se tiene que toda $P \in B_{KL}^\delta(P_0)$ puede ser expresada como $P(\{x_i\}) = P_0(\{x_i\})\alpha_i = p_i^0\alpha_i$ de forma que se debe verificar:

$$\sum_{i=1}^n p_i = 1 \Rightarrow \sum_{i=1}^n p_i^0\alpha_i = 1.$$

$$D_{KL}(P, P_0) = \delta \Rightarrow \sum_{i=1}^n p_i \log\left(\frac{p_i}{p_i^0}\right) = \sum_{i=1}^n p_i^0\alpha_i \log(\alpha_i) = \delta.$$

Es decir, los valores α_i deben verificar que:

$$\sum_{i=1}^n p_i^0\alpha_i = 1; \quad \sum_{i=1}^n p_i^0\alpha_i \log(\alpha_i) = \delta. \quad (3.6)$$

Entonces, se tiene que:

$$\underline{P}_{KL}(f) = \min_{P \in B_{KL}^\delta(P_0)} P(f) = \min_{P \in B_{KL}^\delta(P_0)} \sum_{i=1}^n a_i p_i = \min_{\alpha_1, \dots, \alpha_n} \sum_{i=1}^n a_i p_i^0 \alpha_i$$

sujeto a las condiciones dadas en la Ecuación (3.6).

El Teorema de Weierstrass afirma que si $h(x)$ es una función continua definida sobre un conjunto compacto $K \subseteq \mathbb{R}^n$, entonces el problema de optimización dado por:

$$\begin{aligned} \text{Optimizar : } & h(x) \\ \text{s.a : } & x \in K \end{aligned}$$

tiene al menos una solución. Además, si tanto la función objetivo como las restricciones son convexas, se trata de un problema de optimización convexa. Si además tanto la función objetivo es estrictamente convexa, podemos asegurar que esta solución es única.

Aplicado a nuestro problema, tenemos una función $h(\alpha_1, \dots, \alpha_n)$ que es continua sobre un conjunto compacto $(\alpha_1, \dots, \alpha_n) \in [0, 1]^n$ y también es estrictamente convexa. Como además las restricciones son convexas, nuestro problema de optimización dado por:

$$\begin{aligned} \min_{\alpha_1, \dots, \alpha_n} & \sum_{i=1}^n a_i p_i^0 \alpha_i \\ \text{s.a : } & \sum_{i=1}^n p_i^0 \alpha_i = 1. \\ & \sum_{i=1}^n p_i^0 \alpha_i \log(\alpha_i) = \delta. \end{aligned}$$

tiene una única solución. □

Al contrario que con el modelo Euclídeo, es una tarea realmente difícil encontrar una expresión para las previsiones superiores e inferiores del modelo de Kullback-Leibler, es por ello que en la demostración anterior no se ha realizado la búsqueda de esta expresión y solo se ha demostrado la existencia de la misma.

3.2. Propiedades

En esta sección vamos a estudiar algunas propiedades importantes sobre el modelo de distorsión asociado a la divergencia de Kullback-Leibler, concretamente estudiaremos si este modelo verifica la propiedad de 2-monotonía y demostraremos que el conjunto de puntos extremos del conjunto credal asociado al modelo es igual a la frontera de $B_{KL}^\delta(P_0)$.

Proposición 3.4. *Sea $P_0 \in \mathbb{P}^*(\mathcal{X})$ y $\delta > 0$ de forma que $B_{KL}^\delta(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$. Entonces:*

$$\text{ext}(\mathcal{M}(\underline{P}_{KL})) = \text{fr}(B_{KL}^\delta(P_0)).$$

Demostración. Ya hemos visto anteriormente que si $P \in \text{ext}(\mathcal{M}(\underline{P}_{KL}))$ entonces se tiene que $D_{KL}(P, P_0) = \delta$, es decir, $P \in \text{fr}(B_{KL}^\delta(P_0))$.

Veamos ahora que se cumple el recíproco, sea $P \in \text{fr}(B_{KL}^\delta(P_0))$, es decir, sea P tal que verifica que $D_{KL}(P, P_0) = \delta$ y sea la apuesta $f \in \mathcal{L}(\mathcal{X})$ dada por $f(x) = \sum_{i=1}^n a_i I_{\{x_i\}}(x)$ donde a_i viene dado por $a_i = \log\left(\frac{p_i^0}{p_i}\right)$.

Con esta apuesta f se tiene que:

$$\begin{aligned} P(f) &= \sum_{i=1}^n a_i p_i = \sum_{i=1}^n p_i \log\left(\frac{p_i^0}{p_i}\right) = \sum_{i=1}^n (\log(p_i^0) - \log(p_i)) p_i \\ &= \sum_{i=1}^n -(\log(p_i) - \log(p_i^0)) p_i = -\sum_{i=1}^n p_i \log\left(\frac{p_i}{p_i^0}\right) \\ &= -D_{KL}(P, P_0) = -\delta. \end{aligned}$$

Veamos ahora que para cualquier otra $P^* \in B_{KL}^\delta(P_0)$ tal que verifique que $P^*(f) = \underline{P}_{KL}(f)$ para la apuesta f se tiene que $P^* = P$.

Toda probabilidad $P^* \in B_{KL}^\delta(P_0)$ puede ser expresada como $p_i^* = p_i^0 \alpha_i$ para todo $i = 1, \dots, n$ y por tanto se tiene que la probabilidad de f viene dada por:

$$P^*(f) = \sum_{i=1}^n p_i^* a_i = \sum_{i=1}^n p_i^* \log\left(\frac{p_i^0}{p_i}\right) = \sum_{i=1}^n p_i^* \log(p_i^0) - \sum_{i=1}^n p_i^* \log(p_i).$$

3.3. IMPLEMENTACIÓN

Sumando y restando $\sum_{i=1}^n p_i^* \log(p_i^*)$ se obtiene que:

$$\begin{aligned}
 P^*(f) &= \sum_{i=1}^n p_i^* \log(p_i^0) - \sum_{i=1}^n p_i^* \log(p_i) + \sum_{i=1}^n p_i^* \log(p_i^*) - \sum_{i=1}^n p_i^* \log(p_i^*) \\
 &= \sum_{i=1}^n p_i^* (\log(p_i^*) - \log(p_i)) - \sum_{i=1}^n p_i^* (\log(p_i^*) - \log(p_i^0)) \\
 &= \sum_{i=1}^n p_i^* \log\left(\frac{p_i^*}{p_i}\right) - \sum_{i=1}^n p_i^* \log\left(\frac{p_i^*}{p_i^0}\right) \\
 &= D_{KL}(P^*, P) - D_{KL}(P^*, P_0).
 \end{aligned}$$

Como $D_{KL}(P^*, P) \geq 0$ se tiene que:

$$P^*(f) = D_{KL}(P^*, P) - D_{KL}(P^*, P_0) \geq -D_{KL}(P^*, P_0) \geq -\delta = P(f).$$

Acabamos de ver que $P^*(f) \geq P(f)$ con $f \in \mathcal{L}(\mathcal{X})$. Además, se tiene que $P^*(f) = -\delta = P(f)$ si y solo si $D_{KL}(P^*, P) = 0$, pero esto solo ocurre si $P^* = P$. Por lo tanto P es la única probabilidad que alcanza \underline{P}_{KL} , entonces se tiene que $P \in \text{ext}(\mathcal{M}(\underline{P}_{KL}))$. \square

Veamos que la previsión inferior \underline{P}_{KL} no es 2-monótona.

Proposición 3.5. *Sea $P_0 \in \mathbb{P}^*(\mathcal{X})$, $\delta > 0$ de forma que $B_{KL}^\delta(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$. Entonces la previsión inferior \underline{P}_{KL} no es 2-monótona.*

Demostración. Podemos verlo como una consecuencia directa de la proposición 2.5 y del hecho de que si \underline{P}_{KL} fuese 2-monótona entonces $B_{KL}^\delta(P_0)$ tendría a lo sumo $n!$ puntos extremos.

La Proposición 2.5 nos dice que el número de puntos extremos del conjunto $B_{KL}^\delta(P_0)$ es igual al número de puntos en la frontera de la bola $B_{KL}^\delta(P_0)$, pero sabemos que existen infinitos puntos en la frontera de la bola por lo que el número de puntos extremos va a ser mayor que $n!$, y por tanto \underline{P}_{KL} no puede ser 2-monótona. \square

Al igual que en el modelo Euclídeo, ocurre que, en el modelo de Kullback-Leibler \underline{P}_{KL} no es 2-monótona en general. No es una tarea sencilla demostrar la 2-monotonía en la restricción a sucesos, es por ello que dejamos en este trabajo como tarea abierta el demostrar si la restricción a sucesos de \underline{P}_{KL} verifica, o no, la propiedad de 2-monotonía.

3.3. Implementación

A continuación, vamos a ver un ejemplo de cálculo de la previsión inferior para el modelo de Kullback-Leibler. El cálculo de esta previsión vamos a realizarlo mediante un programa desarrollado en el software **R** y en el entorno

3.3. IMPLEMENTACIÓN

de trabajo RStudio. Para el correcto uso del siguiente programa es necesario instalar el paquete `nloptr` mediante el siguiente código:

```
1  install.packages("nloptr")
2  library(nloptr)
```

El programa que propongo realiza el cálculo de la previsión inferior de una apuesta $f \in \mathcal{L}(\mathcal{X})$ dada por $f(x) = \sum_{i=1}^n a_i I_{\{x_i\}}(x)$ a partir de la probabilidad inicial P_0 y el parámetro de distorsión $\delta > 0$, mediante la optimización del problema planteado en la Proposición 3.3:

```
1  lpr_KL <- function(f,p0,delta){
2    #f es el vector fila que contiene los elementos a_i de la
      apuesta f(x)= \sum_{i=1}^n a_i I_{\{x_i\}}(x).
3    #po es el vector fila que contiene la probabilidad Po de cada
      {x_i}.
4    #delta es el parametro de distorsion.
5    n<-length(p0) # Tamaño del espacio muestral
6    eval_f0 <- function(p){
7      return( sum(f*p) ) # Funcion Objetivo
8    }
9    eval_g_ineq <- function (p) {
10     constr <- c(sum(p*log(p/p0))-delta) # Condición de desigualdad
11     return (constr)
12   }
13   eval_g_eq <- function(p){
14     return ( sum(p)-1 ) # Condición de igualdad
15   }
16   lb <- rep(0,n) # Cota inferior
17   ub <- rep(1,n) # Cota superior
18   local_opts <- list( "algorithm" = "NLOPT_LD_MMA", "xtol_rel" = 1.0
19     e-15 )
20   opts <- list( "algorithm"= "NLOPT_GN_ISRES",
21     "xtol_rel"= 1.0e-15,
22     "maxeval"= 160000,
23     "local_opts" = local_opts,
24     "print_level" = 0 )
25   resul <- nloptr ( x0 = p0,
26     eval_f = eval_f0,
27     lb = lb,
28     ub = ub,
29     eval_g_ineq = eval_g_ineq,
30     eval_g_eq = eval_g_eq,
31     opts = opts
32   )
33   return(resul$objective)
```

Para realizar el cálculo de $\overline{P}_{KL}(f)$ solo es necesario hacer uso de la relación de conjugación entre \underline{P}_{KL} y \overline{P}_{KL} dada en (1.3), únicamente sería necesario realizar el siguiente cálculo:

```
1  #Cálculo de la previsión superior.
2  upr_KL <- -lpr_KL(-f,Po,delta)
```

Sobre el espacio muestral $\mathcal{X} = \{x_1, x_2, x_3\}$ de cardinal tres vamos a tomar el mismo ejemplo que en la Sección 2.3 dado por el parámetro de distorsión $\delta = 0.1$, la probabilidad $P_0 = (0.3, 0.2, 0.5)$ y la apuesta dada por $f = 2I_{\{x_1\}} + 6I_{\{x_2\}} + 17I_{\{x_3\}}$. Utilizando las siguientes instrucciones:

3.3. IMPLEMENTACIÓN

```

1  Po<-c(0.3,0.2,0.5) # Probabilidad P_0
2  delta<-0.1 # Parámetro de distorsión
3  f<-c(2,6,17) # Apuesta
4  lpr_KL(f,Po,delta) # Previsión inferior
5  upr_KL <- -lpr_KL(-f,Po,delta) #Previsión superior

```

se obtiene el siguiente valor para la previsión inferior:

$$\underline{P}_{KL}(f) = 7.26201. \overline{P}_{KL}(f) = 13.2838.$$

La representación gráfica de la bola $B_{KL}^\delta(P_0)$ puede verse en la Figura 3.1. En ella se puede ver que la forma que tiene la bola es claramente distinta que la bola $B_E^\delta(P_0)$ del modelo Euclídeo explicado en el capítulo previo, marcada en la figura con línea discontinua.

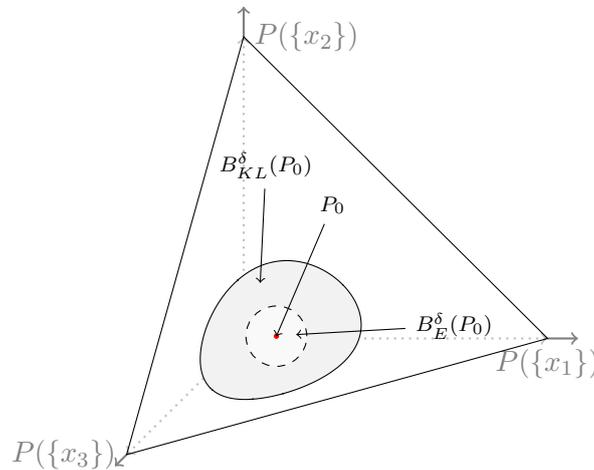


Figura 3.1: Representación gráfica de la bola $B_{KL}^\delta(P_0)$ junto con la bola $B_E^\delta(P_0)$.

Capítulo 4

Comparación de modelos de distorsión

A lo largo de este capítulo vamos a comparar los distintos modelos de distorsión según las diferentes propiedades que cumplen sus funciones de distorsión asociadas o las propiedades que verifican las probabilidades y previsiones inferiores asociadas a cada modelo.

En esta comparación usaremos los dos modelos estudiados en este trabajo, Euclídeo (*ECL*) y de Kullback-Leibler (*KL*), pero también añadiremos a la comparación los modelos de distorsión pari-mutuel (*PMM*), variación total (*TV*) y linear-vacuous (*LV*), introducidos en el Capítulo 1.3 de este trabajo.

4.1. Propiedades de las funciones de distorsión

En esta sección vamos a comparar los diferentes modelos según las propiedades que verifican las funciones de distorsión asociadas a cada modelo.

Las propiedades que estas funciones pueden cumplir son: axioma 1 (definida positiva), axioma 2 (simetría), axioma 3 (desigualdad triangular), axioma 4 (convexidad) y axioma 5 (continuidad) vistos en la Sección 1.3 de este trabajo.

Modelo	Ax. 1	Ax. 2	Ax. 3	Ax. 4	Ax. 5	Resultado
<i>PMM</i>	✓	×	×	✓	✓	[6, Prop. 6]
<i>LV</i>	✓	×	✓	✓	✓	[6, Prop. 10]
<i>TV</i>	✓	✓	✓	✓	✓	[7, Prop. 2.1]
<i>ECL</i>	✓	✓	✓	✓	✓	Prop. 2.1
<i>KL</i>	✓	×	×	✓	✓	Prop. 3.1

Cuadro 4.1: Estudio de las propiedades verificadas por las funciones de distorsión de los diferentes modelos.

4.2. PROPIEDADES DE LAS PROBABILIDADES INFERIORES COHERENTES ASOCIADAS

En la Cuadro 4.1 podemos ver todos los modelos que se comparan en este capítulo y las propiedades que verifica cada función de distorsión asociada a cada modelo, además también se facilita el resultado teórico donde se demuestra que estas propiedades se cumplen.

Las funciones de distorsión del modelo de variación total y del modelo Euclídeo son las únicas funciones que verifican los axiomas 1, 2 y 3, y por tanto, son las únicas funciones de distorsión que son distancias.

Por otro lado tenemos que para los modelos pari-mutuel, linear-vacuous y Kullback-Leibler las funciones de distorsión no son simétricas (axioma 2), además los modelos pari-mutuel y Kullback-Leibler tampoco verifican la desigualdad triangular (axioma 3). Por todo esto, se tiene que las funciones de distorsión de estos tres modelos no son distancias.

Por último, todas las funciones de distorsión son continuas y convexas (axiomas 4 y 5). El cumplimiento de estas dos propiedades es realmente importante pues nos permite saber que las bolas asociadas a cada modelo son cerradas y convexas, es decir, son conjuntos credales y por tanto la bola y su previsión superior contienen la misma información probabilística.

4.2. Propiedades de las probabilidades inferiores coherentes asociadas

Otra forma de comparar los diferentes modelos es según las propiedades que cumplen las probabilidades inferiores asociadas a cada modelo, es decir, si verifican la propiedad de 2-monotonía.

La importancia de la propiedad de 2-monotonía sobre las previsiones y probabilidades inferiores reside en que nos permite relacionar las previsiones inferiores sobre apuestas con las probabilidades inferiores sobre sucesos, de tal forma que, una previsión inferior 2-monótona y su probabilidad inferior 2-monótona resultan equivalentes.

Modelo	2-monotonía con apuestas	2-monotonía con sucesos	Resultado
<i>PMM</i>	✓	✓	[5, Proposición 5]
<i>LV</i>	✓	✓	[6, Sección 5]
<i>TV</i>	✓	✓	[7, Proposición 2.4]
<i>ECL</i>	×	✓	Proposición 2.6 y 2.9
<i>KL</i>	×	??	Proposición 3.5

Cuadro 4.2: Estudio de la propiedad de 2-monotonía en los diferentes modelos de distorsión.

Como podemos observar en el Cuadro 4.2, de los modelos que hemos estudiado en este trabajo y los que han sido estudiados en [6, 7] se tiene

4.3. COMPLEJIDAD

que tanto las previsiones inferiores como las probabilidades inferiores coherentes de los modelos pari-mutuel, linear-vacuous y de variación total son 2-monótonas.

Por otro lado, la previsión inferior del modelo Euclídeo es 2-monótona únicamente cuando consideramos su restricción a sucesos.

Por último, en el modelo de Kullback-Leibler las previsiones inferiores no verifican la propiedad de 2-monotonía para apuestas. En el caso de la 2-monotonía sobre la restricción a sucesos lo hemos marcado como ?? ya que se deja como un problema abierto para su estudio.

4.3. Complejidad

Una característica importante de los modelos de distorsión es la simplicidad de su representación, medida en términos del número de puntos extremos del conjunto credal, cuantos más puntos extremos tiene un conjunto credal consideramos que su complejidad es mayor.

Como ya hemos visto anteriormente en este trabajo, en el caso de que se cumpla la propiedad de 2-monotonía sobre las previsiones inferiores coherentes sabemos que existen como mucho $n!$ puntos extremos [6].

En el caso de los modelos pari-mutuel, linear-vacuous y de variación total los puntos extremos han sido estudiados en [5, 11], la siguiente tabla resume los resultados para estos modelos:

Modelo	Nº máximo de puntos extremos	Resultado
<i>PMM</i>	$\frac{n!}{\lceil \frac{n}{2} \rceil (\lceil \frac{n}{2} \rceil - 1)! (n - \lceil \frac{n}{2} \rceil - 1)!}$	[5, Proposición 2]
<i>LV</i>	n	[11, Sección 3.6.3(b)]
<i>TV</i>	$n(n - 1)$	[7, Proposición 2.5]

Cuadro 4.3: Número máximo de puntos extremos en los diferentes modelos de distorsión.

Por otra parte, se ha concluido en los Capítulos 2 y 3 que el número de puntos extremos de los modelos Euclídeo y de Kullback-Leibler son infinitos, incluso se ha realizado una representación del conjunto credal en ambos casos.

Podemos observar en el Cuadro 4.3 que los modelos más simples resultan ser el modelo linear-vacuous, seguido del modelo de variación total y del modelo pari-mutuel, y por tanto, los modelos más complejos en términos del número de puntos extremos son el modelo Euclídeo y el modelo de Kullback-Leibler.

A continuación, vamos a evaluar toda la información recogida a lo largo de este capítulo y la vamos a indicar en una tabla.

Indicaremos con ++ un resultado muy bueno, + para un resultado

4.3. COMPLEJIDAD

bueno, $-$ para un resultado malo, $--$ para un resultado muy malo y $?$ para un resultado sin conclusiones por falta de información.

Modelo	<i>PMM</i>	<i>LV</i>	<i>TV</i>	<i>ECL</i>	<i>KL</i>
Propiedades de la función de distorsión	-	+	++	++	-
Propiedades de la previsión inferior	+	++	-	-	?
Complejidad	+	++	++	-	-

Cuadro 4.4: Evaluación cualitativa de las propiedades estudiadas de cada modelo.

Atendiendo al Cuadro 4.4 y teniendo en cuenta todos los aspectos que hemos considerado y estudiado en este trabajo, el modelo con mejores propiedades y más estable es el modelo linear-vacuous. A pesar de que la función de distorsión que genera este modelo no es una distancia, su previsión inferior es 2-monótona y su complejidad es baja puesto que el número de puntos extremos coincide con el cardinal del espacio \mathcal{X} . Le sigue el modelo pari-mutuel cuya función de distorsión tampoco es una distancia pero su previsión inferior es 2-monótona tanto para apuestas como en su restricción a sucesos y su complejidad resulta ser mayor que en el modelo linear-vacuous. Por otro lado, el modelo de variación total es el que tiene mejores propiedades sobre la función de distorsión y su complejidad es baja, aunque sigue siendo más simple el modelo linear-vacuous.

Finalmente, tenemos los modelos que hemos estudiado en este trabajo, el modelo Euclídeo y el modelo de Kullback-Leibler. El modelo Euclídeo está generado por la distancia Euclídea, que verifica todos los axiomas propuestos en la Sección 1.3, mientras que, la divergencia de Kullback-Leibler no es una distancia ya que no verifica los axiomas de simetría y desigualdad triangular. Por otro lado, ambos modelos resultan ser muy complejos ya que el número de puntos extremos es infinito, además su previsión inferior resulta no ser 2-monótona para apuestas en ambos casos, únicamente podemos asegurar que la previsión inferior es 2-monótona para sucesos en el caso del modelo Euclídeo, en el modelo de Kullback-Leibler no disponemos de suficiente información.

Capítulo 5

Conclusiones

A lo largo de este trabajo se ha hecho el estudio de los modelos de distorsión generados por la distancia Euclídea y por la divergencia de Kullback-Leibler como complemento al estudio de diversos modelos en [6, 7] con el objetivo de poder comparar todos estos modelos entre sí.

Todos estos modelos son generados por una probabilidad inicial P_0 distorsionada mediante una función, teniendo en cuenta un factor de distorsión δ que indica la cantidad de imprecisión que se desea añadir al modelo.

En el Capítulo 2 hemos estudiado el modelo de distorsión inducido por la distancia Euclídea, este modelo cumple todos los axiomas propuestos en la Sección 1.3, en particular los axiomas de convexidad y continuidad, y por tanto la bola que define este modelo coincide con el conjunto credal del modelo. Además, también se dieron expresiones para el cálculo de la previsión inferior coherente y de la probabilidad inferior coherente de este modelo. Por otro lado, en este capítulo se demostró que los puntos extremos del conjunto credal coinciden con la frontera de la bola definida por el modelo. Finalmente, se programaron códigos para el cálculo de las previsiones superiores e inferiores y para la representación del conjunto credal.

En el Capítulo 3 se ha estudiado el modelo inducido por la divergencia de Kullback-Leibler. Este modelo presenta dificultades al tratar de dar una fórmula explícita para la previsión inferior y para la probabilidad inferior. Se ha visto que la divergencia de Kullback-Leibler no verifica las propiedades de simetría y la desigualdad triangular pero si cumple las propiedades de convexidad y continuidad, lo que nos permite asegurar que la bola definida por el modelo es cerrada y convexa, y por tanto, debe coincidir con el conjunto credal. Además, se ha planteado el cálculo de la previsión inferior mediante un problema de optimización, pero debido a la complejidad de esta función de distorsión se ha optado por únicamente demostrar que existe solución para el problema planteado y que esta solución es única. También se ha visto, al igual que en el modelo Euclídeo, que los extremos del conjunto credal coinciden con la frontera de la bola asociada al modelo. Por último,

se han programado códigos para calcular la previsión inferior y la previsión superior de este modelo mediante la optimización del problema planteado anteriormente y para la representación de su conjunto credal.

Para finalizar este estudio, en el Capítulo 4, se ha realizado una comparación de todos los modelos trabajados a lo largo de este estudio, es decir, de los modelos pari-mutuel, linear-vacuous, variación total, Euclídeo y de Kullback-Leibler. Esta comparación se ha realizado en términos de diferentes propiedades que pueden cumplir estos modelos como por ejemplo las propiedades que cumplen sus funciones de distorsión o la complejidad del modelo. En conclusión, los modelos estudiados en [6, 7] son más estables y tienen, en general, mejores propiedades que los propuestos en este trabajo.

A lo largo de este trabajo se puede encontrar un estudio de las principales propiedades de los modelos trabajados pero no se trata de un estudio exhaustivo de estos modelos por lo que se puede extender a otras propiedades que no han sido estudiadas como la propiedad de k -aditividad. Por supuesto, también es posible continuar este trabajo mediante el estudio de otros modelos generados por otras funciones de distorsión, como por ejemplo la distancia de Wasserstein [10].

Bibliografía

- [1] Choquet, G., 1953-1954. “Theory of capacities”. *Annales De L’Institut Fourier*, (5): 131-295.
- [2] de Finetti, B., 1972. “Probability, induction and statistics: The art of guessing”. New York: John Wiley.
- [3] Huber, P. J. 1981. “Robust statistics”. New York: John Wiley.
- [4] Levin, D. A., Peres, Y. and Wilmer, E. L., 2009. “Markov chains and mixing times”. Providence, RI: American Mathematical Society.
- [5] Montes, I. Miranda, E. and Destercke, S., 2019. “Pari-Mutuel probabilities as an uncertainty model”. *Information Sciences*, 481: 550-573.
- [6] Montes, I., Miranda, E. and Destercke, S., 2020 “Neighbourhood and distortion models. Part I: New results on old models”. *International Journal of General Systems*, 49 (6):602-635.
- [7] Montes, I., Miranda, E. and Destercke, S., 2020. “Neighbourhood and distortion models. Part II: New models and synthesis”. *International Journal of General Systems*.
- [8] Pelessoni, R., Vicig, P. and Zaffalon, M., 2010. “Inference and risk measurement with the Pari-mutuel model”. *International Journal of Approximate Reasoning*, 51: 1145-1158.
- [9] Seidenfeld, T. and Wasserman, L., 1993 “Dilation for sets of probabilities”. *The Annals of Statistics*, 21: 1139–1154.
- [10] Vaserstein, N.L., 1969 “Markov processes over denumerable products of spaces, describing large systems of automata”. *Problemy Peredachi Informatsii*. 5(3): 64–72.
- [11] Walley, P., 1991. “Statistical reasoning with imprecise probabilities”. Chapman and Hall. London.