

Universidad de Valladolid

Facultad de Ciencias
Departamento de Estadística e
Investigación Operativa

Tesis doctoral

*Comportamiento del estimador
máximo verosímil para un parámetro
 k -dimensional en modelos con
restricciones*

Miguel Alejandro Fernández Temprano

Memoria presentada para optar al grado de Doctor en Ciencias Matemáticas, por Miguel Alejandro Fernández Temprano.

Bonifacio Salvador González, Catedrático de Universidad, y Cristina Rueda Sabater, Profesora Titular de Universidad, certifican que la presente memoria ha sido realizada, bajo su dirección, por Miguel Alejandro Fernández Temprano en el Departamento de Estadística e Investigación Operativa de la Universidad de Valladolid.

Valladolid, 12 de Julio de 1995

A mi padre

Contenido

Introducción	7
1 Estimación de Funciones Lineales del Parámetro	21
1.1 El octante positivo	27
1.2 Otros conos rectos.....	41
1.3 Conos circulares.....	47
1.3.1 Prueba del lema 1.13	56
1.4 Conos de orden.....	58
1.4.1 Tree Order.....	62
1.4.2 Orden Total	63
2 Estimación Simultánea de Coordenadas	70
2.1 Caso de un semiespacio	72
2.2 Estimación simultánea en conos de orden	78
2.2.1 Extensiones del Teorema 2.5.....	96
3 Estimadores Alternativos	101
3.1 Comparando el EMV con estimadores basados en otras métricas	104

3.2	Estimadores mixed para el Octante Positivo	112
3.2.1	Estimación global del parámetro	112
3.2.2	Estimación de funciones lineales del parámetro.....	116
3.2.3	Otro tipo de estimadores mixed en O^+	131
4	Un Modelo Uniforme de Escala	137
4.1	Tree Order	138
4.2	Orden Total	142
	Bibliografía	148

Introducción

El tema objeto de esta memoria se enmarca dentro de la Inferencia Estadística con restricciones. Esta rama de la Estadística, que apareció hacia mediados de los años 50, recoge un conjunto de procedimientos diseñados para aprovechar la información a priori que a menudo se tiene, en las aplicaciones reales, sobre los parámetros de interés en un modelo estadístico.

Uno de los tópicos más desarrollados de esta rama es el de los contrastes con restricciones para medias de poblaciones normales. El interés de este tipo de contrastes surge en muchas aplicaciones prácticas en las que se plantea la verificación de ciertas propiedades de un conjunto de medias como la homogeneidad, monotonía o unimodalidad. En este sentido se han logrado notables avances obteniéndose inicialmente la distribución del estadístico razón de verosimilitudes en distintas situaciones en referencias como Bartholomew (61), Barlow et al. (72), Shapiro (85), Raubertas et al. (86) y Robertson et al. (88) entre otros y más adelante propiedades de la función de potencia de los mismos, Perlman (69) y Mukerjee, Robertson y Wright (86). Asimismo se han detectado anomalías en este tipo de tests y caracterizado las situaciones en las que son dominados por otros que no tienen en cuenta parte de la información

adicional proporcionada por las restricciones, Menéndez y Salvador (91, 92) y Menéndez, Rueda y Salvador (91, 92a, 92b). Por otro lado se han considerado y estudiado con ciertos resultados positivos otro tipo de procedimientos alternativos a la razón de verosimilitudes como por ejemplo los tests lineales cuyo estudio aparece entre otros en Abelson y Tukey (63), Robertson et al. (88) y Rueda (89).

El presente trabajo se engloba dentro del otro de los grandes tópicos de la Inferencia con restricciones como es la estimación. Aquí trabajamos sobre todo con el Estimador de Máxima Verosimilitud, el objetivo principal será comparar el comportamiento de dicho estimador con otros estimadores propuestos y el criterio principalmente empleado será el del error cuadrático medio.

Como ocurre con los contrastes de razón de verosimilitudes, también se han detectado anomalías en el comportamiento del EMV. Si bien es cierto que el error cuadrático medio total del Estimador Máximo Verosímil (EMV) restringido es menor que el del EMV no restringido en un modelo normal (Robertson et al. (88)) esta propiedad no siempre se mantiene al estimar funcionales del parámetro. Situaciones en las que un estimador, que es globalmente mejor que otro, se ve después superado por este en la estimación de funciones lineales del parámetro no son exclusivas de problemas con restricciones. A saber, la primera referencia sobre este tipo de cuestiones en el caso

no restringido es Rao y Shinozaki (78).

En el contexto de los problemas con restricciones, el caso particular de la estimación de la i -ésima coordenada ha sido estudiado para diferentes tipos de restricciones y distribuciones. Cuando el modelo es normal y bajo las restricciones de un orden total Lee (81) prueba que cada coordenada del EMV tiene error cuadrático medio menor que la correspondiente del EMV no restringido, resultado que fue posteriormente ampliado en Kelly (89) utilizando el criterio de la probabilidad de cubrimiento. En cambio, por otra parte, Lee (88) establece que en el cono del árbol simple (tree order) la propiedad anterior no se verifica para la raíz del árbol y si, bajo determinadas condiciones, para las otras coordenadas.

Sin embargo no hay muchos resultados referidos a la estimación, individual o simultánea, de funciones lineales del parámetro que son muchas veces de interés, por ejemplo, en el modelo lineal, tampoco resultados generales que expliquen el desigual comportamiento del EMV y pocos utilizando criterios de comparación diferentes al error cuadrático medio. Tiene interés por tanto estudiar en general si los procedimientos de estimación con restricciones suponen o no una mejora respecto de los estimadores clásicos.

En lo que a los criterios de comparación se refiere el criterio más habitual de comparación entre estimadores es, por supuesto, el del error cuadrático

medio. Este criterio presenta ventajas que son sobradamente conocidas. Pero la elección de un criterio de comparación no es una cuestión baladí puesto que un estimador que supera uniformemente a otro con el criterio del error cuadrático medio puede verse uniformemente rebasado por el otro si la función de pérdida empleada no es la cuadrática sino la cuártica, como puede verse en un ejemplo que aparece en Rao y Shinozaki (78). Este tipo de situación lleva a buscar procedimientos de comparación que sean “robustos” en el sentido de que, cuando este procedimiento declara a un estimador mejor que otro no solamente sea mejor bajo una determinada función de pérdida sino bajo una clase más amplia de funciones.

Con este objeto Hwang (85) define la U-dominación o dominación universal, criterio que emplearemos varias veces en esta memoria, de la forma siguiente:

Sea $R_L(\theta, \delta) = E_\theta L(|\theta - \delta(X)|_D)$ el riesgo del estimador δ bajo la función de pérdida L y bajo el error euclídeo generalizado con respecto a la matriz semidefinida positiva D donde $|\theta - \delta(x)|_D = [(\theta - \delta(x))' D (\theta - \delta(x))]^{1/2}$. Entonces δ_1 domina universalmente a δ_2 (bajo el error euclídeo generalizado con respecto a la matriz semidefinida positiva D) si para cada θ y cada función de pérdida no decreciente L , $R_L(\theta, \delta_1) \leq R_L(\theta, \delta_2)$ y la desigualdad es estricta para alguna función de pérdida.

Hwang (85) comprueba además que este criterio es equivalente a la conocida dominación estocástica que se define como sigue:

Se dice que la variable aleatoria Y es estocásticamente menor o igual que

Z y se denota por $Y \leq_d Z$ si para cada número real positivo t se tiene que $\Pr(Y \geq t) \leq \Pr(Z \geq t)$. Entonces un estimador δ_1 domina estocásticamente δ_2 bajo el error euclídeo generalizado con respecto a la matriz semidefinida positiva D si para cada θ , $|\delta_1(X) - \theta|_D \leq_d |\delta_2(X) - \theta|_D$ y la desigualdad es estricta para algún θ .

Otro criterio que utilizaremos y que relaciona la estimación de funciones lineales del parámetro con el comportamiento global y la estimación simultánea del mismo es el definido en Lehmann (83) y que afirma que un estimador δ_1 está “ más concentrado en torno al parámetro θ ” que otro δ_2 si para cualquier combinación lineal de las componentes del parámetro $d'\theta$ el error cuadrático medio de $d'\delta_1$ es menor o igual que el de $d'\delta_2$.

El modelo general que nos proponemos estudiar y que iremos particularizando en determinadas circunstancias es el siguiente:

X_1, \dots, X_k son variables aleatorias con momento de orden 2 finito que siguen un modelo de localización donde $\theta = (\theta_1, \dots, \theta_k)$ es el parámetro de localización de dicho modelo de tal forma que $\theta_i = EX_i$ y $\theta \in C$, donde C es el cono de restricciones. En este modelo vamos a denotar por X^* y X los EMV restringido y no restringido respectivamente. El caso particular que más desarrollaremos es, por supuesto, el modelo normal.

Motivado por los resultados como los de Lee (81, 88) o Kelly (89) men-

cionados antes y la falta de resultados referidos a direcciones que no sean las coordenadas, se plantea en el capítulo I la caracterización de direcciones $d'\theta$ para las que el error cuadrático medio del estimador restringido es mayor que el del EMV no restringido en diferentes situaciones cuando C es un cono recto o circular o un cono de orden. En este capítulo juega un papel fundamental la llamada “dirección central” de un cono. El concepto de dirección central se debe a Abelson y Tukey (63) quienes la definen como aquella que minimiza el máximo ángulo que forma con las direcciones que generan el cono.

Comenzamos estudiando el problema considerando C el octante positivo en \Re^k ; es decir el parámetro θ a estimar verifica $\theta_i \geq 0$ para $i = 1, \dots, k$. En este caso particular la sencillez de los cálculos permite generalizar los resultados a modelos bastante más amplios que el modelo normal habitual. Por otro lado, el hecho de que este cono sea recto (las direcciones que lo generan son ortogonales), permite trasladar los resultados a otros conos rectos cualesquiera cuando la distribución subyacente es normal. Algunos de estos conos han sido objeto de estudio en otros trabajos, por ejemplo los de promedios crecientes o el star-shaped, que tienen bastantes aplicaciones prácticas (Shaked (79), Dykstra y Robertson (82, 83) y Robertson et al. (88)).

Los resultados fundamentales que se obtienen en relación a los conos rectos son de dos tipos. En primer lugar la influencia de la dimensión k a la hora de determinar cual de los dos estimadores es mejor, y en segundo la determinación como “dirección más desfavorable” en la estimación máximo verosímil de la dirección central del cono.

Es precisamente este último resultado relativo a las direcciones centrales el que nos lleva al estudio de los conos circulares $C(\omega, c)$ que vienen definidos precisamente a partir de su dirección central, c , que forma un ángulo ω ($0 \leq \omega \leq \pi/2$) con la generatriz del cono. Otro aspecto de interés en estos conos es que coinciden con los conos poliédricos si la dimensión del espacio, k , es 2. La referencia más completa en lo que se refiere a inferencia en espacios paramétricos restringidos a conos circulares es Akkerboom (90). Suponiendo también normalidad en la distribución subyacente demostramos que si $k \geq 4$ existen conos circulares para los que el error cuadrático medio de $c'X^*$ no es uniformemente mejor que el de $c'X$ y que esto ocurre para cualquier ángulo ω si la dimensión k es suficientemente grande. También realizamos un estudio más detallado del caso particular de interés $k = 2$.

En un estudio de este tipo es obligado tratar también los conos de orden por ser los que tienen mayor aplicación práctica. Una lista completa de referencias y resultados obtenidos hasta el momento aparece en Robertson et al. (88). Entre los conos de orden más ampliamente estudiados cabe destacar el cono del orden total, $C_{OT} = \{\theta \in \mathfrak{R}^k : \theta_1 \leq \dots \leq \theta_k\}$ y el tree order o árbol simple, $C_{ST} = \{\theta \in \mathfrak{R}^k : \theta_0 \leq \theta_i \text{ con } i = 1, \dots, k-1\}$ asociado a problemas en los que se comparan varios tratamientos con un control.

De los resultados de Lee (81, 88) y Kelly (89) parecía deducirse la existencia de alguna diferencia fundamental entre los dos conos mencionados antes. Sin embargo, nosotros probamos que, en la estimación de la dirección central, $c'\theta$, tanto en el caso del tree order como en el del orden total a partir de una

dimensión k tampoco se produce reducción de error cuadrático cuando se utiliza $\mathcal{C}'X^*$ en lugar de $\mathcal{C}'X$ como estimador. Por lo que sabemos es la primera vez que se obtiene un resultado en el que el EMV no domina al no restringido en un cono de orden total. Además utilizando valores calculados por Abelson y Tukey (63) acotamos el valor de k para el que se tiene el resultado anterior, valor que precisamos más mediante simulación.

El capítulo II está dedicado a la estimación simultánea de combinaciones lineales del parámetro en un modelo normal restringido. La estimación simultánea es fundamental en la cuestión de la estimación por intervalos de varias componentes del parámetro, es decir la construcción de regiones de confianza (Seber (77) y Scheffé (59)) y también en los procedimientos de comparación múltiples (Hochberg y Tamhane (87)). Aquí estudiamos el problema de la estimación simultánea de la media de un vector normal bajo dos tipos diferentes de restricciones y hacemos uso de diferentes criterios a la hora de comparar los EMV con y sin restricciones.

Inicialmente supondremos que el espacio paramétrico viene definido por una única restricción lineal (es decir que el parámetro pertenece a un semiespacio). Para este modelo particular contamos con algunos resultados previos respecto a la estimación individual de combinaciones lineales de θ (Rueda y Salvador (95)).

La sencillez del cono de restricciones en este caso permite obtener resultados muy fuertes respecto a los tópicos considerados. Así se demuestra que X^* es “mejor” que X para estimar θ utilizando un criterio que introducimos y que es el más fuerte de los definidos hasta ahora:

δ_1 es mejor que δ_2 si $\Pr(\delta_1 - \theta \in A) \geq \Pr(\delta_2 - \theta \in A)$ para cualquier conjunto convexo y simétrico respecto del origen A .

Como aplicaciones particulares del resultado se obtiene que los intervalos del tipo de Scheffé, Tukey o Bonferroni o incluso los más generales de Bowden (70) tienen mayor probabilidad de cubrimiento si se centran en el EMV restringido en lugar de en el no restringido. Además, según la terminología de Hwang (85) concluimos que X^* domina universalmente a X y que por tanto X es U-inadmisibile.

Por otro lado, en el contexto de la estimación simultánea, se estudia el caso particular de la estimación de las coordenadas cuando el cono de restricciones es un cono de orden. En este caso el criterio de comparación es el de la probabilidad de cubrimiento, ya que, desafortunadamente, no se pueden obtener resultados tan generales como en la situación anterior. De hecho, creemos que resultados tan generales como los de la situación anterior van a ser difícilmente generalizables a conos definidos por 2 o más restricciones. En los casos del tree order y el orden total apuntamos los contraejemplos pertinentes.

Respecto a la estimación simultánea de coordenadas obtenemos resulta-

dos del tipo

$$\Pr(|X_i^* - \theta_i| \leq t_i, i = 1, \dots, k) \geq \Pr(|X_i - \theta_i| \leq t_i, i = 1, \dots, k)$$

para distintos conos de orden, modelos generales no necesariamente de localización y determinadas configuraciones de los t_i . Además creemos que la técnica de demostración utilizada puede aplicarse para obtener más resultados de este tipo en otras situaciones. Encontramos incluso situaciones bastante generales, en lo referido a las restricciones y a la distribución subyacente, en las que se puede asegurar incluso la contención entre los conjuntos que aparecen involucrados en esa desigualdad.

A la vista de los resultados obtenidos en relación a los EMV, en el siguiente capítulo nos planteamos el estudio de otros estimadores que puedan ser mejores que estos en cierto sentido. Un tipo de estimadores alternativos que tratamos aquí son los llamados estimadores mixed, que son estimadores “intermedios” entre el EMV restringido y el no restringido y aparecen primeramente en Katz (63) en un contexto binomial con restricciones y más recientemente en Kumar y Sharma (88) para restricciones de orden total en modelos normales, o en Vijayasree y Singh (93) para dos medias exponenciales ordenadas.

Más concretamente, el capítulo III está dedicado al estudio de estimadores alternativos al EMV y a su comparación con ellos en la estimación de funciones lineales de θ en dos modelos diferentes en lo referido tanto a las

restricciones como a las distribuciones subyacentes. Por lo que respecta a las restricciones, tratamos las situaciones de una sola restricción y la del octante positivo.

En el caso de una sola restricción, retomando los resultados del capítulo anterior consideramos un modelo normal y tratamos de ver hasta que punto el EMV es bueno comparándolo con otros que son proyecciones con métricas diferentes a la dada por la inversa de la matriz de covarianzas Σ . Como justificación de este tipo de estimadores podemos decir que la proyección con otras métricas es un procedimiento similar al empleado por Lee (88) en el contexto del tree order. Obtenemos que aunque el EMV se comporta muy bien en la comparación con estos estimadores no siempre U-domina a todos ellos.

En lo que se refiere al octante positivo empleamos de nuevo, gracias a la sencillez de las restricciones, el modelo más general que describimos anteriormente y utilizamos como criterio de comparación el error cuadrático medio. Consideramos estimadores mixed o intermedios X^α del tipo de los de Kumar y Sharma (88) y los comparamos con X y X^* obteniendo resultados referidos tanto a la estimación global del parámetro θ como a la de funciones lineales $d'\theta$. El resultado más significativo es, quizá, que algunos de estos estimadores X^α siguen estando “más concentrados en torno a θ ” (Lehmann (83)) que X , para valores de k para los que esa propiedad ya no se verifica para X^* . Además, con el objeto de estudiar más a fondo el comportamiento de estos estimadores como en Kumar y Sharma (88) y Vijayasree y Singh (93), llevamos a cabo un

estudio sobre la admisibilidad de estos estimadores y otros estimadores mixed más relacionados con los de Vijayasree y Singh (93), X^β , dentro de la clase definida por ellos mismos en la estimación de funciones lineales de θ .

Por último dedicamos el capítulo IV al modelo de escala definido por k variables aleatorias independientes uniformes en $(0, \theta_i)$. El objetivo fundamental es comprobar hasta que punto los resultados obtenidos en modelos de localización pueden extenderse a otro tipo de modelos, pero hay que notar además que los modelos restringidos de escala también están recibiendo cierta atención en los últimos años. Kushary y Cohen (89) y Hwang y Peddada (94), por ejemplo, tratan modelos generales de este tipo bajo restricciones de orden obteniendo ciertos resultados relativos a la admisibilidad y dominación estocástica de las coordenadas del parámetro.

El modelo considerado en el capítulo IV aparece en Elfessi y Pal (92), que tratan el caso de dos poblaciones, y Misra y Dhariyal (95) quienes estudian la admisibilidad de los estimadores en el orden total proponiendo alternativas. Nosotros consideramos tanto restricciones de orden total como de tree order y de nuevo los resultados obtenidos son del mismo tipo que los comentados en modelos de localización, es decir en ambos casos obtenemos funciones $c'\theta$ para las que el error cuadrático medio de $c'X^*$ puede ser mayor que el de $c'X$ si k es suficientemente grande. Vemos también que, en este modelo, esto no solo

ocurre con el EMV sino que también otros estimadores restringidos que dominan al EMV se ven superados en determinadas condiciones por estimadores no restringidos.

Como conclusiones finales de todo lo comentado anteriormente podemos señalar las siguientes. En primer lugar que, a la hora de determinar si el EMV X^* es mejor o no que X no es tan importante el cono C de restricciones como la dimensión de su linealidad y la del espacio en el que está contenido. Esta consideración nos parece más próxima a la realidad que la de Robertson et al. (88) quienes justifican el hecho de que haya reducción de error en las coordenadas del orden total y no en las del tree order a partir del número de conjuntos inferiores que existen en cada uno. Hemos probado aquí, de hecho, que el comportamiento de ambos con respecto a la dirección central es similar. A la vista de estos resultados tampoco parece válida la alternativa que proponen Hwang y Peddada (94) en la que se sustituye el EMV por otro EMV más restringido.

En segundo lugar observamos que aunque el EMV con restricciones tiene menor error cuadrático medio global que el correspondiente sin restricciones, no se produce esta dominación respecto de criterios de comparación más exigentes y que solo en la situación en la que C es un semiespacio se obtiene un resultado que permite asegurar que, el EMV con restricciones, es mejor que el EMV sin

restricciones. Por tanto de nuevo la dimensión de C y de su linealidad es un factor determinante.

Quedan, por último, abiertas muchas cuestiones de interés, en particular obtener resultados respecto a la estimación simultánea de funciones en modelos generales y también encontrar estimadores mejores en algunas de las situaciones mencionadas a lo largo de este trabajo. En este sentido una técnica que creemos que podría dar buenos resultados, ya que parece que la consideración sistemática de la información adicional de que se dispone como una truncación del espacio paramétrico no es conveniente cuando la dimensión del espacio es grande, es no tratar dicha información adicional como una truncación del espacio paramétrico sino como medio de definir una probabilidad a priori sobre dicho espacio y utilizar por tanto procedimientos bayesianos de estimación.

Por último, quisiera expresar mi sincero agradecimiento a Boni y Cristina por la ayuda y el estímulo que de ellos he recibido durante la elaboración de este trabajo. También mi agradecimiento a Juanma por la realización del estilo con el que esta memoria está escrita (los errores en su uso son míos, por supuesto) y al resto de los compañeros del Departamento por el ánimo y apoyo que he recibido de ellos.

Capítulo 1

Estimación de Funciones Lineales del Parámetro

Este capítulo está dedicado a la comparación entre el EMV restringido y el no restringido en la estimación de funciones lineales de un parámetro de k -dimensional, θ bajo diferentes conos de restricciones y utilizando el criterio del error cuadrático medio. En estas comparaciones va a tener gran importancia la denominada “dirección central” del cono. Este concepto y su desarrollo se debe a Abelson y Tukey (63) quienes la definen como sigue:

Definición 1.1 *Se dice que $c \in \mathfrak{R}^k$ es la dirección central del cono $C \subset \mathfrak{R}^k$ si es la dirección cuyo máximo ángulo con el conjunto de direcciones que generan el cono es mínimo.*

Para comparar dos estimadores multidimensionales además del conocido criterio del error cuadrático medio que mide el comportamiento del estimador $\delta = (\delta_1, \dots, \delta_k)$ para estimar el parámetro $\theta = (\theta_1, \dots, \theta_k)$ a través de $\sum_{i=1}^k E(\delta_i - \theta_i)^2$, utilizaremos el siguiente criterio que Lehmann presenta en su libro *Theory of Point Estimation* (83).

Criterio Se dice que el estimador δ está más concentrado en torno a $\theta \in \mathfrak{R}^k$ que δ^* si

$$M(\theta, \delta^*) - M(\theta, \delta)$$

es una matriz semidefinida positiva, donde $M(\theta, \delta)$ es una matriz cuyo elemento (i, j) es $E[(\delta_i - \theta_i)(\delta_j - \theta_j)]$.

Otra formulación equivalente de este criterio es la siguiente: δ está más concentrado en torno a $\theta \in \mathfrak{R}^k$ que δ^* si y solamente si

$$E[d'(\delta - \theta)]^2 \leq E[d'(\delta^* - \theta)]^2$$

para cualquier $d \in \mathfrak{R}^k$, es decir si al estimar una función lineal cualquiera del parámetro $d'\theta$, el error cuadrático medio del estimador $d'\delta$ es menor que el de $d'\delta^*$.

En el contexto de la inferencia con restricciones se han encontrado situaciones en las que el EMV que utiliza la información adicional disponible no está más concentrado en torno a θ que el EMV no restringido. Lee (88) probó que en un modelo normal y cuando se sabe que el parámetro verifica las restricciones de un tree order, el error cuadrático medio en la coordenada raíz es superior cuando el EMV empleado es el restringido.

Tratamos de ver aquí si esta falta de reducción se produce también en otros contextos, es decir en otro tipo de conos e incluso en modelos más generales que el normal. Comparamos a través del error cuadrático medio el comportamiento de $X^* = P(X/C)$ y X estimador no restringido en lo que a estimación de funciones lineales del parámetro θ se refiere.

El modelo más general de los que consideraremos es el que aparece en la primera sección en la que suponemos que tenemos X_1, \dots, X_k variables aleatorias independientes tales que $E(X_i) = \theta_i$ y $Var(X_i) < \infty$ y con función de densidad $f(x_i, \theta_i) = f(x_i - \theta_i)$ (es decir provenientes de un modelo de localización cuyo parámetro es la media de la distribución). En esta sección suponemos que todas las componentes del parámetro $\theta = (\theta_1, \dots, \theta_k)$ son positivas; o dicho de otro modo sabemos que el parámetro pertenece al octante positivo de \mathfrak{R}^k . Vamos a comparar, en lo que a estimación de funciones lineales se refiere, el estimador equivariante de la familia no restringida, X , con el estimador X^* , proyección sobre el octante positivo. Hay que notar que el estimador X^* es además el Estimador Máximo Verosímil para θ en casos usuales e importantes como la familia normal o doble exponencial y en general siempre que la densidad de X sea monótona decreciente para los valores de X_i menores que la media de la variables.

El resultado principal de esta primera sección puede resumirse diciendo que, a pesar de que en el octante positivo el estimador X^* es globalmente mejor que X precisamente por serlo los correspondientes a las coordenadas individuales (i.e. $E\left(\sum_{i=1}^k (X_i^* - \theta_i)^2\right) \leq E\left(\sum_{i=1}^k (X_i - \theta_i)^2\right)$), no siempre lo es para funciones lineales del parámetro que pueden ser de interés en muchas aplicaciones. Probamos que esta situación depende de la función lineal que se estima y de la dimensión k del espacio. Además este resultado se particulariza para varias situaciones concretas como algunos modelos uniformes, el modelo doble exponencial o el modelo normal. También se demuestra que para $k = 2$

siempre es preferible desde el punto de vista del error cuadrático medio X^* a X para estimar funciones lineales independientemente de la distribución subyacente y damos un modelo en el que para $k = 3$ hay pérdida en algunas direcciones y otros en los que la reducción de error en todas las direcciones se produce hasta una dimensión k_0 arbitrariamente grande.

En cuanto a la estimación de direcciones particulares probamos que efectivamente para algunas como las coordenadas ya mencionadas o la comparación de medias es siempre preferible utilizar X^* y que en cualquier caso la dirección más desfavorable desde este punto de vista es la dirección central del cono.

En la siguiente sección se utilizan los resultados demostrados en el octante positivo para estudiar la estimación de funciones lineales del parámetro en conos rectos generales cuando la distribución subyacente de las variables es normal. Obtenemos en esta situación que el estimador X^* es mejor que X para estimar cualquier función lineal cuando $k \leq 4 + p$ donde p es la dimensión de la linealidad (mayor subespacio contenido en el cono) del cono recto que estamos considerando. Además obtenemos que se mantienen, en esta situación, resultados como el antes mencionado de la dirección diagonal del cono como situación más desfavorable.

Estos resultados se aplican a los casos particulares de los conos star-shaped y de los promedios crecientes que, como mencionan Dykstra y Robertson (83), son útiles cuando se sabe, por ejemplo, que una función de regresión sigue una tendencia genéricamente monótona pero no que cumple las restricciones de un orden total. En estas situaciones puede muchas veces utilizarse la

suposición de que la función es monótona en promedio que es menos restrictiva y permite que el orden se invierta en un rango pequeño de valores del conjunto de parámetros.

El hecho de que la dirección central del cono sea, en algunas circunstancias la dirección más desfavorable motiva el estudio de los problemas de estimación de funciones lineales del parámetro en unos conos particulares, denominados conos circulares, definidos precisamente a partir de su dirección central c para los cuales el estudio de $E(c'X^*)^2$ parece más viable y que consideraremos solamente en el modelo normal.

La referencia básica en problemas de inferencia con restricciones dadas por conos circulares es Akkerboom (90). El trabajo de Akkerboom está referido al estudio de contrastes de hipótesis para estos conos y su utilidad en otros casos no circulares. Sin embargo y a pesar de encontrarnos en un contexto diferente algunas de las propiedades referentes a la estructura de estos conos y a distribuciones de estadísticos definidos en relación a ellos nos serán de utilidad.

En primer lugar se estudia el comportamiento en $\theta = 0$ de $E_\theta(c'X^*)^2 - E_\theta(c'X)^2$ en función de la dimensión del espacio obteniéndose que el estimador X^* es mejor que X cuando $k < 4$. Sin embargo para $k = 4$ existen conos circulares en los que la situación se invierte.

En segundo lugar en función del ángulo ω del cono donde $0 \leq \omega \leq \pi/2$ obtenemos también que $E_0(c'X^*)^2 - E_0(c'X)^2 > 0$ si la dimensión del espacio k es suficientemente grande.

Por último en el caso $k = 2$, en el que todos los conos son circulares, se hace un desarrollo más detallado obteniendo que cuando $\theta = 0$ no siempre la dirección central del cono es la dirección de estimación más desfavorable sino que existen ángulos para los que esa dirección es la ortogonal a la central y que existe un ángulo para el que $E_0 (d' X^*)^2$ es constante cuando d es un vector unitario cualquiera de \mathfrak{R}^2 .

En la sección cuarta de este capítulo tratamos el caso de los conos de orden bajo normalidad que es uno de los más frecuentes en la literatura. Llamamos conos de orden a aquellos que vienen definidos por relaciones de orden entre los parámetros. Dentro de los conos de orden los más estudiados por sus aplicaciones son el orden total $C_{OT} = \{\theta_1 \leq \theta_2 \leq \dots \leq \theta_k\}$ y el tree order $C_{ST} = \{\theta_0 \leq \theta_i \text{ con } i = 1, \dots, k - 1\}$.

Probamos primeramente algunas propiedades relativas a conos convexos y cerrados cualesquiera que vamos a utilizar después para los conos de orden. Cuestiones relativas a estos conos más generales y que nos serán de utilidad en esta sección pueden encontrarse en referencias como Raubertas, Lee y Nordheim (86) y Menéndez, Rueda y Salvador (92).

Demostramos que los estimadores correspondientes a dos funciones lineales son independientes si las direcciones de estimación pertenecen una a la linealidad del cono $L_S(C)$ y otra al ortogonal a dicha linealidad. Como consecuencia probamos también que la ganancia o pérdida cuadrática de X^* con respecto a X cuando estimamos una función lineal determinada por una dirección d es la misma que cuando la función a estimar viene determinada por

la dirección $P(d/L_S(C)^\perp)$.

Para el caso del tree order la dirección central resulta ser la proyección del vector que determina la primera coordenada $(1, 0, \dots, 0)$ sobre el ortogonal de la linealidad del cono y obtenemos que se tiene pérdida en el error cuadrático medio para las mismas situaciones en las que esta pérdida existía para la primera coordenada.

En el caso del orden total, para ver si efectivamente, el EMV X^* es mejor que X desde el punto de vista de Lehmann y teniendo en cuenta los resultados obtenidos en las secciones precedentes consideramos la dirección central de este cono c . Demostramos aquí que en el caso del orden total, cuando se considera esta dirección $E(c'(X^* - \theta))^2 > E(c'(X - \theta))^2$ para ciertos valores del parámetro cuando la dimensión del espacio es suficientemente grande. Además obtenemos mediante simulación que cuando $k = 7$ ya se tiene este tipo de situación. Nos parece importante señalar aquí que este tipo de resultado no había sido encontrado hasta ahora en el cono del orden total.

1.1 El octante positivo

Sean X_1, \dots, X_k v.a.i. con momento de orden 2, provenientes de un modelo de localización en el que el parámetro es la media de la distribución y verificando $E(X) = E((X_1, \dots, X_k)) = \theta \in O_k^+$, donde O_k^+ representa el octante positivo de \mathfrak{R}^k que se define de la forma siguiente

Definición 1.2 *Denominaremos octante positivo de \mathfrak{R}^k al conjunto*

$$O_k^+ = \{ \theta \in \mathfrak{R}^k / \theta_i \geq 0, i = 1, \dots, k \}$$

En esta sección también emplearemos la siguiente notación

Nota 1.1 *Representaremos por*

$$F(x) = P_0(X_i < x)$$

$$f(x) = F'(x)$$

$$X_i^+ = X_i I_{\{X_i \leq 0\}}$$

$$X_i^- = -X_i I_{\{X_i < 0\}}$$

En este caso el EMV X^* , proyección sobre el octante positivo definido por

Definición 1.3

$$X_i^* = P(X / O_k^+) = X_i^+ = X_i I_{\{X_i \geq 0\}} \text{ para } i = 1, \dots, k$$

Probaremos el siguiente resultado:

Teorema 1.1 *Si $\lim_{x \rightarrow \infty} x^2 F(-x) = 0$ entonces*

$$E\left((c'(X^* - \theta))^2\right) \leq E\left((c'(X - \theta))^2\right) \quad \forall c \in \mathfrak{R}^k, \forall \theta \in O_k^+$$

si y solamente si

$$k \leq 1 + \frac{E_0\left(\left(X_i^-\right)^2\right)}{\left(E_0\left(X_i^-\right)\right)^2} \quad (1.1)$$

Para la demostración de este teorema necesitaremos un par de lemas que probamos previamente.

Lema 1.2 *Dados cuatro números reales positivos a, b, c, d se verifica que:*

$$\frac{a+b}{c+d} \geq \frac{a}{c} \iff \frac{b}{d} \geq \frac{a}{c}$$

Demostración. La demostración es trivial. ■

$$\text{Sean } s(\mu) = E_{\mu}(X_i^-) \text{ y } h(\mu) = E_{\mu}\left((X_i^-)^2\right) + 2\mu E_{\mu}(X_i^-)$$

Lema 1.3 *Supongamos que $s(\theta_i) \neq 0$ para algún i . Entonces la función*

$$f_k(\theta) = \frac{\sum_{i=1}^k c_i^2 h(\theta_i)}{\sum_{i=1}^k c_i^2 s^2(\theta_i)}$$

alcanza su mínimo dentro del octante positivo en 0 independientemente de c , es decir que verifica

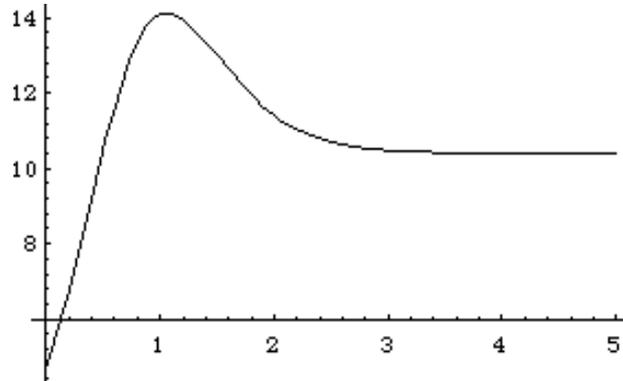
$$f_k(\theta) \geq f_k(0) \quad \forall \theta \in O_k^+$$

Nota 1.2 *Es conveniente tener en cuenta que*

$$s(\theta_i) = 0 \iff h(\theta_i) = 0 \iff F(-\theta_i) = 0$$

Nota 1.3 *En lo sucesivo denotaremos X_i por Y cuando sea conveniente para evitar subíndices innecesarios.*

Antes de la demostración presentamos una gráfica del tipo de función con el que estamos tratando



Sección típica de la función $f_k(\theta)$. Concretamente esta gráfica corresponde a un modelo normal bidimensional; se representa $f_2(1, \theta_2)$ cuando $c_1 = c_2 = 1$

Demostración. Primeramente especificamos más los valores de $s(\mu)$ y $h(\mu)$.

$$\begin{aligned}
 s(\mu) &= E_\mu(Y^-) = -\int_{-\infty}^0 y f(y - \mu) dy \\
 &= -\left(\int_{-\infty}^0 (y - \mu) f(y - \mu) dy + \mu \int_{-\infty}^0 f(y - \mu) dy\right) \quad (1.2) \\
 &= -\int_{-\infty}^{-\mu} y f(y) dy - \mu F(-\mu)
 \end{aligned}$$

mientras que

$$\begin{aligned}
 E_\mu\left((Y^-)^2\right) &= \int_{-\infty}^0 y^2 f(y - \mu) dy \\
 &= \int_{-\infty}^0 (y - \mu)^2 f(y - \mu) dy + 2\mu \int_{-\infty}^0 y f(y - \mu) dy + \mu^2 \int_{-\infty}^0 f(y - \mu) dy \\
 &= \int_{-\infty}^{-\mu} y^2 f(y) dy + 2\mu \int_{-\infty}^{-\mu} y f(y) dy + \mu^2 F(-\mu)
 \end{aligned}$$

de donde

$$\begin{aligned} h(\mu) &= E_{\mu} \left((Y^-)^2 \right) + 2\mu E_{\mu} (Y^-) \\ &= \int_{-\infty}^{-\mu} y^2 f(y) dy - \mu^2 F(-\mu) \end{aligned} \quad (1.3)$$

Además vamos a necesitar

$$s'(\mu) = -F(-\mu) \quad (1.4)$$

y

$$h'(\mu) = -2\mu F(-\mu) \quad (1.5)$$

El lema se va demostrar actuando por inducción sobre k . Probamos entonces el resultado para $k = 1$ para lo que basta ver que $f_1(\theta) = \frac{h(\theta)}{s^2(\theta)}$ es creciente en θ .

$$f_1'(\theta) = \frac{h'(\theta) s(\theta) - 2h(\theta) s'(\theta)}{s^3(\theta)}$$

y de (1.4) y (1.5) tenemos

$$\begin{aligned} h'(\theta) s(\theta) - 2h(\theta) s'(\theta) &= 2F(-\theta) (h(\theta) - \theta s(\theta)) \\ &= 2F(-\theta) \left(E_{\theta} \left((Y^-)^2 \right) + \theta E_{\theta} (Y^-) \right) \end{aligned}$$

de donde es inmediato que $f_1(\theta) \geq 0$ por ser Y^- una variable positiva y $\theta \in O^+$.

Además $f_1(\theta) > 0$ puesto que estamos suponiendo $s(\theta) = 0 \Leftrightarrow F(-\theta) = 0$.

Actuando por inducción suponemos ahora que $f_{k-1}(\theta) \geq f_{k-1}(0)$ en las condiciones descritas y vamos a probar que $f_k(\theta) \geq f_k(0)$. Supongamos por contra que existe $\theta^0 \in O_k^+$ tal que $f_k(\theta^0) < f_k(0)$.

Fijamos ahora las $k - 1$ primeras variables de la función $f_k(\theta)$, $\theta_1, \dots, \theta_{k-1}$ y consideramos $g(\theta_k) = f_k(\theta_1, \dots, \theta_k)$. Veamos que esta función alcanza su mínimo cuando $\theta_k = 0$ cualesquiera que sean $\theta_i \in O^+$ para $i = 1, \dots, k - 1$.

Excluimos primeramente la situación $s(\theta_1) = \dots = s(\theta_{k-1}) = 0$ puesto que entonces estaríamos de nuevo en el caso $k = 1$ y habríamos terminado.

En el resto de las situaciones lo que vamos a probar para ver esto es, primeramente que $g'(0) > 0$ con lo que en 0 hay un mínimo local, y después que esa función tiene a lo sumo otro extremo en O_k^+ que debería ser por tanto un máximo con lo que el único competidor de 0 como mínimo es el límite cuando $\theta \rightarrow \infty$ que veremos que es mayor que $g(0)$.

$$g'(\theta_k) = \frac{c_k^2 h'(\theta_k) \left(\sum_{i=1}^k c_i^2 s^2(\theta_i) \right) - 2 \left(\sum_{i=1}^k c_i^2 h(\theta_i) \right) c_k^2 s(\theta_k) s'(\theta_k)}{\left(\sum_{i=1}^k c_i^2 s^2(\theta_i) \right)^2} \quad (1.6)$$

de donde tenemos

$$g'(0) = \frac{2c_k^2 F(0) \left(\left(\sum_{i=1}^{k-1} c_i^2 h(\theta_i) \right) + c_k^2 h(0) \right) s(0)}{\left(\sum_{i=1}^{k-1} c_i^2 s^2(\theta_i) + c_k^2 s^2(0) \right)^2} > 0$$

notar que si $F(0) = 0$ el problema que consideramos carece de sentido puesto que serían variables positivas.

Además de (1.6) tenemos $g'(\theta_k) = 0$ si y solo si

$$-2c_k^2 F(-\theta_k) \left(\theta_k \left(\sum_{i=1}^k c_i^2 s^2(\theta_i) \right) - s(\theta_k) \left(\sum_{i=1}^k c_i^2 h(\theta_i) \right) \right) = 0$$

es decir que, o bien

$$F(-\theta_k) = 0$$

o bien

$$c_k^2 \theta_k s^2(\theta_k) + \theta_k A - s(\theta_k) c_k^2 h(\theta_k) - s(\theta_k) B = 0$$

donde $A = \sum_{i=1}^{k-1} c_i^2 s^2(\theta_i)$ y $B = \sum_{i=1}^{k-1} c_i^2 h(\theta_i)$. Entonces utilizando (1.2) y (3) tenemos:

$$-c_k^2 \left(\int_{-\infty}^{-\theta_k} x(x + \theta_k) f(x) dx \right) + \frac{\theta_k}{s(\theta_k)} A = B \quad (1.7)$$

Si ahora llamamos $t(\theta_k)$ al lado izquierdo de la expresión (1.7) y probamos que la derivada de esa función es positiva será estrictamente creciente y la solución de la ecuación (1.7) será única.

$$\begin{aligned} t'(\theta_k) &= -c_k^2 \left(\int_{-\infty}^{-\theta_k} x f(x) dx \right) + A \frac{s(\theta_k) - \theta_k s'(\theta_k)}{s^2(\theta_k)} = \\ &= c_k^2 \left(\int_{-\infty}^{-\theta_k} -x f(x) dx \right) + A \frac{\int_{-\infty}^{-\theta_k} -x f(x) dx}{s^2(\theta_k)} > 0 \end{aligned}$$

utilizando (1.2) y (1.4). Notar que de nuevo $\int_{-\infty}^{-\theta_k} -x f(x) dx = 0 \Leftrightarrow F(-\theta_k) = 0$.

Entonces hemos probado que o bien existe un punto θ_k^1 verificando $F(-\theta_k^1) = 0$ con lo cual $F(-\theta_k) = 0$ para todo $\theta_k \geq \theta_k^1$ o bien el único extremo que puede haber es un máximo. En cualquiera de los dos casos el único competidor de 0 como mínimo es $\lim_{\theta_k \rightarrow \infty} g(\theta_k)$. Pero ahora tenemos:

$$\lim_{\theta_k \rightarrow \infty} g(\theta_k) = \frac{\sum_{i=1}^{k-1} c_i^2 h(\theta_i)}{\sum_{i=1}^{k-1} c_i^2 s^2(\theta_i)}$$

puesto que $\lim_{\theta_k \rightarrow \infty} h(\theta_k) = \lim_{\theta_k \rightarrow \infty} s(\theta_k) = 0$ ya que $\lim_{\theta \rightarrow \infty} \theta^2 F(-\theta) = 0$ por hipótesis. Por otra parte

$$g(0) = \frac{\sum_{i=1}^{k-1} c_i^2 h(\theta_i) + c_k^2 h(0)}{\sum_{i=1}^{k-1} c_i^2 s^2(\theta_i) + c_k^2 s^2(0)}$$

con lo que obtenemos $g(0) \leq \lim_{\theta_k \rightarrow \infty} g(\theta_k)$ sin más que utilizar el lema 1.2 y la hipótesis de inducción.

Entonces si existe el punto $\theta^0 \in O_k^+$ tal que $f_k(\theta^0) < f_k(0)$ tenemos también que $f_k(\theta_1^0, \dots, \theta_{k-1}^0, 0) < f_k(0)$ es decir que $g(0) < f_k(0)$ pero esto es falso utilizando de nuevo el lema 1.2 y la hipótesis de inducción. ■

Demostración del Teorema :

Notemos primeramente que si $s(\theta_i) = 0$ para todos los subíndices entonces $X^* = X$ y el problema no tiene interés.

Por otra parte tenemos

$$E(c'(X - \theta))^2 = \sum_{i=1}^k c_i^2 E(X_i - \theta_i)^2 \quad (1.8)$$

mientras que

$$E(c'(X^* - \theta))^2 = \sum_{i=1}^k c_i^2 E(X_i^* - \theta_i)^2 + 2 \sum_{i < j} c_i c_j E(X_i^* - \theta_i) E(X_j^* - \theta_j) \quad (1.9)$$

Ahora bien como:

$$X_i - \theta_i = (X_i^+ - \theta_i) - X_i^- = (X_i^* - \theta_i) - X_i^-$$

se tiene

$$E(X_i^* - \theta_i) = E(X_i - \theta_i) + E_{\theta_i}(X^-) = E_{\theta_i}(X^-) = s(\theta_i) \quad (1.10)$$

y de

$$(X_i - \theta_i)^2 = ((X_i^+ - \theta_i) - X_i^-)^2 = (X_i^-)^2 + 2\theta_i X_i^- + (X_i^* - \theta_i)^2$$

obtenemos

$$E(X_i - \theta_i)^2 - E(X_i^* - \theta_i)^2 = E_{\theta_i}(X^-)^2 + 2\theta_i E_{\theta_i}(X^-) = h(\theta_i) \quad (1.11)$$

Finalmente restando (1.9) de (1.8) y teniendo en cuenta (1.10) y (1.11)

tenemos

$$\begin{aligned} E(c'(X - \underline{\theta}))^2 - E(c'(X^* - \underline{\theta}))^2 &= \sum_{i=1}^k c_i^2 h(\theta_i) - 2 \sum_{i < j} c_i c_j s(\theta_i) s(\theta_j) \\ &= \sum_{i=1}^k c_i^2 h(\theta_i) - ((d' \cdot 1)^2 - (d' \cdot d)) \\ &= \sum_{i=1}^k c_i^2 h(\theta_i) + (d' \cdot d) - k(d' \cdot d) \cos^2(d, 1) \end{aligned}$$

donde $d_i = c_i s(\theta_i)$. Obtenemos ahora inmediatamente la siguiente desigualdad

$$E(c'(X - \theta))^2 - E(c'(X^* - \theta))^2 \geq 0$$

si solamente si

$$k \leq \frac{1}{\cos^2(d, 1)} \left(1 + \frac{\sum_{i=1}^k c_i^2 h(\theta_i)}{(d' \cdot d)} \right) \quad (1.12)$$

Ahora bien $\frac{\sum_{i=1}^k c_i^2 h(\theta_i)}{(d' \cdot d)} = f_k(\theta)$ y el lema 1.3 prueba que esta función alcanza su mínimo en 0 independientemente del valor de c . Como además si $\theta = 0$, $\cos^2(d, 1) = \cos^2(c, 1)$ y esta función alcanza su máximo, que vale 1, cuando $c = \lambda \cdot (1, \dots, 1)$, entonces el lado derecho de la desigualdad (1.12) alcanza su mínimo cuando $c = \lambda \cdot (1, \dots, 1)$ y $\theta = 0$ y este mínimo vale

$$1 + \frac{h(0)}{s^2(0)} = 1 + \frac{E_0\left(\left(X_i^-\right)^2\right)}{\left(E_0\left(X_i^-\right)\right)^2}$$

■

Si ahora nos dedicamos a estimar algunas direcciones particulares obtenemos los siguientes resultados.

Corolario 1.4 *La dirección más desfavorable para la estimación cuando se comparan X^* y X es la diagonal del cono en cualquier caso.*

Demostración. Evidente de la demostración del teorema. ■

Corolario 1.5 *Si intentamos estimar una dirección c tal que $\cos(c, 1) = cte$ (es decir independiente de k) entonces existe k_0 tal que*

$$E\left((c'(X^* - \theta))^2\right) > E\left((c'(X - \theta))^2\right) \text{ para todo } k \geq k_0$$

Demostración. Evidente de (1.12) eligiendo

$$k > \frac{1}{\cos^2(c, 1)} \left(1 + \frac{E_0\left(\left(X_i^-\right)^2\right)}{\left(E_0\left(X_i^-\right)\right)^2} \right)$$

■

Corolario 1.6 *Para las direcciones $c' = (0, \dots, \overset{i}{1}, \dots, 0) = e'_i$, $i = 1, \dots, k$ y $c' = e'_i - e'_j$ para $i, j = 1, \dots, k$, $i \neq j$ se tiene que*

$$E\left((c'(X^* - \theta))^2\right) \leq E\left((c'(X - \theta))^2\right) \text{ para todo } k > 0$$

Demostración. La primera afirmación es evidente de (1.11) aunque puede deducirse también de la demostración del teorema 3.9 de Moors (85). ■

Demostración. En cuanto a la segunda referida a la comparación de medias de dos poblaciones basta notar que (1.9) se transforma en

$$\begin{aligned} E\left(\left(X_i^* - X_j^* - (\theta_i - \theta_j)\right)^2\right) &= E\left(\left(X_i^* - \theta_i\right)^2\right) + E\left(\left(X_j^* - \theta_j\right)^2\right) - \\ &- 2E\left(X_i^* - \theta_i\right)E\left(X_j^* - \theta_j\right) \end{aligned}$$

y tener en cuenta las ecuaciones (1.10) y (1.11). ■

Nota 1.4 *No se puede concluir del teorema que fijado c se alcance en $\theta = 0$ el mínimo de $E((c'(X^* - \theta))^2) - E((c'(X - \theta))^2)$ o dicho de otro modo que en no es cierto que para un c cualquiera*

$$E((c'(X^* - \theta))^2) \leq E((c'(X - \theta))^2) \Leftrightarrow k \leq \frac{1}{\cos^2(c, 1)} \left(1 + \frac{E_0((X_i^-)^2)}{(E_0(X_i^-))^2} \right)$$

en el Ejemplo 1.5 aparece un contraejemplo referido a esta situación.

Ejemplo 1.1 La condición $\lim_{\theta \rightarrow \infty} \theta^2 F(-\theta) = 0$ se verifica en particular cuando $P(X < k) = 0$ para algún $k \in \mathfrak{R}$, como en el caso de las variables del tipo $U(\theta - a, \theta + a)$. En esta situación

$$s(\theta) = E_\theta(Y^-) = - \int_{\theta-a}^0 y \frac{1}{2a} dy = \frac{(\theta - a)^2}{4a}$$

$$E_\theta((Y^-)^2) = \int_{\theta-a}^0 y^2 \frac{1}{2a} dy = \frac{(\theta - a)^3}{6a}$$

$$h(\theta) = E_\theta((Y^-)^2) + 2\theta E_\theta(Y^-) = \frac{(\theta - a)^3}{6a} + 2\theta \frac{(\theta - a)^2}{4a}$$

Por tanto

$$f_k(0) = \frac{h(0)}{s^2(0)} = \frac{a^3/6a}{(a^2/4a)} = \frac{8}{3}$$

y en éste modelo la estimación de funciones lineales es mejor con X^* mientras $k \leq 3$.

Ejemplo 1.2 En el modelo normal, es decir cuando $Y \rightsquigarrow N(\theta, \sigma^2)$ tenemos que claramente $\lim_{\theta \rightarrow \infty} \theta^2 F(-\theta) = 0$ y además

$$\begin{aligned} E_{\theta}(Y^-) &= \int_{-\infty}^0 -y \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y-\theta)^2}{2\sigma^2}\right\} dy \\ &= \sigma\varphi\left(\frac{\theta}{\sigma}\right) - \theta\Phi\left(-\frac{\theta}{\sigma}\right) = s_{\sigma}(\theta) \end{aligned}$$

$$E_{\theta}\left(\left(Y^-\right)^2\right) = (\sigma^2 + \theta^2)\Phi\left(-\frac{\theta}{\sigma}\right) - \theta\sigma\varphi\left(-\frac{\theta}{\sigma}\right)$$

$$h_{\sigma}(\theta) = E_{\theta}\left(\left(Y^-\right)^2\right) + 2\theta E_{\theta}(Y^-) = (\sigma^2 - \theta^2)\Phi\left(-\frac{\theta}{\sigma}\right) + \theta\sigma\varphi\left(-\frac{\theta}{\sigma}\right)$$

Entonces

$$f_k(0) = \frac{h_{\sigma}(0)}{s_{\sigma}^2(0)} = \frac{\sigma^2/2}{\sigma^2/2\pi} = \pi$$

con lo que en el modelo normal el EMV, que en este caso es X^* , es mejor para estimar cualquier función lineal del parámetro cuando este está en O_k^+ cuando $k \leq 1 + \pi$ es decir cuando $k < 5$. Por contra si $k \geq 5$ entonces existe $\underline{c} \in \mathfrak{R}^k$ para el que es preferible utilizar X como base de la estimación en lugar de X^* .

Ejemplo 1.3 Si $X \rightsquigarrow DExp(\theta, \lambda)$ también tenemos que X^* es el EMV y además

$$\begin{aligned} \lim_{\theta \rightarrow \infty} \theta^2 F(-\theta) &= \lim_{\theta \rightarrow \infty} \theta^2 \left(\int_{-\infty}^{-\theta} \frac{1}{2} \lambda \exp\{-\lambda|x|\} dx \right) \\ &= \lim_{\theta \rightarrow \infty} \frac{1}{2} \theta^2 \exp\{-\lambda\theta\} = \lim_{\theta \rightarrow \infty} \frac{1}{\lambda^2 \exp\{\lambda\theta\}} = 0 \end{aligned}$$

$$s(0) = - \int_{-\infty}^0 \frac{1}{2} \lambda x \exp\{\lambda x\} dx = \frac{1}{2\lambda}$$

$$h(0) = \int_{-\infty}^0 \frac{1}{2} \lambda x^2 \exp\{\lambda x\} dx = \frac{1}{2\lambda^2}$$

con lo cual

$$f_k(0) = \frac{h(0)}{s^2(0)} = 2$$

Por tanto X^* es mejor para estimar cualquier función lineal del parámetro cuando este está en O_k^+ cuando $k \leq 3$.

Ejemplo 1.4 Del lado izquierdo de (1.1) se deduce inmediatamente que para cualquier modelo de localización y $k = 2$ es siempre mejor el estimador X^* que X puesto que $\frac{E_0((X_i^-)^2)}{(E_0(X_i^-))^2} \geq 1$. Además esa cota se alcanza en este ejemplo.

Sean a, b, ε tres números reales tales que $a, \varepsilon > 0$ y $b < 0$ y sea $p \in (0, 1)$.

Definimos Y tal que $f(y) = \begin{cases} p & \text{si } y \in (a - 1/2, a + 1/2) \\ \frac{1-p}{2\varepsilon} & \text{si } y \in (b - \varepsilon, b + \varepsilon) \\ 0 & \text{en el resto de los casos} \end{cases}$. Tenemos entonces

$$E_\theta(Y) = ap + b(1 - p)$$

$$E_\theta(Y^-) = -(1 - p)b$$

$$E_\theta((Y^-)^2) = \int_{b-\varepsilon}^{b+\varepsilon} y^2 \frac{1-p}{2\varepsilon} dy = (b^2 + \varepsilon^2)(1 - p)$$

Si queremos que $E(Y) = \theta$ valga 0 debe suceder $p = \frac{-b}{a-b}$ con lo cual

$$f_k(0) = \frac{h(0)}{s^2(0)} = \frac{E_0((Y^-)^2)}{(E_0(Y^-))^2} = \frac{(b^2 + \varepsilon^2)(1 - p)}{b^2(1 - p)^2} = \left(1 + \frac{\varepsilon^2}{b^2}\right) \frac{a - b}{a}$$

y eligiendo b y ε suficientemente pequeños conseguimos que $f_k(0) < 2$ y estamos en la situación buscada. Es decir que para $k = 3$ se pierde en algunas direcciones si se utiliza el estimador X^* .

Además se puede conseguir una distribución para la que no se pierda hasta un $k_0 \geq 2$ fijado sin más que elegir b suficientemente grande en:

$$k_0 = 1 + \left(1 + \frac{\varepsilon^2}{b^2}\right) \left(\frac{a-b}{a}\right)$$

Ejemplo 1.5 Tomamos $c_1 = \dots = c_{k-1} = 1$ y $c_k = -(k-1)$ con lo que $\cos(c, 1) = 0$. Mientras que si elegimos $\theta_1 = \dots = \theta_{k-1} = 0$ pero θ_k suficientemente grande tenemos que $d_i = s(0)$ si $1 \leq i \leq k-1$ y $d_k \xrightarrow{\theta_k \rightarrow \infty} 0$ con lo que utilizando (1.12) tenemos

$$\frac{1}{\cos^2(d, 1)} \left(1 + \frac{\sum_{i=1}^k c_i^2 h(\theta_i)}{(d' \cdot d)}\right) \xrightarrow{\theta_k \rightarrow \infty} \frac{k}{k-1} \left(1 + \frac{E_0 \left((X_i^-)^2 \right)}{\left(E_0 (X_i^-) \right)^2}\right)$$

y este último termino será menor que k simplemente tomando $k > 2 + \frac{E_0 \left((X_i^-)^2 \right)}{\left(E_0 (X_i^-) \right)^2}$.

Ejemplo 1.6 Modelo lineal normal

Consideremos un modelo lineal habitual, esto es tal que $Y = X\beta + \varepsilon$ donde $\varepsilon \rightsquigarrow N_n(0, \sigma^2 I)$ con lo cual $\hat{\beta} \rightsquigarrow N_k(\beta, \sigma^2 (X'X)^{-1})$, como es bien sabido, y supongamos que tenemos medios para situar los regresores al nivel que deseemos con lo que podremos conseguir que $X'X = I$.

Es habitual en situaciones de este tipo el que se conozca el carácter de la relación que pueda existir entre los regresores y la variable respuesta o dicho de otro modo el signo de β_i para cada i . Entonces haciendo una transformación trivial podemos suponer que $\beta_i \geq 0$ para $i = 1, \dots, k$.

En resumen tenemos $\hat{\beta} \rightsquigarrow N_k(\beta, \sigma^2 I)$ y $\beta_i \geq 0$ para $i = 1, \dots, k$ con lo que utilizando el Ejemplo 1.2 obtenemos que la estimación del valor de

la variable respuesta a un nivel cualquiera de las variables regresoras $\widehat{Y}_i = x_i' \widehat{\beta} = \sum_{j=1}^k x_{ij} \widehat{\beta}_j$, que es una función lineal de los parámetros, será mejor desde el punto de vista del error cuadrático medio si y solamente si el número de regresores es estrictamente menor que 5.

Nota 1.5 *Es claro que, si en lugar de suponer las restricciones que hemos empleado en esta sección sabemos que $\theta_i \geq a_i$ para $i = 1, \dots, k$ donde los a_i son números reales conocidos todos los resultados aquí demostrados siguen siendo ciertos haciendo las traslaciones correspondientes y el cambio de variable $Y_i = X_i - a_i$ para $i = 1, \dots, k$ con lo que $E(Y_i) \geq 0$.*

Con otras palabras, el problema es invariante frente a este tipo de traslaciones.

Además en el caso de un modelo de localización y escala de parámetros θ_i y σ_i^2 respectivamente también es claro que el caso en el que los parámetros de escala sean conocidos puede reducirse al estudiado en esta sección sin más que considerar el cambio de variable $Y_i = X_i/\sigma_i$ puesto que si se sabía que $E(X_i) \geq 0$ también es $E(Y_i) = E(X_i)/\sigma_i \geq 0$ para $i = 1, \dots, k$.

Es decir que este problema también es invariante en relación a estos cambios de escala.

1.2 Otros conos rectos

Definición 1.4 *Decimos que un cono $C = \{x \in \mathfrak{R}^k / a_i' x \geq 0, i = 1, \dots, n\}$ es recto cuando $a_i' \cdot a_j = 0$ para $i \neq j$ y $1 \leq i, j \leq n$.*

Para los resultados relativos a conos rectos generales que probaremos nos será de utilidad también el concepto de linealidad de un cono.

Definición 1.5 Dado un cono $C = \{x \in \mathfrak{R}^k / a'_i x \geq 0, i = 1, \dots, n\}$ llamamos linealidad de C y lo denotamos por $L_S(C)$ al subespacio de mayor dimensión contenido en C , es decir

$$L_S(C) = \{x \in \mathfrak{R}^k / a'_i x = 0, i = 1, \dots, n\}$$

En este caso el modelo a considerar no será tan general como en la sección anterior sino que vamos a suponer que $X_i \rightsquigarrow N(\theta_i, \sigma^2)$, $i = 1, \dots, k$ son variables aleatorias independientes tales que $\theta = (\theta_1, \dots, \theta_k) \in C$ donde $C = \{x \in \mathfrak{R}^k / a'_i x \geq 0, i = 1, \dots, n\}$ ($k \geq n$) cono recto cualquiera. Probamos entonces el siguiente resultado

Teorema 1.7

$$E((c'(X^* - \theta))^2) \leq E((c'(X - \theta))^2) \quad \forall c \in \mathfrak{R}^k, \forall \theta \in C$$

si y solamente si

$$k \leq 4 + p \tag{1.13}$$

donde $p = k - n = \dim L_S(C)$.

Veamos primeramente un lema que contiene los cálculos necesarios

Lema 1.8 Sean $Y_i \rightsquigarrow N(\mu_i, \sigma^2)$, $i = 1, \dots, k$ variables aleatorias independientes tales que $\mu = (\mu_1, \dots, \mu_k) \in C'$ donde $C' = \{x \in \mathfrak{R}^k / x_i \geq 0, i = 1, \dots, n\}$

es decir el octante positivo de \mathfrak{R}^n sumergido en \mathfrak{R}^k . En estas condiciones

$$\begin{aligned} & E\left(\left(b'(Y^* - \underline{\mu})\right)^2\right) - E\left(\left(b'(Y - \underline{\mu})\right)^2\right) \\ &= E\left(\left(\sum_{i=1}^k b_i(Y_i^* - \mu_i)\right)^2\right) - E\left(\left(\sum_{i=1}^k b_i(Y_i - \mu_i)\right)^2\right) \\ &= E\left(\left(\sum_{i=1}^n b_i(Y_i^* - \mu_i)\right)^2\right) - E\left(\left(\sum_{i=1}^n b_i(Y_i - \mu_i)\right)^2\right) \end{aligned} \quad (1.14)$$

para cualquier $b \in \mathfrak{R}^k$.

Demostración. La prueba es directa teniendo en cuenta que $Y_i^* = Y_i$ para $i = n + 1, \dots, k$ y que por la definición de las variables Y_i^* estas siguen siendo independientes con lo que

$$\begin{aligned} & E\left(\left(\sum_{i=1}^k b_i(Y_i^* - \mu_i)\right)^2\right) \\ &= \sum_{i=1}^k b_i^2 E\left((Y_i^* - \mu_i)^2\right) + \sum_{i,j=1}^k b_i b_j E(Y_i^* - \mu_i) E(Y_j^* - \mu_j) \\ &= \sum_{i=1}^k b_i^2 E\left((Y_i^* - \mu_i)^2\right) + \sum_{i,j=1}^n b_i b_j E(Y_i^* - \mu_i) E(Y_j^* - \mu_j) \end{aligned}$$

ya que $E(Y_i^* - \mu_i) = 0$ para $i = n + 1, \dots, k$. Como además

$$E\left(\left(\sum_{i=1}^k b_i(Y_i - \mu_i)\right)^2\right) = \sum_{i=1}^k b_i^2 E\left((Y_i - \mu_i)^2\right)$$

obtenemos inmediatamente (1.14). ■

Ahora damos la demostración del teorema.

Demostración. Elegimos una base ortonormal B de \mathfrak{R}^k de la siguiente forma

$$B = \left\{ \frac{a_i}{\|a_i\|}, i = 1, \dots, n \right\} \cup \text{base ortonormal de } L_S(C)$$

y llamamos M a la matriz que tiene por filas los vectores de la base B . Por construcción esta matriz verifica que $MM' = M'M = I$.

Si $X \rightsquigarrow N_k(\theta, \sigma^2 I)$ donde $\underline{\theta} \in C$ entonces $Y = MX \rightsquigarrow N_k(M\theta, \sigma^2 MM')$ con $M\theta = \mu \in MC = C' = \{Mx / x \in C\} = \{x \in \Re^k / x_i \geq 0, i = 1, \dots, n\}$.

Además el estimador máximo verosímil para μ será

$$Y^* = P(Y/MC) = P(MX/MC) = M \cdot P(X/C) = MX^*$$

por las propiedades de las proyecciones.

Concretando más

$$Y_i^* = \begin{cases} Y_i^+ & \text{cuando } i = 1, \dots, n \\ Y_i & \text{cuando } i = n + 1, \dots, k \end{cases}$$

Ahora

$$\begin{aligned} & E\left((c'(X^* - \theta))^2\right) - E\left((c'(X - \theta))^2\right) \\ &= E\left(\left((Mc)'(MX^* - M\theta)\right)^2\right) - E\left(\left((Mc)'(MX - M\theta)\right)^2\right) \\ &= E\left(\left(b'(Y^* - \mu)\right)^2\right) - E\left(\left(b'(Y - \mu)\right)^2\right) \\ &= E\left(\left(\sum_{i=1}^n b_i(Y_i^* - \mu_i)\right)^2\right) - E\left(\left(\sum_{i=1}^n b_i(Y_i - \mu_i)\right)^2\right) \end{aligned} \quad (1.15)$$

donde la primera igualdad es cierta porque $M'M = I$ y la tercera proviene del lema previo. Si tenemos en cuenta el Ejemplo 1.2 de la sección anterior obtenemos inmediatamente (1.13). ■

Corolario 1.9 *Las direcciones de estimación más desfavorables para X^* se pueden escribir ahora como*

$$\alpha \cdot M' \left(\left(1, \dots, \overset{n}{1}, \lambda_{n+1}, \dots, \lambda_k \right)' \right) \quad (1.16)$$

donde $\alpha, \lambda_{n+1}, \dots, \lambda_k$ son números reales cualesquiera.

Demostración. Evidente teniendo en cuenta (1.15) y aplicando el corolario 1.4 de la sección anterior. Notar que los vectores escritos en (1.16) son la transformación inversa de vectores que pueden escribirse como un elemento de la diagonal del octante positivo de \mathfrak{R}^n sumergido en \mathfrak{R}^k , $(1, \dots, \overset{n}{1}, 0, \dots, 0)'$ más uno cualquiera de los vectores de $L_S(C') = \{x \in \mathfrak{R}^k / x_1 = \dots = x_n = 0\}$. ■

Ejemplo 1.7 Cono de los promedios crecientes y cono star-shaped

El cono de los promedios crecientes se define como

$$C_1 = \left\{ \theta_1 \leq \frac{\theta_1 + \theta_2}{2} \leq \dots \leq \frac{\theta_1 + \dots + \theta_k}{k} \right\}$$

y puede escribirse también como

$$C_1 = \{x \in \mathfrak{R}^k / a'_i x \geq 0, i = 1, \dots, k-1\}$$

eligiendo

$$a'_i = \left(-1, \dots, \overset{i}{-1}, i, 0, \dots, 0 \right) \text{ para } i = 1, \dots, k-1 \quad (1.17)$$

de donde es obvio que $a'_i \cdot a_j = 0$ si $i \neq j$ y el cono es recto.

Además $L_S(C_1) = \{x \in \mathfrak{R}^k / x' = \lambda(1, \dots, 1)'\}$ y por tanto $p = \dim L_S(C_1) = 1$.

La matriz de transformación M en este caso vale

$$M = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & \dots & 0 \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & 0 & \dots & 0 \\ -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & \frac{3}{\sqrt{12}} & \dots & 0 \\ \vdots & & & & & \\ \frac{-1}{\sqrt{k(k-1)}} & \frac{-1}{\sqrt{k(k-1)}} & \frac{-1}{\sqrt{k(k-1)}} & \frac{-1}{\sqrt{k(k-1)}} & \dots & \frac{k-1}{\sqrt{k(k-1)}} \\ -\frac{1}{\sqrt{k}} & -\frac{1}{\sqrt{k}} & -\frac{1}{\sqrt{k}} & -\frac{1}{\sqrt{k}} & \dots & -\frac{1}{\sqrt{k}} \end{pmatrix}$$

En este caso como $\dim L_S(C_1) = 1$ tenemos que los estimadores basados en X^* tienen menor error cuadrático medio que los basados en X para estimar cualquier función lineal del parámetro si y solamente si $k \leq 5$.

En cuanto a las direcciones más desfavorables, para generar la diagonal de C_1 podemos utilizar un elemento de la diagonal del octante positivo de \mathfrak{R}^n sumergido en \mathfrak{R}^k como se hace en el corolario. Si elegimos por ejemplo $\underline{b}' = (1, \dots, 1, 0)$ obtenemos que $M'\underline{b} = \underline{d}$ vale

$$\left(-\left(\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{6}} + \dots + \frac{1}{\sqrt{k(k-1)}} \right), \frac{1}{\sqrt{2}} - \left(\frac{1}{\sqrt{6}} + \dots + \frac{1}{\sqrt{k(k-1)}} \right), \dots, \frac{-(k-1)}{\sqrt{k(k-1)}} \right)'$$

con lo cual esas direcciones serán: $\underline{d} + \lambda (1, \dots, 1)$, $\lambda \in \mathfrak{R}$.

En cuanto al cono star-shaped podemos definirlo como

$$C_2 = \left\{ \theta_1 \leq \frac{\theta_1 + \theta_2}{2} \leq \dots \leq \frac{\theta_1 + \dots + \theta_k}{k} \leq 0 \right\}$$

y reescribirlo como

$$C_2 = \left\{ x \in \mathfrak{R}^k / d'_i x \geq 0, i = 1, \dots, k \right\}$$

donde los $k - 1$ primeros vectores están definidos en 1.17 y $a'_k = -(1, \dots, 1)$. Ahora la dimensión de la linealidad de C_2 es 0 y la matriz M es la misma que la de C_1 . La diagonal de este cono y dirección más desfavorable será la definida por el vector \underline{d} definido para el cono anterior.

1.3 Conos circulares

Definición Se dice que $C(\omega, \underline{c})$ es un cono circular de \mathfrak{R}^k con eje \underline{c} y ángulo ω donde $\underline{c} \in \mathfrak{R}^k$ y $0 \leq \omega \leq \pi/2$ si $C(\omega, \underline{c}) = \{\xi \in \mathfrak{R}^k : \underline{c}'\xi \geq \|\xi\| \cos(\omega)\}$

En esta sección supondremos que $X \rightsquigarrow N_k(\theta, \sigma^2 I)$ donde $\theta \in C$. Además supondremos, sin pérdida de generalidad, que $\sigma^2 = 1$ y $\|c\| = 1$ donde $\|\cdot\|$ es la norma euclídea habitual en \mathfrak{R}^k . El EMV sin restricciones es por tanto X y sea como hasta ahora X^* el EMV con restricciones en esta situación.

Para determinar $E_0(c'X^*)^2$ se va a utilizar la distribución del estadístico $\|X^*\|^2$ cuando $\theta = 0$. Esta distribución, como es habitual, es una distribución Chi-BarSquared pero esto no puede determinarse como en otras ocasiones haciendo uso de la independencia entre la norma de la proyección y la región sobre la que se proyecta y agrupando después los sumandos según la dimensión de dicha región, puesto que los conos que tratamos son circulares y no cabe hablar de dimensión de caras del cono, ni siquiera de caras de dicho cono. Este argumento está implícito en los desarrollos de Akkerboom (90) parte de los cuales vamos a utilizar en la demostración del lema anterior y que se agrupan en los dos siguientes lemas.

Lema 1.10 (Akkerboom (90)) Sean $U = c'X$ y $V = \left(\|X\|^2 - (c'X)^2\right)^{1/2}$ entonces

$$\|X^*\|^2 = \begin{cases} u^2 + v^2 & \text{si } u \geq (u^2 + v^2)^{1/2} \cos \omega \\ (u \cos \omega + v \sin \omega)^2 & \text{si } \begin{aligned} &-(u^2 + v^2)^{1/2} \sin \omega < u \\ &u < (u^2 + v^2)^{1/2} \cos \omega \end{aligned} \\ 0 & \text{si } u \leq (u^2 + v^2)^{1/2} \cos \omega \end{cases}$$

Lema 1.11 (Akkerboom (90)) La distribución conjunta de U y V cuando $\theta = 0$, $\sigma^2 = 1$ viene dada por

$$q(u, v) = C_k \cdot \exp \left\{ -\frac{1}{2} (u^2 + v^2) \right\} \cdot v^{k-2} \text{ para } u \in \Re \text{ y } 0 \leq v < \infty$$

donde la constante de normalización C_k está definida en (1.20).

Además si denotamos por Ψ_f la función de distribución de una χ_f^2 , $\Psi_0 = 1_{[0, \infty)}$,

$B_{p,q}$ la función de distribución de una Beta con parámetros p y q y $B(p, q) =$

$\Gamma(p) \Gamma(q) / \Gamma(p + q)$. Entonces:

$$\Pr_0 \left(\|X^*\|^2 \leq t \right) = \sum_{j=0}^k \delta_j \Psi_j(t) \quad (1.18)$$

donde los δ_j están dados por

$$\begin{aligned} \delta_0 &= \frac{1}{2} B_{1/2(k-1), 1/2}(\cos^2 \omega) = \Pr_0(X \in C^P) \\ \delta_k &= \frac{1}{2} B_{1/2(k-1), 1/2}(\sin^2 \omega) = \Pr_0(X \in C) \\ \delta_j &= \frac{1}{2} \binom{k-2}{j-1} \frac{B\left(\frac{1}{2}j, \frac{1}{2}(k-j)\right)}{B\left(\frac{1}{2}, \frac{1}{2}(k-1)\right)} \cdot \sin^{j-1} \omega \cdot \cos^{k-j-1} \omega \end{aligned} \quad (1.19)$$

para $1 \leq j \leq k-1$.

Enunciamos ahora el lema que necesitaremos para los desarrollos posteriores.

Lema 1.12 *Si $\theta = 0$ entonces*

$$E(c'X^*)^2 = \cos^2 \omega \left(\sum_{j=1}^{k-1} \delta_j \cdot j \right) + C_k \int_0^\infty \int_0^{u \tan \omega} u^2 \exp \left\{ -\frac{1}{2} (u^2 + v^2) \right\} v^{k-2} du dv$$

donde los δ_j están definidos en (1.19) para $1 \leq j \leq k-1$ y

$$C_k = \left\{ 2^{\frac{k-2}{2}} \Gamma \left(\frac{1}{2} \right) \Gamma \left(\frac{k-1}{2} \right) \right\}^{-1} \quad (1.20)$$

Demostración. Teniendo en cuenta la partición de \mathfrak{R}^k en C , C^P y $\mathfrak{R}^k - (C \cup C^P) = R_0$

$$E(c'X^*)^2 = \int_C (c'X^*)^2 f(x) dx + \int_{R_0} (c'X^*)^2 f(x) dx + \int_{C^P} (c'X^*)^2 f(x) dx$$

utilizando las propiedades de la proyección y que C es un cono circular de ángulo ω se obtiene

$$= \int_C (c'X)^2 f(x) dx + \int_{R_0} \cos^2 \omega \cdot \|X^*\|^2 f(x) dx = I1 + I2 \quad (1.21)$$

Ahora bien de (1.18), tenemos que

$$\begin{aligned} & \Pr \left(\|X^*\|^2 > t \text{ y } X \in R_0 \right) \\ &= \Pr \left(\|X^*\|^2 > t \right) - \Pr \left(\|X^*\|^2 > t \text{ y } X \in C \cup C^P \right) \\ &= \Pr \left(\|X^*\|^2 > t \right) - \Pr \left(\|X^*\|^2 > t \text{ y } X \in C \right) \\ &= 1 - \sum_{j=0}^k \delta_j \Psi_j(t) - \left(\Pr(X \in C) - \Pr \left(\|X^*\|^2 \leq t \text{ y } X \in C \right) \right) \\ &= 1 - \sum_{j=0}^k \delta_j \Psi_j(t) - (\delta_k - \delta_k \Psi_k(t)) \\ &= 1 - (\delta_0 + \delta_k) - \sum_{j=1}^{k-1} \delta_j \Psi_j(t) = \sum_{j=1}^{k-1} \delta_j (1 - \Psi_j(t)) \end{aligned} \quad (1.22)$$

Con (1.22) podemos calcular $I2$:

$$\begin{aligned}
 I2 &= \cos^2 \omega \cdot E \left(1_{R_0} (X) \cdot \|X^*\|^2 \right) \\
 &= \cos^2 \omega \cdot \int_0^\infty \Pr \left(X \in R_0 \text{ y } \|X^*\|^2 > t \right) dt \\
 &= \cos^2 \omega \cdot \int_0^\infty \left(\sum_{j=1}^{k-1} \delta_j (1 - \Psi_j(t)) \right) dt \\
 &= \cos^2 \omega \cdot \left(\sum_{j=1}^{k-1} \delta_j \cdot E \left(\chi_j^2 \right) \right) \\
 &= \cos^2 \omega \cdot \left(\sum_{j=1}^{k-1} \delta_j \cdot j \right)
 \end{aligned}$$

Ahora para $I1$ de los lemas 1.10 y 1.11 tenemos:

$$\begin{aligned}
 I1 &= \int_C (\underline{c}'X)^2 f(x) dx = \iint_{u \geq (u^2+v^2)^{1/2} \cos \omega} u^2 \cdot q(u, v) du dv \\
 &= \int_0^\infty \int_0^{u \tan \omega} C_k u^2 \exp \left\{ -\frac{1}{2} (u^2 + v^2) \right\} \cdot v^{k-2} du dv
 \end{aligned} \tag{1.23}$$

■

Además se puede calcular completamente el valor de la integral $I1$. La expresión final que aparece en el siguiente lema se obtiene en el apéndice a esta sección.

Lema 1.13 *Cuando k es impar*

$$I1 = \frac{1}{2} - \cos^3 \omega \sum_{p=0}^{r-1} \frac{(2(r-p)-1)!}{2^{2(r-p)-1} \cdot ((r-p-1)!)^2} \cdot \text{sen}^{2(r-p-1)} \omega \tag{1.24}$$

donde $r = (k-1)/2$.

Cuando k es par y mayor que 2

$$I1 = \frac{1}{\pi} (\omega + \text{sen } \omega \cos \omega) - \cos^3 \omega \sum_{p=1}^{r-1} \frac{(r-p)! \cdot (r-p-1)! \cdot 2^{2(r-p)-1}}{(2(r-p)-1)! \cdot \pi} \cdot \text{sen}^{2(r-p)-1} \omega \tag{1.25}$$

donde $r = k/2$.

Si $k = 2$ obtenemos simplemente

$$I1 = \frac{1}{\pi} (\omega + \operatorname{sen} \omega \cos \omega) \quad (1.26)$$

Los lemas anteriores son necesarios para la obtención de los siguientes resultados relativos a X y X^* .

Teorema 1.14 Sea $C(\omega)$ un cono circular de ángulo ω en \mathfrak{R}^k .

Si $k < 4$ entonces $E(c'X^*)^2 < E(c'X)^2$

Si $k = 4$ existe $\omega \in (0, \pi/2)$ tal que $E(c'X^*)^2 > E(c'X)^2$.

Demostración. Hay que notar que $E(c'X)^2 = 1$ independientemente de la dirección c y de la dimensión del espacio con el que se trata, puesto que en esa situación el problema es invariante.

Para $k = 2$ de los lemas 1.12 y 1.13 se obtiene:

$$E(c'X^*)^2 = \frac{1}{\pi} (\omega + \operatorname{sen} \omega \cos \omega) + \frac{1}{2} \cos^2 \omega \quad (1.27)$$

y derivando con respecto a ω

$$\frac{\partial}{\partial \omega} E(c'X^*)^2 = \cos \omega \left(\frac{2}{\pi} \cos \omega - \operatorname{sen} \omega \right)$$

cuyo único cero en el intervalo $(0, \pi/2)$ es el punto

$$\omega_2 = \arctan \left(\frac{2}{\pi} \right)$$

utilizando aquí y en el resto de la sección la determinación $(-\pi/2, \pi/2)$ del arco tangente. Por otra parte mediante la derivada segunda comprobamos la

condición de máximo de ω_2

$$\frac{\partial^2}{\partial \omega^2} E(c'X^*)^2 = -\frac{4}{\pi} \cos \omega \sin \omega - \cos 2\omega < 0$$

y ahora $E(c'X^*)^2 \Big|_{\omega=\omega_2} = 0.6804 < 1$.

Si $k = 3$ entonces de los lemas 1.12 y 1.13 $E(c'X^*)^2 = \frac{1}{2} + \cos^2 \omega \sin \omega$

y utilizando los mismos argumentos

$$\frac{\partial}{\partial \omega} E(c'X^*)^2 = \cos \omega (\cos^2 \omega - 2 \sin^2 \omega)$$

la raíz de esta ecuación en el intervalo considerado es ahora

$$\omega_3 = \arctan \left(\frac{1}{\sqrt{2}} \right)$$

y como antes

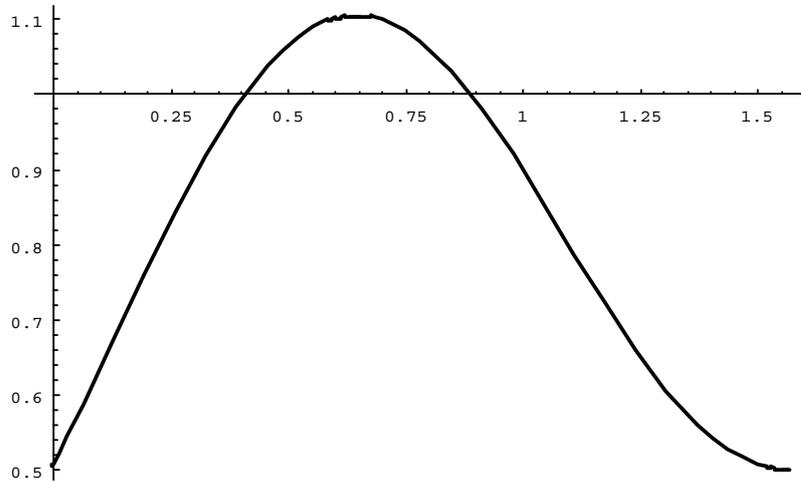
$$\frac{\partial^2}{\partial \omega^2} E(c'X^*)^2 = \sin \omega (-5 \cos^2 \omega - 2 \cos 2\omega) < 0$$

mientras que $E(c'X^*)^2 \Big|_{\omega=\omega_3} = 0.8849 < 1$.

En cambio cuando $k = 4$

$$E(c'X^*)^2 = \frac{1}{\pi} (\omega + \sin \omega \cos \omega) + \frac{2}{\pi} \cos^3 \omega \sin \omega + \cos^2 \omega \left(\frac{1}{2} + \sin^2 \omega \right)$$

En estas circunstancias la ecuación $\frac{\partial}{\partial \omega} E(c'X^*)^2 = 0$ no tiene una solución explícita sencilla pero utilizando métodos numéricos obtenemos que cuando $\omega = 0.413701$ y $\omega = 0.889535$ se tiene $E(c'X^*)^2 = 1$ y el máximo de esta función se alcanza en $\omega = 0.647884$ y es $E(c'X^*)^2 = 1.10363$. La gráfica de esta función puede observarse en la siguiente figura. ■



Error Cuadrático Medio en dimensión 4 bajo $\theta = 0$ para conos circulares de ángulo $0 \leq \omega \leq \pi/2$

Cabe preguntarse ahora si esta situación depende solamente de la dimensión o también del ángulo del cono que se considera. A ello responde el siguiente teorema

Teorema 1.15 *Para cualquier ángulo $\omega \in (0, \pi/2)$, existe $k \in \mathbb{N}$ tal que $E(c'X^*)^2 > E(c'X)^2$.*

Demostración. De (1.21) tenemos que

$$E(c'X^*)^2 \geq \cos^2 \omega \cdot \left(\sum_{j=1}^{k-1} \delta_j \cdot j \right)$$

Si consideramos $k = 2r$ es decir una dimensión par y nos quedamos solamente

con los sumandos $j = 2p + 1$ impares tenemos, puesto que $\delta_j > 0$

$$E(c'X^*)^2 \geq \cos^2 \omega \cdot \left(\sum_{\substack{j=1 \\ j \text{ impar}}}^{k-1} \delta_j \cdot j \right) \quad (1.28)$$

Ahora de (1.19)

$$\begin{aligned} 2\delta_j &= \frac{(k-2)\cdots(k-j-1)}{(j-1)!} \cdot \frac{(j/2-1)\cdots 1/2}{\left(\frac{k-3}{2}\right)\cdots\frac{k-j}{2}} \cdot \text{sen}^{j-1} \omega \cdot \cos^{k-j-1} \omega \\ &= \frac{(2(r-1))\cdots(2(r-p-1))}{2p \cdot (2p-1) \cdots 1} \cdot \frac{\frac{2p-1}{2} \cdot \frac{2p-3}{2} \cdots \frac{1}{2}}{\frac{2r-3}{2} \cdot \frac{2r-5}{2} \cdots \frac{2(r-p)-1}{2}} \cdot \text{sen}^{2p} \omega \cdot \cos^{2(r-p-1)} \omega \\ &= \frac{(r-1) \cdots (r-p)}{p!} \cdot \text{sen}^{2p} \omega \cdot \cos^{2(r-p-1)} \omega \\ &= \binom{r-1}{p} \cdot (\text{sen}^2 \omega)^p \cdot (\cos^2 \omega)^{r-p-1} \end{aligned}$$

Y entonces el lado derecho de (1.28) es

$$\begin{aligned} &\cos^2 \omega \left(\sum_{p=0}^{r-1} \binom{r-1}{p} (\text{sen}^2 \omega)^p \cdot (\cos^2 \omega)^{r-p-1} \right) \\ &= \cos^2 \omega \left((r-1) \text{sen}^2 \omega + \frac{1}{2} \right) \end{aligned} \quad (1.29)$$

puesto que el sumatorio no es más que $E\left(Z + \frac{1}{2}\right)$ donde Z es una variable aleatoria binomial de parámetros $r-1$ y $\text{sen}^2 \omega$.

El resultado se obtiene puesto que la expresión (1.29) tiende a ∞ cuando $r \rightarrow \infty$ para cualquier $\omega \in (0, \pi/2)$. ■

Hasta ahora hemos centrado nuestra atención en la dirección central del cono circular por dos razones esenciales, la primera porque es más manejable y la segunda por la supuesta condición de dirección más desfavorable que encontramos en el cono recto. Para evaluar esta conjetura hemos estudiado el cono circular en \mathfrak{R}^2 y hemos determinado cual es en este caso bajo $\theta = 0$ la dirección más desfavorable.

Lema 1.16 *Sea $d \in \mathfrak{R}^2$ unitario. Si ω_1 representa el ángulo que forman c y d :*

$$E(d'X^*)^2 = \frac{1}{\pi} (\omega + \cos \omega \sin \omega \cos 2\omega_1) + \frac{1}{2} (\cos^2 \omega_1 \cos^2 \omega + \sin^2 \omega \sin^2 \omega_1) \quad (1.30)$$

para $0 \leq \omega \leq \pi/2$ y $0 \leq \omega_1 \leq \pi/2$.

Demostración. Si en (1.27) denotamos $g(\omega) = \frac{1}{\pi} (\omega + \sin \omega \cos \omega)$ y además llamamos $h(\omega_1, \omega) = E(d'X^*)^2$ entonces $E(c'X^*)^2 = h(0, \omega)$. Tenemos entonces que si d es una dirección interior a C podemos escribir:

$$h(\omega_1, \omega) = \frac{1}{2} (h(0, \omega + \omega_1) + h(0, \omega - \omega_1)) \quad (1.31)$$

Y cuando la dirección que se considera d es exterior al cono:

$$h(\omega_1, \omega) = \frac{1}{2} (g(\omega + \omega_1) - g(\omega_1 - \omega)) + \frac{1}{4} (\cos^2(\omega_1 + \omega) + \cos^2(\omega_1 - \omega))$$

y como $g(-\omega) = -g(\omega)$ esta última expresión queda:

$$\frac{1}{2} (h(0, \omega + \omega_1) + h(0, \omega - \omega_1))$$

es decir que obtenemos de nuevo (1.31).

Hay que notar aquí que a pesar de que $0 \leq \omega + \omega_1 \leq \pi$ la expresión que define $g(\omega)$ sigue siendo válida en esta situación.

Tenemos por tanto tras algunos cálculos

$$h(\omega_1, \omega) = \frac{1}{2\pi} (\omega + \sin \omega \cos \omega \cos 2\omega_1) + \frac{1}{4} (\cos^2(\omega + \omega_1) + \cos^2(\omega - \omega_1))$$

de donde se obtiene (1.30). ■

Teorema 1.17 *Sea $C(c, \omega)$ un cono circular en \mathfrak{R}^2 de ángulo ω y dirección central c . Entonces*

$$E_0 (c' X^*)^2 > E_0 (d' X^*)^2, \forall d \in \mathfrak{R}^2$$

si y solamente si

$$\omega < \omega_0 = \frac{1}{2} \left(\pi + \arctan \left(-\frac{\pi}{2} \right) \right) = 1.0688$$

Demostración. Es suficiente derivar (1.30) respecto ω_1 con lo que obtenemos tras algunos cálculos:

$$\frac{\partial}{\partial \omega_1} h(\omega_1, \omega) = \sin 2\omega_1 \left(-\frac{1}{\pi} \sin 2\omega - \frac{1}{2} \cos 2\omega \right)$$

de donde se obtiene inmediatamente el enunciado del teorema puesto que el único 0 de $-\frac{1}{\pi} \sin 2\omega - \frac{1}{2} \cos 2\omega$ para $0 \leq \omega \leq \pi/2$ es claramente ω_0 y esa función es positiva para $\omega > \omega_0$ y negativa para $\omega < \omega_0$. ■

Nota 1.6 *Los conos descritos por el teorema anterior incluyen todos los conos agudos de \mathfrak{R}^2 .*

Nota 1.7 *Cuando $\omega > \omega_0 = \frac{1}{2} \left(\pi + \arctan \left(-\frac{\pi}{2} \right) \right)$ la dirección más desfavorable para la estimación es la ortogonal al eje, o sea $\arg \max_d E (d' X^*)^2 = c^\perp$. Cuando $\omega = \omega_0$ entonces $E (d' X^*)^2$ es constante como función de d .*

1.3.1 Prueba del lema 1.13

Demostración. De (1.23) tenemos que

$$\begin{aligned} I1 &= \int_0^\infty \int_0^{u \tan \omega} C_k u^2 \exp \left\{ -\frac{1}{2} (u^2 + v^2) \right\} \cdot v^{k-2} du dv \\ &= C_k \int_0^\infty u^2 \exp \left\{ -\frac{1}{2} u^2 \right\} \left(\int_0^{u \tan \omega} v^{k-2} \exp \left\{ -\frac{1}{2} v^2 \right\} dv \right) du \end{aligned}$$

y si denotamos $f(u, \omega, k) = \int_0^{u \tan \omega} v^{k-2} \exp\left\{-\frac{1}{2}v^2\right\} dv$, integrando por partes tenemos

$$\begin{aligned} f(u, \omega, k) &= \\ &= -v^{k-3} e^{-(1/2)v^2} \Big|_0^{u \tan \omega} + \int_0^{u \tan \omega} (k-3) v^{k-4} \exp\left\{-\frac{1}{2}v^2\right\} dv \\ &= -(u \tan \omega)^{k-3} e^{-(1/2)(u \tan \omega)^2} + \int_0^{u \tan \omega} (k-3) v^{k-4} \exp\left\{-\frac{1}{2}v^2\right\} dv \end{aligned}$$

e iterando el proceso se obtiene que cuando k es impar y $r = (k-1)/2$ la integral $f(u, \omega, k)$ es igual a

$$-\sum_{p=0}^{r-1} 2^p (u \tan \omega)^{2(r-(p+1))} \exp\left\{-\frac{1}{2}(u \tan \omega)^2\right\} \frac{r-1!}{r-(p+1)!} + 2^{r-1} (r-1)!$$

y cuando k es par, $r = k/2$ y $k > 2$

$$\begin{aligned} &-\sum_{p=0}^{r-1} (u \tan \omega)^{k-1-2p} \exp\left\{-\frac{1}{2}(u \tan \omega)^2\right\} \frac{k-3!}{k-2p!} \frac{r-p-1!}{2^p(r-1)!} + \\ &+ \frac{\sqrt{2\pi}(k-3)!}{2^{r-1}(r-1)!} \left(\Phi(u \tan \omega) - \frac{1}{2}\right) \end{aligned}$$

Si $k = 2$ obtenemos simplemente $f(u, \omega, 2) = \sqrt{2\pi} \left(\Phi(u \tan \omega) - \frac{1}{2}\right)$.

Ahora para finalizar los cálculos vamos a necesitar las integrales:

$$\int_0^\infty \exp\left\{-\frac{1}{2}t^2\right\} t^{2r} dt = \sqrt{\frac{\pi}{2}} \frac{(2r-1)!}{2^{r-1}(r-1)!} \quad (1.32)$$

y

$$\int_0^\infty \exp\left\{-\frac{1}{2}t^2\right\} t^{2r+1} dt = 2^r \cdot r! \quad (1.33)$$

Entonces en el caso impar $I1 = C_k \int_0^\infty u^2 \exp\left\{-\frac{1}{2}u^2\right\} f(u, \omega, k) du$ queda

$$\begin{aligned} &C_k \left[-\sum_{p=0}^{r-1} 2^p (\tan \omega)^{2(r-(p+1))} \frac{r-1!}{r-(p+1)!} \left(\int_0^\infty u^{2(r-p)} \exp\left\{-\frac{1}{2}u^2(1+\tan^2 \omega)\right\} du \right) \right] + \\ &+ C_k 2^{r-1} (r-1)! \sqrt{\frac{\pi}{2}} \end{aligned}$$

donde $r = (k-1)/2$.

Haciendo ahora un cambio de variable tenemos que

$$\int_0^\infty u^{2(r-p)} \exp \left\{ -\frac{1}{2}u^2 (1 + \tan^2 \omega) \right\} du = \int_0^\infty t^{2(r-p)} \exp \left\{ -\frac{1}{2}t^2 \right\} \cos^{2(r-p)+1} \omega dt$$

y utilizando (1.20) y (1.32) obtenemos (1.24).

En el caso par para obtener $I1 = C_k \int_0^\infty u^2 \exp \left\{ -\frac{1}{2}u^2 \right\} f(u, \omega, k) du$ es necesario calcular

$$C_k \left[-\sum_{p=0}^{r-1} (\tan \omega)^{k-1-2p} \frac{k-3!}{k-2p!} \frac{r-p-1!}{2^p(r-1)!} \left(\int_0^\infty \exp \left\{ -\frac{1}{2}u^2 (1 + \tan^2 \omega) \right\} u^{2(r-p)+1} du \right) \right] + C_k \frac{\sqrt{2\pi}(k-3)!}{2^{r-1}(r-1)!} \left(\int_0^\infty \left(\Phi(u \tan \omega) - \frac{1}{2} \right) u^2 \exp \left\{ -\frac{1}{2}u^2 \right\} du \right)$$

donde $r = k/2$.

Basta ahora hacer el mismo cambio de variable que en el caso impar y utilizar esta vez (1.33). Finalmente

$$\int_0^\infty \Phi(u \tan \omega) u^2 \exp \left\{ -\frac{1}{2}u^2 \right\} du = \frac{1}{\sqrt{2\pi}} \left(\frac{\pi}{2} + \omega + \cos \omega \sin \omega \right)$$

y (1.20) de nuevo nos ayuda a conseguir (1.25).

Para el caso $k = 2$ tenemos

$$\begin{aligned} I1 &= c_2 \int_0^\infty \sqrt{2\pi} \left(\Phi(u \tan \omega) - \frac{1}{2} \right) u^2 \exp \left\{ -\frac{1}{2}u^2 \right\} du \\ &= \frac{1}{\pi} \left(\frac{\pi}{2} + \omega + \cos \omega \sin \omega - \frac{\sqrt{2\pi}}{2} \int_0^\infty u^2 \exp \left\{ -\frac{1}{2}u^2 \right\} du \right) \\ &= \frac{1}{\pi} (\omega + \cos \omega \sin \omega) \end{aligned}$$

con lo que obtenemos (1.26). ■

1.4 Conos de orden

Veamos primero algunas propiedades de los estimadores cuando las restricciones están definidas por un cono convexo y cerrado general. Sean por tanto

$X_i \rightsquigarrow N(\theta_i, \sigma^2)$ independientes verificando $\theta = (\theta_1, \dots, \theta_k) \in C$ cono convexo y cerrado. Denotamos como hasta ahora por $X^* = P(X/C)$ es el EMV bajo las restricciones impuestas. Enumeramos ahora una serie de lemas de interés que nos van a ser de utilidad. Sus demostraciones pueden encontrarse en las referencias originales correspondientes.

Lema 1.18 (Raubertas, Lee y Nordheim (86)) Sean C_1 y C_2 conos convexos y cerrados en \mathfrak{R}^k con $C_2 \subset C_1$. Entonces si C_1 o C_2 es un subespacio lineal

$$P(P(x/C_1)/C_2) = P(x/C_2)$$

para todo $x \in \mathfrak{R}^k$.

Lema 1.19 (Menéndez, Rueda y Salvador (92)) Sean C y L un cono convexo y cerrado y un subespacio lineal respectivamente. Para cada $x \in \mathfrak{R}^k$, la relación

$$P(P(x/C)/L) = P(P(x/L)/C) \quad (1.34)$$

se verifica si y solamente si x satisface las tres condiciones siguientes

$$P(P(x/L)/C) \in L \quad (1.35)$$

$$P(P(x/C)/L) \in C \quad (1.36)$$

$$P(P(x/C)/C \cap L) = P(x/C \cap L) \quad (1.37)$$

Lema 1.20 (Menéndez, Rueda y Salvador (92)) Sean C y L verificando (1.34). Entonces:

$$P(P(x/C)/L^\perp) = P(P(x/L^\perp)/C) = P(x/L^\perp \cap C), \quad \forall x \in \mathfrak{R}^k$$

Hay que notar que el lema 1.18 en su enunciado original solo se refiere a conos poliédricos pero como ya se anuncia en Menéndez, Rueda y Salvador (92) el resultado es válido también para conos convexos y cerrados. Estos lemas previos vamos a utilizarlos para demostrar la siguiente proposición.

Proposición 1.21 1) $\forall d \in L_S(C)$ se tiene que $d'X = d'X^*$.

2) Si $d \in L_S(C)$ y $e \in L_S(C)^\perp$ entonces $d'X^*$ y $e'X^*$ son independientes.

Demostración. Para la primera parte tenemos

$$X = P(X/L_S(C)) + P(X/L_S(C)^\perp)$$

y del primeros de los lemas anteriores

$$X^* = P(X/C) = P(X/L_S(C)) + P(P(x/C)/L_S(C)^\perp) \quad (1.38)$$

con lo cual como $d \in L_S(C)$

$$d'X = d'P(X/L_S(C)) = d'X^*$$

Para probar la segunda parte construimos una base ortogonal de \mathfrak{R}^k de tal forma que los primeros elementos de dicha base sean a su vez una base de $L_S(C)$ y el resto lo sean de $L_S(C)^\perp$. Sea $B = \{c_1, \dots, c_k\}$ dicha base de tal modo que si la dimensión de $L_S(C)$ es p , $\{c_1, \dots, c_p\}$ es una base de $L_S(C)$ y $\{c_{p+1}, \dots, c_k\}$ es base ortogonal de $L_S(C)^\perp$. Entonces si el punto x tiene coordenadas (x_1, \dots, x_k) en esta base tenemos del apartado anterior que $d'X^*$ solo depende de (x_1, \dots, x_p) .

Por otra parte del lema 1.18 obtenemos que en estas circunstancias se verifica (1.35) y (1.36) y (1.37) son obvios puesto que $L_S(C) \subset C$. Entonces de los lemas 1.19 y 1.20 obtenemos

$$P\left(P(x/C)/L_S(C)^\perp\right) = P\left(P(x/L_S(C)^\perp)/C\right) = P\left(x/C \cap L_S(C)^\perp\right)$$

y por tanto considerando (1.38)

$$e'x^* = e'P\left(P(x/L_S(C)^\perp)/C\right)$$

donde $P\left(P(x/L_S(C)^\perp)/C\right)$ es un vector cuyas componentes no nulas dependen solamente de x_{p+1}, \dots, x_k puesto que claramente $P\left(x/L_S(C)^\perp\right) = (0, \dots, 0, x_{p+1}, \dots, x_k)$.

Por tanto $e'X^*$ depende solamente de X_{p+1}, \dots, X_k y es independiente de $d'X^*$.

■

Podemos ahora utilizar esta proposición para probar el siguiente

Teorema 1.22 *Sea $d \in \mathfrak{R}^k$ y sea $e = P\left(d/L_S(C)^\perp\right)$ entonces*

$$E(d'X - d'\theta)^2 - E(d'X^* - d'\theta)^2 = E(e'X - e'\theta)^2 - E(e'X^* - e'\theta)^2$$

Demostración. Si denotamos por $f = P(d/L_S(C))$ entonces $d = e + f$ y

$$E(d'X - d'\theta)^2 = E(e'X - e'\theta)^2 + E(f'X - f'\theta)^2$$

mientras que

$$E(d'X^* - d'\theta)^2 = E(e'X^* - e'\theta)^2 + E(f'X^* - f'\theta)^2 + E((e'X^* - e'\theta)(f'X^* - f'\theta))$$

entonces utilizando la proposición 1.21 tenemos

$$E((e'X^* - e'\theta)(f'X^* - f'\theta)) = E(e'X^* - e'\theta)E(f'X - f'\theta) = 0$$

$$E(f'X^* - f'\theta)^2 = E(f'X - f'\theta)^2$$

y obtenemos el resultado. ■

Tratamos ahora más específicamente el caso de los conos de orden, es decir que θ pertenece a un cono de orden. Sea $c \in L_S(C)^\perp$, $\|c\| = 1$ la dirección central del cono de orden. Esta dirección tiene la propiedad de que el máximo ángulo entre c y las direcciones generadoras del cono es mínimo. Abelson y Tukey (63) determinaron la existencia y unicidad de esta dirección para situaciones como las aquí tratadas.

1.4.1 Tree Order

Veamos ahora que sucede en el caso del tree order, es decir que suponemos que el vector de medias verifica $\theta \in C' = \{\theta_0 \leq \theta_i \text{ para } i = 1, \dots, k-1\}$. La dirección central del tree order ha sido calculada por Robertson et al. (88) y resulta ser $c = \lambda(k-1, -1, \dots, -1)$ con $\lambda = \|(k-1, -1, \dots, -1)\|^{-1}$.

Tenemos entonces:

Teorema 1.23

$$E(c'(X^* - \theta))^2 > E(c'(X - \theta))^2$$

si y solamente si

$$E(X_0^* - \theta_0)^2 > E(X_0 - \theta_0)^2 \tag{1.39}$$

Demostración. Es inmediato comprobar que

$$\begin{aligned} a &= (1, 0, \dots, 0) = P(a/L_S(C')) + P(a/L_S(C')^\perp) \\ &= \frac{1}{k}(1, \dots, 1) + \frac{1}{k}(k-1, -1, \dots, -1) = c_1 + c_2 \end{aligned}$$

El teorema 1.22 resuelve entonces la cuestión. ■

Nota 1.8 Según las simulaciones de Lee (88) si los tamaños muestrales en cada población son iguales (1.39) ocurre ya cuando $k = 9$ mientras que si en la población correspondiente a la raíz del árbol se toma un tamaño muestral $n_0 > 3.5n$ donde n es el tamaño común a las demás poblaciones (1.39) aparece cuando $k = 6$.

1.4.2 Orden Total

Tratamos ahora el caso del orden total es decir que suponemos ahora que el vector de medias $\theta \in C = \{\theta_1 \leq \theta_2 \leq \dots \leq \theta_k\}$ y que además $\omega_i = \omega_j$ para cualesquiera i, j . En esta situación la dirección central, para la que no existe una formula cerrada sencilla, fue calculada por Abelson y Tukey (63) quienes además determinaron los valores de

$$r_{\min}^2 = \min \left\{ \cos^2(c, d) \right\}, \text{ para } d \in \left\{ \text{Generadores de } C \cap L_S(C)^\perp \right\} \quad (1.40)$$

para distintos valores de k .

Vamos a probar que si la dimensión del espacio que consideramos k es suficientemente grande $c'X^*$ tiene mayor error cuadrático medio que $c'X$ para estimar $c'\theta$ cuando $\theta = 0 \in C$. Para ello van a ser necesarios dos lemas que pasamos a demostrar previamente al resultado anunciado.

Lema 1.24 *Si representamos por X_A^* representa la proyección de X sobre el cono A .*

$$E \left(\left\| X_{L_S(C)^\perp \cap C}^* \right\|_k^2 \right) = \sum_{a=2}^k \frac{1}{a}$$

Demostración. $\left\| X_{L_S(C)^\perp \cap C}^* \right\|_k^2 = \left\| X_{L_S(C)}^* - X_C^* \right\|^2$ (lemas previos) es el estadístico Razón de Verosimilitudes para contrastar la linealidad contra el cono del orden total y tiene una distribución bajo $\theta = 0$ Chi-bar-squared (Robertson et al. (88)). Es decir que

$$E \left(\left\| X_{L_S(C)^\perp \cap C}^* \right\|_k^2 \right) = \sum_{l=1}^k \Pr(l, k) E \left(\chi_{l-1}^2 \right) \quad (1.41)$$

donde

$$\Pr(1, k) = \frac{1}{k} \quad \Pr(k, k) = \frac{1}{k!} \quad y$$

$$\Pr(l, k) = \frac{1}{k} \Pr(l-1, k-1) + \frac{k-1}{k} \Pr(l, k-1) \quad \text{para } l = 2, \dots, k-1$$

cuando $\omega_i = \omega_j$ que es el caso que estamos considerando (Robertson et al. (88), Corolario A, pag. 81).

De Robertson et al. (88) pag. 444 es claro que

$$E \left(\left\| X_{L_S(C)^\perp \cap C}^* \right\|_3^2 \right) = \frac{1}{2} + \frac{1}{3}$$

Por otra parte desarrollando la expresión a la derecha de (1.41) se obtiene

$$E \left(\left\| X_{L_S(C)^\perp \cap C}^* \right\|_k^2 \right) = \sum_{l=1}^k \Pr(l, k) (l-1)$$

y utilizando los valores de $\Pr(l, k)$

$$\begin{aligned}
&= \Pr(k, k) (k-1) + \sum_{l=2}^{k-1} \Pr(l, k) (l-1) \\
&= \frac{k-1}{k!} + \sum_{l=2}^{k-1} \left[\frac{1}{k} \Pr(l-1, k-1) + \frac{k-1}{k} \Pr(l, k-1) \right] (l-1) \\
&= \frac{k-1}{k!} + \sum_{l=2}^{k-1} \frac{1}{k} \Pr(l-1, k-1) (l-1) + \sum_{l=2}^{k-1} \frac{k-1}{k} \Pr(l, k-1) (l-1) \\
&= \frac{k-1}{k!} + \frac{1}{k} \left[\sum_{l=2}^{k-1} \Pr(l-1, k-1) (l-2) + \Pr(l-1, k-1) \right] + \frac{k-1}{k} E \left\| X_{L_S(C)^+ \cap C}^* \right\|_{k-1}^2 \\
&= \frac{1}{k!} + \frac{1}{k} \left[E \left(\left\| X_{L_S(C)^+ \cap C}^* \right\|_{k-1}^2 \right) + \left(1 - \frac{1}{(k-1)!} \right) \right] + \frac{k-1}{k} E \left(\left\| X_{L_S(C)^+ \cap C}^* \right\|_{k-1}^2 \right) \\
&= E \left(\left\| X_{L_S(C)^+ \cap C}^* \right\|_{k-1}^2 \right) + \frac{1}{k}
\end{aligned}$$

con lo que actuando por inducción sobre k resolvemos el problema. \blacksquare

Lema 1.25 Para r_{\min}^2 definido en (1.40) se tiene que

$$r_{\min}^2 > \frac{2}{3 + \ln(k-1)} \quad (1.42)$$

Demostración. De Abelson y Tukey (63) sabemos que $r_{\min}^2 = \frac{1}{\sum_{j=1}^k d_j^2}$

donde

$$d_j = \{(j-1)[1 - ((j-1)/k)]\}^{\frac{1}{2}} - \{j(1 - j/k)\}^{\frac{1}{2}}$$

es decir $d_j = z(j-1) - z(j)$ donde

$$z(t) = \{t(1 - t/k)\}^{\frac{1}{2}} \text{ para } 0 \leq t \leq k$$

Notar además que

$$-d_{j+1} = d_{k-j} \text{ para } j = 0, \dots, k-1 \text{ y } d_{[k/2]+1} = 0 \text{ para } k \text{ impar} \quad (1.43)$$

Como

$$z'(t) = \frac{1 - 2t/k}{2 \{t(1 - t/k)\}^{\frac{1}{2}}}$$

verifica $z'(t) \geq 0$ cuando $1 \leq t \leq k/2$ y $z'(t) < 0$ cuando $k/2 < t \leq k$ y además

$$z''(t) = \frac{-\frac{2}{k} \left[2 \{t(1-t/k)\}^{\frac{1}{2}} \right] - 2(1-2t/k) z'(t)}{4 \{t(1-t/k)\}} < 0$$

tenemos que $d_j = z(j-1) - z(j) < z'(j-1)$. Teniendo en cuenta (1.43)

$$\begin{aligned} \sum_{j=1}^k d_j^2 &< 2d_1^2 + 2 \sum_{j=2}^{[k/2]} \frac{(1-2(j-1)/k)^2}{4 \{t(1-(j-1)/k)\}} \\ &< 2d_1^2 + 2 \frac{(2/k-1)^2}{4(1-1/k)} + 2 \int_1^{[k/2]} \frac{(1-2t/k)^2}{4 \{t(1-t/k)\}} dt \\ &\leq 2d_1^2 + \frac{(2-k)^2}{2k(k-1)} + \int_1^{k-1} \frac{(1-2t/k)^2}{4 \{t(1-t/k)\}} dt \\ &= \frac{(2-k)^2}{2k(k-1)} + 1 + \frac{1}{2} \ln(k-1) \\ &< \frac{3}{2} + \frac{1}{2} \ln(k-1) \end{aligned}$$

de donde se deduce (1.42). ■

Teorema 1.26 *Para $\theta = 0$ existe k_0 tal que para cualquier $k \geq k_0$ se verifica que $E(c'X^*)^2 > E(c'X)^2$.*

Demostración. Como $c \in L_S(C)^\perp$ (demostración de la Proposición 1.21, pag. 60)

$$c'X^* = c'X_{L_S(C)^\perp \cap C}^* \quad (1.44)$$

Por otro lado de (1.40) tenemos que

$$\forall x \in C \cap L_S(C)^\perp, \frac{\langle c, x \rangle}{\|x\|^2} \geq r_{\min}^2 \quad (1.45)$$

Entonces de (1.44) y (1.45)

$$\begin{aligned} E(c'X^*)^2 &= E\left(c'X_{L_S(C)^\perp \cap C}^*\right)^2 = E\left(\frac{(c'X_{L_S(C)^\perp \cap C}^*)^2}{\|X_{L_S(C)^\perp \cap C}^*\|_k^2} \cdot \|X_{L_S(C)^\perp \cap C}^*\|_k^2\right) \\ &\geq r_{\min}^2 \cdot E\left(\|X_{L_S(C)^\perp \cap C}^*\|_k^2\right) \end{aligned} \quad (1.46)$$

y ahora utilizando los lemas 1.24 y 1.25 en (1.46) y la conocida desigualdad

$$\ln k < \sum_{a=1}^k \frac{1}{a}$$

tenemos que

$$E(c'X^*)^2 > \frac{2}{3 + \ln(k-1)} \left(\sum_{a=2}^k \frac{1}{a}\right) > \frac{\left(\sum_{a=1}^k \frac{1}{a}\right) + 3 + \left(\sum_{a=2}^k \frac{1}{a}\right) - 4}{3 + \left(\sum_{a=1}^k \frac{1}{a}\right)} \quad (1.47)$$

y (1.47) es mayor que $1 = E(c'X)^2$ cuando

$$\sum_{a=2}^k \frac{1}{a} > 4$$

■

A la vista del teorema 1.22 podemos señalar otras direcciones d en las que el error cuadrático medio utilizando $d'X^*$ como estimador de $d'\theta$ va a ser superior para k suficientemente grande y $\theta = 0$ que si se utiliza $d'X$.

Corolario 1.27 Para $\theta = 0$ y $k \geq k_0$, $E(d'X^*)^2 > E(d'X)^2$ si $d = c + \lambda(1, \dots, 1)$, $\lambda \in \mathfrak{R}$.

Demostración. Evidente de los Teoremas 1.22 y 1.26 puesto que $L_S(C) = \{\lambda(1, \dots, 1)\}$. ■

Nota 1.9 *Utilizando los valores exactos para r_{\min}^2 que aparecen en Abelson y Tukey pág. 1358 tenemos que cuando*

$$k = 15, r_{\min}^2 = .433 \text{ y por tanto } E(c'X^*)^2 > r_{\min}^2 \cdot E\left(\|X_{L_S(C)^\perp}^*\|_{15}^2\right) = 1.00379$$

mientras que si

$$k = 14, r_{\min}^2 = .439 \text{ y por tanto } E(c'X^*)^2 > r_{\min}^2 \cdot E\left(\|X_{L_S(C)^\perp}^*\|_{14}^2\right) = 0.988436$$

Por otra parte para precisar más la dimensión en la que se produce este efecto se efectuaron simulaciones en las condiciones que venimos manejando; esto es en la dirección central y con $\underline{\theta} = 0$. Los resultados obtenidos para el ECM tras 100.000 repeticiones del experimento tomando $X_i \rightsquigarrow N(0, 1)$ y los valores de los coeficientes del vector unitario de la dirección central del cono aparecen en la tabla.

Los resultados indican que ese efecto se produce cuando $k = 7$ pero no son lo suficientemente claros para $k = 6$ como para tomar una decisión al respecto.

j	Valores de c_j						
	k=4	k=5	k=6	k=7	k=8	k=9	k=10
1	-.698	-.690	-.682	-.674	-.668	-.662	-.656
2	-.108	-.155	-.181	-.196	-.206	-.214	-.218
3	.108	.000	-.052	-.083	-.103	-.117	-.127
4	.698	.155	.052	.000	-.032	-.053	-.069
5		.690	.181	.083	.032	.000	-.022
6			.682	.196	.103	.053	.022
7				.674	.206	.117	.069
8					.668	.214	.127
9						.662	.218
10							.656
ECM	.813	.921	.999	1.074	1.140	1.200	1.253

Tabla 1: Simulación en la dirección central del orden total para diferentes dimensiones

Capítulo 2

Estimación Simultánea de Coordenadas

Este capítulo está dedicado a la estimación simultánea de combinaciones lineales del parámetro en un modelo normal restringido. Concretamente utilizamos el criterio de la probabilidad de cubrimiento, que es más fuerte que el del error cuadrático medio y que ha sido tratado en artículos recientes por Hwang (85), Kelly (89), en Hwang y Peddada (94) y Rueda y Salvador (95) quienes prueban que cuando se tiene una única restricción la probabilidad de cubrimiento de los EMV de cada una de las coordenadas es superior a la de los estimadores estandar.

Primeramente consideramos el caso en el que existe una única restricción, es decir aquel en el que el conjunto de valores posibles del parámetro es un semiespacio $C = \{\theta \in \mathfrak{R}^k : a'\theta \geq 0\}$, y obtenemos un resultado muy general bajo hipótesis de normalidad. Concretamente en la sección 2.1 probamos que Para cualquier $A \subset \mathfrak{R}^k$ convexo y simétrico con respecto al origen

$$\Pr(X - \theta \in A) \leq \Pr(X^* - \theta \in A)$$

Desarrollamos en el contexto del modelo lineal aplicaciones a los intervalos de confianza simultaneos habituales entre los que se incluyen los intervalos del tipo de Scheffé, Tukey o Bonferroni o incluso los más generales de Bowden (70) y a cuestiones relacionadas con la admisibilidad de los estimadores obteniendose que según el criterio de Hwang (85) X^* U-domina a X .

En la sección segunda tratamos el problema cuando las restricciones son de orden. Desafortunadamente en estas condiciones no pueden conseguirse resultados tan generales como en el caso anterior (obviamente la situación es bastante más complicada ahora). Sin embargo probamos otros resultados parciales que, a nuestro juicio, son de interés. Primeramente demostramos una propiedad general que curiosamente no se había encontrado hasta ahora, probamos que para cualquier cono de orden y cualquier situación en la que el algoritmo Minimum Lower Sets de cálculo del EMV sea aplicable $A_t \subset A_t^*$ para cualquier $t > 0$ donde

$$\begin{aligned} A_t^* &= \{x \in \mathfrak{R}^k : |x_i^* - \theta_i| \leq t, i = 1, \dots, k\} \\ A_t &= \{x \in \mathfrak{R}^k : |x_i - \theta_i| \leq t, i = 1, \dots, k\} \end{aligned}$$

es decir que los intervalos de confianza de longitud fijada y la misma para todas las coordenadas tienen mayor probabilidad de cubrimiento si se centran en X^* en lugar de en el EMV no restringido X . Intentamos también la extensión de esta propiedad de contención a otras situaciones diferentes a los conos de orden obteniendo otras en \mathfrak{R}^2 donde se verifica alguna similar. Seguidamente notamos que la contención no se mantiene para los conos de orden si los valores

de t varían en cada coordenada e intentamos encontrar para que situaciones puede seguir asegurándose que

$$\Pr (|X_i - \theta_i| \leq t_i, i = 1, \dots, k) \leq \Pr (|X_i^* - \theta_i| \leq t_i, i = 1, \dots, k)$$

Desarrollamos una técnica que nos permite, en un modelo normal, dar una respuesta parcial a esta cuestión encontrando determinados tipos de restricciones, que incluyen siempre al orden total, y determinadas configuraciones de los t_i para los que puede demostrarse la desigualdad anterior. En el caso del orden total el resultado óptimo hubiera sido el probar la desigualdad para valores cualesquiera de los t_i . No hemos podido sin embargo conseguir una demostración analítica de este resultado aunque como puede comprobarse en las simulaciones que ofrecemos como final de este capítulo todo parece indicar que dicho resultado es efectivamente cierto.

2.1 Caso de un semiespacio

Sea $C = \{\theta \in \mathfrak{R}^k : a'\theta \geq 0\}$. Suponiendo un modelo normal, vamos a probar que, bajo este tipo de restricción, cuando utilizamos como región de confianza un conjunto convexo y simétrico con respecto al valor de un cierto estimador la probabilidad de cubrimiento simultánea de dicho conjunto es mayor cuando se utiliza como valor central el Estimador Máximo Verosímil en lugar del Estimador que no tiene en cuenta la información adicional proporcionada por la restricción. Para la demostración necesitaremos un par de lemas. La prueba del primero de estos lemas puede encontrarse en Tong (90), Teorema 2.4.4

pag.70.

Lema 2.1 Tong (1990)

Sea $X \rightsquigarrow N_k(\theta, \Sigma)$, Σ definida positiva. Entonces para cualquier número entero positivo m , conjuntos A_1, \dots, A_m de \mathfrak{R}^k , y números reales $\alpha_1, \dots, \alpha_m$ tales que $\alpha_i \geq 0$ y $\sum \alpha_i = 1$, tenemos que

$$\Pr \left(X \in \sum_{i=1}^m \alpha_i A_i \right) \geq \prod_{i=1}^m [\Pr (X \in A_i)]^{\alpha_i}.$$

Lema 2.2 Supongamos que $X \rightsquigarrow N_k(\theta, \Sigma)$, Σ definida positiva y conocido y θ verificando, $a'\theta \geq 0$ donde $a \in \mathfrak{R}^k$. Entonces para cualquier $A \subset \mathfrak{R}^k$ convexo y simétrico respecto al origen tenemos que

Demostración.

$$\Pr (X - \theta \in A / a'X = z) \text{ es no decreciente para } z \leq 0$$

Sea V la raíz cuadrada positiva y simétrica de Σ^{-1} y sea $Z = V(X - \theta)$ con lo que tenemos

$$\begin{aligned} \Pr_{\theta} (X - \theta \in A / a'X = z) &= \Pr_{\theta} (X - \theta \in A / a'(X - \theta) = z - a'\theta) \\ &= \Pr_0 (Z \in B / b'Z = z') \end{aligned}$$

Donde $B = VA$ es un conjunto convexo y simétrico, $b = V^{-1}a$ y $z' = z - a'\theta \leq 0$, puesto que $z \leq 0$ y $a'\theta \geq 0$. Es suficiente entonces demostrar la propiedad para $\theta = 0$ y $\Sigma = I$. Notar que, en este caso la función $g(z)$ que definimos en (2.1) es simétrica.

Consideramos entonces la transformación ortogonal dada por la matriz M tal que, $Z = MX \rightsquigarrow N_k(0, I)$ y además $Z_1 = a'X$, y definimos $C = MA$. Entonces:

$$g(z) = \Pr(Z \in C / Z_1 = z) = \Pr((Z_2, \dots, Z_k) \in C_z) \quad (2.1)$$

donde $C_z = \{(z_2, \dots, z_k) \in \mathfrak{R}^{k-1} / (z, z_2, \dots, z_k) \in C\}$

Sea z' con $z < z' \leq 0$. Entonces existe $\lambda \in (0, 1)$ tal que $z' = \lambda z + (1 - \lambda)(-z)$. Si consideramos el conjunto $B_{z'} = \lambda C_z + (1 - \lambda)C_{-z}$, de la convexidad de C podemos deducir inmediatamente que $B_{z'} \subset C_{z'}$. Por tanto

$$g(z') = \Pr((Z_2, \dots, Z_k) \in C_{z'}) \geq \Pr((Z_2, \dots, Z_k) \in B_{z'}) \quad (2.2)$$

Ahora del lema 2.1 y la simetría de $g(z)$ tenemos

$$\begin{aligned} \Pr((Z_2, \dots, Z_k) \in B_{z'}) &\geq [\Pr((Z_2, \dots, Z_k) \in C_z)]^\lambda [\Pr((Z_2, \dots, Z_k) \in C_{-z})]^{1-\lambda} \\ &= \Pr((Z_2, \dots, Z_k) \in C_z) = g(z) \end{aligned} \quad (2.3)$$

con lo cual el resultado se obtiene de (2.2) y (3). ■

Enunciamos ahora el resultado principal de esta sección.

Teorema 2.3 *Supongamos $X \rightsquigarrow N_k(\theta, \Sigma)$, $a \in \mathfrak{R}^k$, $a'\theta \geq 0$, y Σ definida positiva y conocida. Entonces para cualquier $A \subset \mathfrak{R}^k$ convexo y simétrico con respecto al origen tenemos que:*

$$\Pr(X - \theta \in A) \leq \Pr(X^* - \theta \in A)$$

Demostración. Si denotamos $X^a = P_{\Sigma^{-1}}(X / a'x = 0)$ tenemos que podemos escribir $X^* = X^a I_{[a'x < 0]} + X I_{[a'x \geq 0]}$. Entonces es suficiente probar que

$$\Pr(X - \theta \in A, a'X < 0) \leq \Pr(X^a - \theta \in A, a'X < 0)$$

Y esto podemos deducirlo de

$$\begin{aligned} \Pr(X - \theta \in A, a'X < 0) &= \int_{-\infty}^0 \Pr(X - \theta \in A / a'X = z) dF_{a'X}(z) \\ &\leq \int_{-\infty}^0 \Pr(X - \theta \in A / a'X = 0) dF_{a'X}(z) \\ &= \int_{-\infty}^0 \Pr(X^a - \theta \in A / a'X = 0) dF_{a'X}(z) \\ &= \Pr(X^a - \theta \in A, a'X < 0) \end{aligned}$$

Donde la desigualdad se obtiene aplicando el lema 2.2 y la última igualdad es consecuencia de la independencia entre X^a y $a'X$. ■

Corolario 2.4 *Sea R el estimador usual de σ^2 y para cada $r > 0$ sea $A(r)$ un conjunto convexo y simétrico. Entonces*

$$\Pr(X - \theta \in A(R)) \leq \Pr(X^* - \theta \in A(R))$$

Demostración. El resultado es directo condicionando en r en el teorema anterior puesto que R es independiente de X y también de X^* . ■

Las siguientes aplicaciones ilustran las implicaciones del teorema anterior y su corolario en cuanto a la superioridad de X^* frente a X para estimar θ o cualquier función lineal suya bajo diferentes criterios de dominación.

Aplicación 2.1 Estimación de coordenadas

Sea $A_i = \{x_i \in \mathfrak{R} : |x_i| \leq t_i\}$ con $t_i > 0$ para $1 \leq i \leq k$. Entonces del teorema tenemos que

$$\Pr(|X_i - \theta_i| \leq t_i) \leq \Pr(|X_i^* - \theta_i| \leq t_i) \text{ para } 1 \leq i \leq k$$

Es decir que $|X_i - \theta_i|$ es estocásticamente menor que $|X_i^* - \theta_i|$ para $1 \leq i \leq k$. De aquí puede deducirse no solo que el error cuadrático medio del estimador X_i^* es menor que el de X_i sino que además para cualquier función de pérdida de la forma $\ell(\delta; \theta) = \rho(|\delta - \theta|)$ con $\rho(t)$ no constante y no decreciente en $t > 0$ tenemos que $\ell(X_i; \theta_i)$ es estocásticamente menor que $\ell(X_i^*; \theta_i)$. Con argumentos similares obtenemos que para cualquier $c \in \mathfrak{R}^k$, $|c'X - c'\theta|$ es estocásticamente menor que $|c'X^* - c'\theta|$ y por tanto que X^* está más concentrado alrededor de θ que X , utilizando el criterio de Lehmann (83) página 291.

Este resultado ya aparece en Rueda y Salvador (95).

Aplicación 2.2 Intervalos de confianza simultáneos de longitud constante

Para cada $c \in \mathfrak{R}^k$ sea $t_c > 0$ y sea $A = \bigcap_{c \in \mathfrak{R}^k} \{x \in \mathfrak{R}^k : |c'x| \leq t_c\}$. Entonces según el teorema tenemos que

$$\Pr(|c'X - c'\theta| \leq t_c, \forall c \in \mathfrak{R}^k) \leq \Pr(|c'X^* - c'\theta| \leq t_c, \forall c \in \mathfrak{R}^k)$$

con lo cual los intervalos de confianza simultáneos para todas las funciones lineales del parámetro y de longitud constante centrados en el estimador máximo verosímil tienen mayor probabilidad de cubrimiento que los centrados en el estimador estandar. Obviamente este mismo tipo de argumentación puede

utilizarse para obtener un intervalo simultáneo para un subconjunto de direcciones de interés.

Aplicación 2.3 Intervalos de confianza simultáneos del tipo de Scheffé, de Tukey o Bonferroni

Sean R el estimador estandar de σ^2 y para cada $r > 0$

$$A(r) = \bigcap_{c \in \mathfrak{R}^k} \{x \in \mathfrak{R}^k : |c'x| \leq rt_c\}$$

donde $t_c > 0, \forall c \in \mathfrak{R}^k$. Entonces del corolario deducimos

$$\Pr(|c'X - c'\theta| \leq Rt_c, \forall c \in \mathfrak{R}^k) \leq \Pr(|c'X^* - c'\theta| \leq Rt_c, \forall c \in \mathfrak{R}^k)$$

Por tanto los intervalos de confianza simultáneos habituales también tienen una probabilidad de cubrimiento superior cuando están centrados en el estimador máximo verosímil en lugar de utilizarse el estandar. Como en la aplicación anterior podemos mejorar también los intervalos obtenidos para un subconjunto de direcciones de particular interés como por ejemplo las del tipo $c_i = e_i - e_j$ para $1 \leq i, j \leq k$ con $i \neq j$.

Se puede comprobar inmediatamente que este resultado también es válido para el procedimiento general de Bowden (70) cuyo caso particular más interesante y en ciertos sentidos óptimo, como probó Bohrer (73), es el S-método de Scheffé.

Aplicación 2.4 Dominación universal

Sea B una matriz simétrica definida positiva y para cada $r > 0, t > 0$ sea $A(r) = \{x \in \mathfrak{R}^k : x'Bx \leq tr\}$. De nuevo del corolario deducimos

$$\Pr((X - \theta)' B (X - \theta) \leq Rt) \leq \Pr((X^* - \theta)' B (X^* - \theta) \leq Rt)$$

Utilizando la terminología introducida por Hwang (85) sobre la dominación estocástica bajo el error euclídeo generalizado y la dominación universal concluimos que X^* domina universalmente a X y por tanto que X es U-inadmisibile.

Ejemplo 2.5 Modelo Lineal Normal

Tanto el teorema 2.3 como el corolario 2.4 tienen un interés claro en el caso de la estimación en un modelo lineal $Y = X\beta + \varepsilon$. Bajo las hipótesis del modelo lineal normal de regresión es bien sabido que $\hat{\beta} \rightsquigarrow N_k(\beta, \sigma^2 (X'X)^{-1})$ donde σ^2 es la varianza común y desconocida de los errores ε_i . En esta situación si además se dispone de la información adicional de que $a'\beta \geq 0$, se puede concluir que una estimación simultánea de los parámetros β del modelo tendrá una probabilidad de cubrimiento superior si se emplea el EMV $\hat{\beta}^* = P_{X'X}(\hat{\beta}/a'X \geq 0)$ en lugar del estimador habitual $\hat{\beta} = (X'X)^{-1} X'Y$.

Si utilizamos el S-método de Scheffé para obtener una banda de confianza para el hiperplano de regresión de la aplicación tercera obtenemos que dicha banda de confianza tiene mayor probabilidad de cubrimiento si la calculamos a partir del EMV en lugar de utilizar el estimador estandar.

2.2 Estimación simultánea en conos de orden

En esta sección se van a demostrar resultados acerca de la probabilidad de cubrimiento simultánea para conos de orden. Mientras no se advierta lo contrario, en esta sección tendremos que

X es un vector aleatorio normal k dimensional de media θ perteneciente al

cono de orden C y de matriz de covarianzas diagonal $D = \text{diag}(\tau_1, \dots, \tau_k)$. C denotará un cono definido por la relación de orden \preceq sobre el conjunto de subíndices $\{1, \dots, k\}$. Es decir que

$$C = \left\{ \theta \in \mathfrak{R}^k : \theta_i \leq \theta_j \text{ para cualesquiera } i, j \in \{1, \dots, k\} \text{ tales que } i \preceq j \right\}$$

En la demostración de los resultados de esta sección se hará uso del algoritmo Minimum Lower Sets que está descrito en Robertson et al. (88) páginas 24-25. Este algoritmo está diseñado para calcular $x^* = P_W(x/C)$. Notar que $P_W(x/C)$ coincide con la “regresión isotónica” de x con pesos ω_i que para el caso normal son $\omega_i = 1/\tau_i$. Además Robertson et al., páginas 30-39, desarrolla la “regresión isotónica generalizada” que nos permite que algunos de los resultados que aquí probamos tengan validez para distribuciones diferentes de la normal. Para la aplicación del algoritmo es necesario introducir el concepto de conjunto inferior:

$L \subset \{1, 2, \dots, k\}$ es un conjunto inferior si $\forall i \in L, j \preceq i$ implica $j \in L$

además para un conjunto cualquiera de índices B y un punto $x \in \mathfrak{R}^k$ denotaremos el promedio en B con pesos ω_i por

$$Av_x(B) = \frac{\sum_{i \in B} \omega_i x_i}{\sum_{i \in B} \omega_i}$$

cuando no haya ambigüedad acerca del punto escribiremos simplemente $Av(B)$. Es conveniente recordar, puesto que haremos un uso frecuente de ello, que la función promedio que acabamos de definir cumple la Propiedad del Valor

Medio de Cauchy es decir que

$$D \subset B \text{ y } Av(D) \leq Av(B) \implies Av(D) \leq Av(B) \leq Av(B - D) \quad (2.4)$$

Como esbozo del funcionamiento del algoritmo podemos decir que consiste en construir primeramente el conjunto inferior de promedio mínimo más grande posible y repetir el proceso sobre el conjunto de índices que no han entrado en los conjuntos inferiores de promedio mínimo definidos en los pasos anteriores. El valor la coordenada i -ésima de la regresión isotónica es el del promedio correspondiente al único conjunto dado por el algoritmo en el que esa coordenada está incluida.

Nota 2.1 *En este contexto no se pueden obtener teoremas tan generales como el de la sección anterior. Es decir que si C es un cono de orden definido por la relación \preceq en general no es cierto que si A es un conjunto convexo y simétrico respecto al origen*

$$\Pr(X - \theta \in A) \leq \Pr(X^* - \theta \in A) \quad (2.5)$$

Esto es obvio de Lee (88) sin más que considerar un tree order, tomando k suficientemente grande y

$$A = \{x \in \mathfrak{R}^k : |x_0| \leq t_0\}$$

Tampoco puede obtenerse un teorema de ese tipo para el caso particular de un orden total puesto que a pesar de que para este cono Kelly (89) probó que para

$$1 \leq i \leq k$$

$$\Pr(|X_i - \theta_i| \leq t_i) \leq \Pr(|X_i^* - \theta_i| \leq t_i)$$

en el primer capítulo hemos visto que para k suficientemente grande y d dirección central del cono

$$E\left((d'X - d'\theta)^2\right) > E\left((d'X^* - d'\theta)^2\right) \quad (2.6)$$

mientras que si una expresión del tipo (2.5) fuera cierta para este cono utilizando los argumentos empleados para obtener las aplicaciones del capítulo anterior tendríamos que

$$\Pr(|c'X - c'\theta| \leq t) \leq \Pr(|c'X^* - c'\theta| \leq t)$$

para cualquier $c \in \mathfrak{R}^k$ y para todo $t > 0$ entrando en contradicción con (2.6).

Sin embargo si podemos obtener otro tipo de resultados como los que siguen.

Teorema 2.5 Para cualquier $t > 0$ se verifica que el conjunto

$$A = \left\{x \in \mathfrak{R}^k : |x_i - \theta_i| \leq t, \quad i = 1, 2, \dots, k\right\}$$

está contenido en

$$A^* = \left\{x \in \mathfrak{R}^k : |x_i^* - \theta_i| \leq t, \quad i = 1, 2, \dots, k\right\}$$

Demostración. Elegimos $x \in A$. Podemos obtener x^* a través del algoritmo Minimum Lower Sets. Sea B_1 el primer conjunto inferior dado por el algoritmo y elijamos $i \in B_1$. Se considera:

$$K_{\text{sup}}^i = \{r \in B_1 : i \preceq r\}$$

de tal forma que $B_1 - K_{\text{sup}}^i$ sigue siendo un conjunto inferior contenido en B_1 . Entonces por el desarrollo del algoritmo $Av(B_1 - K_{\text{sup}}^i) \geq Av(B_1)$ y de (2.4) tenemos:

$$Av(B_1 - K_{\text{sup}}^i) \geq Av(B_1) = x_i^* \geq Av(K_{\text{sup}}^i) \geq \inf_{r \in K_{\text{sup}}^i} (\theta_r - t) = \theta_i - t \quad (2.7)$$

Definimos además:

$$K_{\text{inf}}^i = \{r \in B_1 : r \preceq i\}$$

que es también un conjunto inferior contenido en B_1 y por tanto:

$$x_i^* = Av(B_1) \leq Av(K_{\text{inf}}^i) \leq \sup_{r \in K_{\text{inf}}^i} (\theta_r + t) = \theta_i + t \quad (2.8)$$

Entonces de (2.7) y (2.8) hemos probado que $|x_i^* - \theta_i| \leq t$.

Si ahora repetimos este proceso en cada uno de los elementos de este conjunto y sobre cada uno de los conjuntos que el algoritmo define obtenemos el resultado, teniendo en cuenta que cada uno de los conjuntos que el algoritmo define pueden ser considerados como el primer minimum lower set del conjunto de coordenadas que no han sido tratadas antes. ■

Nota 2.2 *El teorema anterior puede demostrarse con valores de t_i diferentes para coordenadas no relacionadas por el orden establecido por \preceq . Además cuando se estiman simultáneamente coordenadas relacionadas es fácil ver que el único resultado posible de este tipo se obtiene cuando las precisiones de las estimaciones son iguales, es decir $t_i = t$ para cualquier subíndice que se esté estimando.*

Corolario 2.6

$$\begin{aligned} & \Pr \left(x \in \mathfrak{R}^k : |x_i - \theta_i| \leq t, \ i = 1, 2, \dots, k \right) \\ & \leq \Pr \left(x \in \mathfrak{R}^k : |x_i^* - \theta_i| \leq t, \ i = 1, 2, \dots, k \right) \end{aligned}$$

Este corolario es consecuencia inmediata del teorema y es válido no solo en el caso en el que $X_i \rightsquigarrow N(\theta_i, \tau_i)$ independientes sino también en todas las situaciones que verifican el Teorema 1.5.2 de Robertson et al. (88) página 34, entre las que están las siguientes:

1. Cuando la distribución de X_i es geométrica, es decir $\Pr(X_i = x) = \theta^x (1 - \theta)$ para $x = 0, 1, 2, \dots$ y se pretende estimar el parámetro θ del vector X de componentes X_i , puede utilizarse el algoritmo haciendo una serie de transformaciones que se describen en la referencia anterior.
2. Si $X_i \rightsquigarrow \gamma(\tau_i, 1/\theta_i)$ es decir con densidad $f(x) = \frac{x^{\tau_i-1} e^{-x/\theta_i}}{\theta_i^{\tau_i} \Gamma(\tau_i)}$ para $x > 0$. Si suponemos que los τ_i son conocidos entonces el estimador máximo verosímil puede calcularse como la regresión isotónica de X_i/τ_i con $\omega_i = \tau_i$.
3. Para $X_i \rightsquigarrow N(\tau_i, \theta_i)$ con τ_i conocidos el EMV restringido de las varianzas es la regresión isotónica de $\hat{\theta}$ con pesos $\omega_i = 1$ donde $\hat{\theta}_i = (X_i - \tau_i)^2$.
4. Cuando la distribución de X_i es de Poisson de parámetro $\lambda_i T_i$ con T_i conocidos también puede calcularse el estimador de $\lambda = (\lambda_1, \dots, \lambda_k)$ a través de la regresión isotónica de x_i/T_i con pesos $\omega_i = T_i$.

Podemos interpretar este corolario diciendo que bajo restricciones de orden cualesquiera los intervalos de confianza simultáneos de longitud fijada e igual para todas las coordenadas tienen mayor probabilidad de cubrimiento si están centrados en el EMV X^* que si lo están en el estimador estandar X .

Podríamos pensar en extender el resultado del teorema 2.5 a restricciones más generales que las definidas por conos de orden. Esta situación se trata en el apéndice a esta sección.

A continuación en el teorema 2.7 y corolarios vamos a demostrar resultados acerca de la probabilidad de cubrimiento simultánea cuando C es un cono de orden que tiene un único elemento minimal o maximal. Diremos que i es un elemento minimal del cono de orden definido por la relación \preceq en el conjunto $\{1, \dots, k\}$ si no existe ningún elemento $j \neq i$ en ese conjunto tal que $j \preceq i$. La definición de elemento maximal es análoga, es decir i será maximal si no existe $j \neq i$ tal que $i \preceq j$.

Como casos particulares importantes de este tipo de conos están el cono del orden total u orden simple $C = \{\theta \in \mathfrak{R}^k : \theta_1 \leq \dots \leq \theta_k\}$ que tiene como único minimal el elemento 1 y como único maximal el elemento k , el de la unimodalidad $C = \{\theta \in \mathfrak{R}^k : \theta_1 \leq \dots \leq \theta_p \geq \dots \geq \theta_k\}$ en el que p es el único maximal o el tree order $C = \{\theta \in \mathfrak{R}^k : \theta_0 \leq \theta_i \text{ para } i = 1, \dots, k-1\}$ donde 0 es el único minimal.

Sea $X \rightsquigarrow N_k(\theta, \sigma^2 D)$ con D diagonal y conocida y $\theta \in C$ cono de orden.

Teorema 2.7 *Supongamos que C tiene un único minimal que denotamos por 0. Entonces para cualquier $t_0 \geq t \geq 0$:*

$$\begin{aligned} \Pr(|X_0 - \theta_0| \leq t_0; |X_i - \theta_i| \leq t, i \neq 0) &\leq \\ &\leq \Pr(|X_0^* - \theta_0| \leq t_0; |X_i^* - \theta_i| \leq t, i \neq 0) \end{aligned}$$

Demstración. Primero ordenamos los valores de los parámetros de forma consistente con el orden parcial definido por \preceq , y de tal forma que $\theta_i \leq \theta_{i+1}$ para $i \geq 0$. Denotamos además por A y A^* los conjuntos cuyas probabilidades estamos comparando. Es decir:

$$\begin{aligned} A &= \{x : |x_0 - \theta_0| \leq t_0; |x_i - \theta_i| \leq t, i \neq 0\} \\ A^* &= \{x : |x_0^* - \theta_0| \leq t_0; |x_i^* - \theta_i| \leq t, i \neq 0\} \end{aligned}$$

Para probar el resultado es suficiente ver que $\Pr(A \cap \overline{A^*}) \leq \Pr(A^* \cap \overline{A})$ puesto que entonces

$$\Pr(A) = \Pr(A \cap \overline{A^*}) + \Pr(A \cap A^*) \leq \Pr(A^* \cap \overline{A}) + \Pr(A^* \cap A) = \Pr(A^*) \quad (2.9)$$

Sean $B_1^{x^0}, \dots, B_{h_{x^0}}^{x^0}$ los conjuntos de nivel asociados con un punto dado x^0 , donde evidentemente $0 \in B_1^{x^0}$. Si p es el elemento mayor de $B_1^{x^0}$ (considerando el orden total que hemos definido) tenemos que:

$$Av(B_1^{x^0}) = x_p^{0*} \geq x_p^0 \geq \theta_p - t_p \geq \theta_i - t_i, \text{ para } i \in B_1^{x^0} \quad (2.10)$$

donde $t_i = t$ para $i > 0$.

Elegimos $x \in A$ entonces del teorema 2.5 sabemos que

$$|x_i^* - \theta_i| \leq t_i, \text{ para } i = 0 \text{ y para } i \notin B_1^x \quad (2.11)$$

Supongamos que $x \notin A^*$, entonces de (2.10) y (2.11) deducimos que

$$\exists i \in B_1^x - \{0\} \text{ tal que } x_i^* = Av(B_1^x) > \theta_i + t$$

pero como hemos ordenado los parámetros, $\theta_1 + t \leq \theta_i + t$ si $i \in B_1^x - \{0\}$ y entonces $Av(B_1^x) > \theta_1 + t$. Por otra parte, como $x \in A$ sabemos que $x_1 \leq \theta_1 + t$, y por tanto de las propiedades del algoritmo MLS y la hipótesis del único minimal en el orden obtenemos $1 \in B_1^x$ y

$$Av_x(0, 1) \geq Av(B_1^x) > \theta_1 + t \quad (2.12)$$

Ahora podemos escribir x_0 y x_1 como:

$$x_0 = \theta_0 + t + \varepsilon \text{ y } x_1 = \theta_1 + t - \varepsilon_1 \text{ donde } \varepsilon \geq 0 \text{ y } 0 \leq \varepsilon_1 \leq 2t \quad (2.13)$$

Notar que si $\varepsilon < 0$ obtendríamos $x \in A^*$ utilizando el teorema 2.5.

Entonces para cada $x \in A \cap \overline{A^*}$ definimos otro punto y de la siguiente forma:

$$y_0 = x_0 - \zeta; \quad y_1 = x_1 - \zeta; \quad y_i = x_i \text{ si } i > 1$$

donde

$$\zeta = \begin{cases} 2t & \text{si } Av_x(0, 1) - x_1 \leq 2t \\ Av_x(0, 1) - x_1 & \text{si } Av_x(0, 1) - x_1 > 2t \end{cases} \quad (2.14)$$

Es fácil comprobar que esta transformación es inyectiva. Probamos además que

$$\zeta \leq 2t + 2 \frac{\omega_0 \varepsilon - \omega_1 \varepsilon_1}{\omega_0 + \omega_1} \quad (2.15)$$

De (2.12), $Av_x(0, 1) = \frac{\omega_0 \theta_0 + \omega_1 \theta_1}{\omega_0 + \omega_1} + t + \frac{\omega_0 \varepsilon - \omega_1 \varepsilon_1}{\omega_0 + \omega_1} > \theta_1 + t$ y como $\theta_0 \leq \theta_1$ obtenemos $\omega_0 \varepsilon - \omega_1 \varepsilon_1 > 0$. Esto prueba (2.15) cuando $\zeta = 2t$. De otro lado cuando $\zeta = Av_x(0, 1) - x_1$:

$$\zeta = \frac{\omega_0 \theta_0 + \omega_1 \theta_1}{\omega_0 + \omega_1} + t + \frac{\omega_0 \varepsilon - \omega_1 \varepsilon_1}{\omega_0 + \omega_1} - (\theta_1 + t - \varepsilon_1)$$

$$\begin{aligned}
&= \frac{\omega_0}{\omega_0 + \omega_1} (\theta_0 - \theta_1) + \frac{\omega_0}{\omega_0 + \omega_1} (\varepsilon + \varepsilon_1) \\
&\leq \frac{\omega_0}{\omega_0 + \omega_1} (\varepsilon + \varepsilon_1) \\
&= 2t + \frac{\omega_0}{\omega_0 + \omega_1} (\varepsilon + \varepsilon_1 - 2t) - \frac{\omega_1}{\omega_0 + \omega_1} 2t \\
&\leq 2t + \frac{\omega_0 \varepsilon - \omega_1 \varepsilon_1}{\omega_0 + \omega_1}
\end{aligned}$$

teniendo en cuenta que como notamos en (2.13), $\varepsilon_1 \leq 2t$. Utilizando de nuevo $\omega_0 \varepsilon - \omega_1 \varepsilon_1 > 0$ probamos finalmente (2.15).

El punto y verifica claramente que $y \notin A$ a.s. porque $y_1 - \theta_1 < -t$ excepto en algunas de las ocasiones en las que se verifica que $x_1 = \theta_1 + t$, conjunto este que tiene probabilidad 0. Este punto verifica también que $y \in A^*$. Para probar esta afirmación actuamos en cuatro pasos.

Primeramente es claro que $0 \in B_1^y$ pero $B_1^y \neq \{0\}$ porque $y_1 < y_0$. Además $1 \in B_1^y$ ya que en caso contrario tendríamos que

$$Av(0) \geq Av(B_1^y) \geq Av(B_1^y - \{0\}) \geq \inf(y_j : j \in B_1^y - \{0\}) > \theta_1 - t$$

y como $y_1 < \theta_1 - t$ entonces $Av(B_1^y \cup \{1\}) < Av(B_1^y)$ llegándose a contradicción.

Segundo $Av_y(0, 1) \in [\theta_1 - t, \theta_1 + t]$ puesto que:

Cuando $\zeta = 2t$ de (2.12) y (2.14) tenemos:

$$\theta_1 + t < Av_x(B_1^x) \leq Av_x(0, 1) \leq x_1 + 2t \leq \theta_1 + 3t$$

y cuando $\zeta = Av_x(0, 1) - x_1$ es obvio que $Av_y(0, 1) = x_1 \in [\theta_1 - t, \theta_1 + t]$.

Tercero, para $i \in B_1^y$ tenemos que $y_i^* \in [\theta_i - t_i, \theta_i + t_i]$, donde $t_i = t$ si $i > 0$, ya que si $B_1^y = \{0, 1\}$ acabamos de probar que $Av_y(0, 1) \in [\theta_1 - t, \theta_1 + t]$ y en

otro caso

$$Av_y(B_1^y) = y_i^* \leq Av_y(0, 1) \leq \theta_1 + t \leq \theta_i + t \text{ si } i \neq 0$$

$$Av_y(B_1^y) = y_0^* \leq y_0 \leq \theta_0 + t_0$$

y de (2.10) deducimos que $y_i^* \geq \theta_i - t$ para cualquier $i \in B_1^y$

Finalmente cuando $i \notin B_1^y$ tenemos que $x_i = y_i$ y por tanto hemos probado que $y \in A^*$.

Además si f es la función de densidad de un vector aleatorio de distribución $N_k(\theta, D)$, se tiene que $f(y) \geq f(x)$:

$$\begin{aligned} \ln f(x) - \ln f(y) &= \frac{1}{2} \left(\sum_{i=0}^k \omega_i (y_i - \theta_i)^2 - \sum_{i=0}^k \omega_i (x_i - \theta_i)^2 \right) \\ &= \frac{1}{2} \left(\omega_0 (t + \varepsilon - \zeta)^2 + \omega_1 (t - \varepsilon_1 - \zeta)^2 - (\omega_0 (t + \varepsilon)^2 + \omega_1 (t - \varepsilon_1)^2) \right) \\ &= \zeta ((\omega_0 + \omega_1) \zeta - 2((\omega_0 + \omega_1)t + (\omega_0\varepsilon - \omega_1\varepsilon_1))) \leq 0 \end{aligned} \tag{2.16}$$

donde la desigualdad es consecuencia de (2.15).

Definimos ahora $B = \{x : Av_x(0, 1) - x_1 \leq 2t\}$ y las transformaciones $g, h : \mathfrak{R}^k \longrightarrow \mathfrak{R}^k$

$$g(x) = y \text{ donde } y_0 = x_0 - 2t, y_1 = x_1 - 2t, y_i = x_i \text{ si } i > 1$$

$$h(x) = y \text{ donde } y_0 = x_0 + x_1 - Av_x(0, 1), y_1 = 2x_1 - Av_x(0, 1), y_i = x_i \text{ si } i > 1$$

Cada una de estas transformaciones es obviamente continua, con derivadas parciales continuas e inversible. Es inmediato también comprobar que el determinante jacobiano de cada una de estas transformaciones es 1 y por tanto

tenemos

$$\begin{aligned} \Pr(A \cap \overline{A^*} \cap B) &= \int_{A \cap \overline{A^*} \cap B} f(x) \, dx = \int_{g(A \cap \overline{A^*} \cap B)} f(g^{-1}(y)) \, dy \\ &\leq \int_{g(A \cap \overline{A^*} \cap B)} f(y) \, dy = \Pr[g(A \cap \overline{A^*} \cap B)] \end{aligned}$$

ya que de (2.16) $f(y) \geq f(g^{-1}(y)) = f(x)$ cuando $y \in A \cap \overline{A^*} \cap B$. Con los mismos argumentos podemos probar que

$$\Pr(A \cap \overline{A^*} \cap \overline{B}) \leq \Pr[h(A \cap \overline{A^*} \cap \overline{B})]$$

Además puesto que la transformación original era inyectiva tenemos que

$$g(A \cap \overline{A^*} \cap B) \cap h(A \cap \overline{A^*} \cap \overline{B}) = \emptyset$$

Como también hemos demostrado que $g(A \cap \overline{A^*} \cap B) \cup h(A \cap \overline{A^*} \cap \overline{B}) \subset A^* \cap \overline{A}$ podemos concluir

$$\Pr(A \cap \overline{A^*}) \leq \Pr[g(A \cap \overline{A^*} \cap B)] + \Pr[h(A \cap \overline{A^*} \cap \overline{B})] \leq \Pr(A^* \cap \overline{A})$$

y por tanto hemos probado (2.9). ■

Corolario 2.8 *Supongamos que C tiene un único maximal que denotamos por k . Entonces para cualquier $t_k \geq t \geq 0$:*

$$\begin{aligned} \Pr(|X_k - \theta_k| \leq t_k; |X_i - \theta_i| \leq t, i \neq k) &\leq \\ &\leq \Pr(|X_k^* - \theta_k| \leq t_k; |X_i^* - \theta_i| \leq t, i \neq k) \end{aligned}$$

Demostración. Basta considerar el cono $-C$ en el que el subíndice k es el correspondiente al único elemento minimal y utilizar el teorema 2.7

apoyándonos en que según las formulas min-max de Robertson et al. (88)

$$X_C^* = -X_{-C}^* \quad \blacksquare$$

Corolario 2.9 *Si i es el único elemento minimal (maximal) de un cono C , entonces $\forall t \geq 0$*

$$\Pr(|X_j - \theta_j| < t, j \neq i) \leq \Pr(|X_j^* - \theta_j| < t, j \neq i)$$

sin mas que hacer $t_0 \rightarrow \infty$.

Ejemplo 2.6 Supongamos un experimento consistente en $k - 1$ tratamientos y un grupo de control. Cada uno de estos grupos viene representado por una variable aleatoria $N(\theta_i, \omega_i)$ para $i = 0, \dots, k - 1$ donde $i = 0$ se refiere al grupo de control. En ciertas aplicaciones (como por ejemplo en Dunnett (55)) se sabe que los parámetros verifican $\theta_0 \leq \theta_i$ si $i = 1, \dots, k - 1$. Esta ordenación, como es bien sabido, se denomina simple tree order, o tree order o árbol simple. Del corolario anterior podemos deducir que para estimar simultáneamente las medias de los tratamientos con la misma precisión t es preferible utilizar el EMV que el estimador estandar.

Tratamos ahora sobre posibles extensiones o generalizaciones de este tipo de resultados.

Nota 2.3 *Cuando $t_0 < t$ no tienen porque verificarse desigualdades como la anterior puesto que como demostró Lee (88) se tiene que, bajo las restricciones correspondientes a un tree order, cuando k es suficientemente grande*

$$\Pr(|X_0 - \theta_0| < t_0) > \Pr(|X_0^* - \theta_0| < t_0)$$

Además ni siquiera es cierta la desigualdad del teorema para valores de t_0 relativamente cercanos al valor de u puesto que utilizando simulación se ha comprobado que tomando $k = 14$ y $t_0 = 1$, $t = 3$ se tiene que $P(A) = 0.65$ mientras que $P(A^*) \simeq 0.46$

Nota 2.4 Podría pensarse, una vez visto el teorema 2.7, en intentar uno similar que involucrara simultáneamente todos los elementos minimales de un cono de orden, es decir intentar probar que:

$$\begin{aligned} \Pr(|X_j - \theta_j| \leq t_j, j \in T; |X_i - \theta_i| \leq t, i \notin T) &\leq \\ &\leq \Pr(|X_j^* - \theta_j| \leq t_j, j \in T; |X_i^* - \theta_i| \leq t, i \notin T) \end{aligned}$$

donde T es el conjunto de elementos minimales del orden que define el cono.

Sin embargo la conjetura anterior no es cierta como puede comprobarse utilizando de nuevo el contraejemplo de Lee con el orden invertido.

Por otra parte, utilizando este mismo tipo de argumentos, es posible demostrar resultados parciales para caso de ordenes específicos, en particular los resultados para el orden total son siempre de máximo interés por ser el cono de orden de aplicación más extendida. En este sentido se obtiene el siguiente resultado.

Teorema 2.10 Sea C el cono del orden total en \mathfrak{R}^k . Entonces si $i \in \{1, \dots, k\}$ se tiene que cuando $0 \leq t \leq t_i$:

$$\Pr(|X_i - \theta_i| \leq t_i; |X_j - \theta_j| \leq t, j \neq i) \leq \Pr(|X_i^* - \theta_i| \leq t_i; |X_j^* - \theta_j| \leq t, j \neq i)$$

Demostración. Elegimos $x \in A = \{x \in \mathfrak{R}^k : |x_i - \theta_i| \leq t, i = 1, 2, \dots, k\}$ de tal forma que $x \notin A^*$ donde $A^* = \{x \in \mathfrak{R}^k : |x_i^* - \theta_i| \leq t, i = 1, 2, \dots, k\}$.

Entonces es claro que no puede verificarse $|x_i - \theta_i| \leq t$ puesto que ese punto verificaría las condiciones del teorema 2.5 y por tanto se tendría que $x \in A^*$.

Supongamos entonces que, por ejemplo, $-t_i \leq x_i - \theta_i < -t$. La situación en la que $t < x_i - \theta_i \leq t_i$ es completamente simétrica y se trata de la misma manera.

Consideramos ahora los conjuntos de nivel correspondientes a este punto que hemos seleccionado. Es claro que los conjuntos de nivel que no contienen al subíndice i no nos dan problemas utilizando de nuevo el teorema 2.5. Queda entonces por examinar el conjunto de nivel correspondiente a dicho subíndice.

Denotemos por $B = \{i - p, \dots, i, i + 1, \dots, i + q\}$ este conjunto de nivel. Es claro que si $p = 0$ entonces también $q = 0$ puesto que en estas condiciones $x_i < x_j$, $\forall j > i$ y por tanto en esta situación el problema está resuelto (de hecho $x \in A^*$).

Veamos primeramente que $Av(B)$ no se escapa por la derecha de los intervalos.

Esto es claro de:

$$Av(B) \leq x_{i-p} \leq \theta_{i-p} + t \leq \theta_j + t \quad \text{si } j > i - p$$

Por tanto si $x \notin A^*$ ese promedio debe escaparse por la izquierda de alguno de los intervalos. Es claro, por otra parte que no va escapar del intervalo correspondiente al subíndice i (haciendo de nuevo referencia al teorema 2.5, por ejemplo). Veamos ahora que, en estas condiciones, $Av(B) < \theta_{i-1} - t$ y que además $q = 0$.

Si $q > 0$ entonces puesto que $Av(B)$ escapa por la izquierda de algún intervalo debe escapar, al menos, del último puesto que $\theta_{i+q} \geq \theta_j$, para $j < i + q$.

Entonces:

$$Av(B) < \theta_{i+q} - t \leq x_{i+q}$$

con lo cual por la propiedad de valor medio de Cauchy:

$$Av(i-p, \dots, i+q-1) < Av(B)$$

lo que es una contradicción con el algoritmo minimum lower sets. Hemos probado entonces que $B = \{i-p, \dots, i\}$ y que además $Av(B) < \theta_{i-1} - t$.

Consideramos ahora el cono $C = \{\theta_1 \leq \dots \leq \theta_i\}$ y le aplicamos el esquema del teorema 2.7 al punto $x^C = (x_1, \dots, x_i)$ obteniéndose un nuevo punto y^C que verifica $y^C \notin A^C$ c. s., $y^{*C} \in A^C$ y $f^C(y^C) \geq f^C(x^C)$.

Si denotamos ahora por $y = (y_1, \dots, y_i, x_{i+1}, \dots, x_k)$ vamos a ver que este punto es suficiente para solucionar el problema puesto que verifica:

1. $y \notin A$ c. s. obviamente puesto que $y \notin A^C$ c. s..
2. $y \in A^*$. Para probar esto supongamos que vamos ejecutando el algoritmo minimum lower sets sobre este punto. Cuando se llega al subíndice i lo que se tiene de momento es el punto y^{*C} que pertenece a A^C . Además del teorema 2.7 se tiene que la componente i -ésima verifica $\theta_i - t_i \leq y_i^{*C} \leq \theta_i + t$. (Notar que $\theta_i \geq \theta_{i-1}$ y se ha probado que $y^C \in A^C$). Puede suceder entonces:
 - $y_i^{*C} < \theta_i - t$ con lo cual como $Av_y(i+1, \dots, i+p) \geq \theta_i - t$ para cualquier $p > 0$ se tiene que el siguiente conjunto de nivel empieza en $i+1$ y por tanto el teorema 2.5 aplicado al resto de los subíndices concluye la cuestión.
 - $|y_i^{*C} - \theta_i| \leq t$. Entonces, como por el algoritmo MLS se verifica que

$y^* = (y_1^*, \dots, y_i^*, x_{i+1}, \dots, x_k)^*$, y este último punto está en A de nuevo por el teorema 2.5 hemos terminado.

3. $f(y) \geq f(x)$ porque $f(y) - f(x) = K \cdot (f^C(y^C) - f^C(x^C))$ con $K > 0$.

Se concluye entonces el teorema como el teorema 2.7 puesto no es difícil comprobar que estas transformaciones conservan las buenas propiedades que tenían las definidas en ese teorema previo. ■

Los resultados anteriores nos llevan a preguntarnos si una propiedad similar a las anteriores será válida para un orden total cuando se consideran valores positivos arbitrarios para los t_i . No hemos podido resolver completamente esta cuestión pero presentamos unos estudios de simulación que parecen indicar que dicha conjetura puede ser cierta. En estas simulaciones se han utilizado 100.000 observaciones de un vector aleatorio $X \rightsquigarrow N_{10}(\theta, I)$.

$\theta = (0, \dots, 0)$												
t_0	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	Stand.	Tree	Total
1	2.5	2.5	3	3	3.5	3.5	4	4	4.5	.66	.60	.80
4.5	1	2.5	2.5	3	3	3.5	3.5	4	4	.66	.74	.93
1	2.5	2.5	2.75	2.75	3	3	3.25	3.25	3.5	.65	.59	.80
3.5	1	2.5	2.5	2.75	2.75	3	3	3.25	3.25	.65	.74	.92
1	2.5	2.75	3	3.25	3.5	3.75	4	4.25	4.5	.67	.58	.80
4.5	1	2.5	2.75	3	3.25	3.5	3.75	4	4.25	.67	.74	.93
1	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	.61	.56	.80

$\theta = (0, .1, \dots, .1)$												
t_0	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	Stand.	Tree	Total
1	2.5	2.5	3	3	3.5	3.5	4	4	4.5	.66	.64	.81
4.5	1	2.5	2.5	3	3	3.5	3.5	4	4	.66	.74	.92
1	2.5	2.5	2.75	2.75	3	3	3.25	3.25	3.5	.65	.63	.81
3.5	1	2.5	2.5	2.75	2.75	3	3	3.25	3.25	.65	.74	.92
1	2.5	2.75	3	3.25	3.5	3.75	4	4.25	4.5	.67	.62	.80
4.5	1	2.5	2.75	3	3.25	3.5	3.75	4	4.25	.67	.74	.92
1	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	.61	.60	.80

$\theta = (0, .1, .1, .1, .1, .2, .2, .2, .2, .2)$												
t_0	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	Stand.	Tree	Total
1	2.5	2.5	3	3	3.5	3.5	4	4	4.5	.66	.66	.81
4.5	1	2.5	2.5	3	3	3.5	3.5	4	4	.66	.74	.92
1	2.5	2.5	2.75	2.75	3	3	3.25	3.25	3.5	.65	.65	.80
3.5	1	2.5	2.5	2.75	2.75	3	3	3.25	3.25	.65	.74	.92
1	2.5	2.75	3	3.25	3.5	3.75	4	4.25	4.5	.67	.64	.81
4.5	1	2.5	2.75	3	3.25	3.5	3.75	4	4.25	.67	.74	.92
1	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	.61	.62	.80

$\theta = (0, .25, .25, .25, .25, .5, .5, .5, .5, .5)$												
t_0	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	Stand.	Tree	Total
1	2.5	2.5	3	3	3.5	3.5	4	4	4.5	.66	.72	.82
4.5	1	2.5	2.5	3	3	3.5	3.5	4	4	.66	.74	.91
1	2.5	2.5	2.75	2.75	3	3	3.25	3.25	3.5	.65	.72	.82
3.5	1	2.5	2.5	2.75	2.75	3	3	3.25	3.25	.65	.74	.91
1	2.5	2.75	3	3.25	3.5	3.75	4	4.25	4.5	.67	.72	.82
4.5	1	2.5	2.75	3	3.25	3.5	3.75	4	4.25	.67	.74	.91
1	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	.61	.69	.81

2.2.1 Extensiones del Teorema 2.5

El teorema 2.5 parece poder extenderse a otros conos que guarden ciertas características comunes con los conos de orden. En particular, todas las restricciones que definen los conos de orden son contrast, es decir que cualquier cono de orden puede escribirse como $C = \{x \in \mathfrak{R}^k : a'_i x \geq 0, i = 1, \dots, p\}$ donde $\sum_{j=1}^k a_{ij} = 0$ para $i = 1, \dots, p$ de tal forma que $(1, \dots, 1)' \in L(a_i, i = 1, \dots, p)^\perp$. En este sentido una posible generalización natural de dicho teorema sería a conos del tipo $C = \{a'_i x \geq 0, i = 1, \dots, p\} \subset \mathfrak{R}^k$ donde $\sum_{j=1}^k a_{ij} = 0$ para $i = 1, \dots, p$ y $t \in L(a_i)^\perp$ con $t_i > 0$. Sin embargo esto no es cierto como probamos a continuación.

Ejemplo 2.7 Sea $X \rightsquigarrow N_3(\theta, \sigma^2 I_3)$ donde $\theta \in C = \left\{ \theta \in \mathfrak{R}^3 : \frac{\theta_1 + \theta_2}{2} \leq \theta_3 \right\}$ cono definido por una única restricción contrast. Si $\theta = (2, 3, 3) \in C$ y elegimos

$t_1 = t_2 = t_3 = 1$, es decir $t \in L(a)^\perp$ entonces para $x = (1, 4, 2)$ tenemos que $x^* = P(x/C) = \left(\frac{5}{6}, \frac{23}{6}, \frac{7}{3}\right)$ y mientras que $|x_i - \theta_i| = 1$ para $i = 1, 2, 3$ en cambio $|x_1^* - \theta_1| = \frac{7}{6} > 1$.

Aunque la conjetura anterior no sea cierta vamos a ver que el resultado del teorema 2.3 puede extenderse a conos definidos por una restricción cualquiera en \mathfrak{R}^2 aunque esa restricción no sea contrast bajo ciertas condiciones.

Vamos a considerar un vector aleatorio X de dimensión 2 de media $\theta \in C = \{x \in \mathfrak{R}^2 / a'x \geq 0\}$ donde a es un vector de \mathfrak{R}^2 y tal que el EMV $X^* = P_{\Sigma^{-1}}(X/C)$ con $\Sigma = \begin{pmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{pmatrix}$.

Teorema 2.11 *Supongamos que $s_{12} \leq 0$. Entonces para cualquier $t \in L(a)^\perp$ con $t_1, t_2 > 0$ se tiene:*

$$\{x : |x_i - \theta_i| \leq t_i, i = 1, 2\} \subset \{x : |x_i^* - \theta_i| \leq t_i, i = 1, 2\}$$

Demostración. Suponemos sin pérdida de generalidad que $a_i \neq 0, i = 1, 2$ puesto que en otro caso el problema se reduce a dimensión 1 y es trivial.

Como $t \in L(a)^\perp$ es claro que a lo sumo uno de los vértices del rectángulo $R = \{x \in \mathfrak{R}^2 / |x_i - \theta_i| \leq t_i, i = 1, 2\}$ está fuera del cono C la región definida por la restricción. Véase la figura (??) al final de esta demostración.

Evidentemente si $R \subset C$ se tiene la igualdad puesto que en esa región $x = x^*$.

Supongamos sin pérdida de generalidad que $V_2 = (\theta_1 + t_1, \theta_2 - t_2)'$ es el vértice que queda fuera de C . Entonces de $a'V_2 < 0$ tenemos $a_1(\theta_1 + t_1) + a_2(\theta_2 - t_2) < 0 \Leftrightarrow a'\theta + (a_1t_1 - a_2t_2) < 0$ luego $a_1t_1 - a_2t_2 < 0$ para valores de t_1, t_2 suficien-

temente grandes. Por tanto debe suceder $a_1 < 0, a_2 > 0$ y puede considerarse $a'X < 0 \Leftrightarrow (-1, \alpha)X < 0$. Recordar que estamos suponiendo $a_i \neq 0, i = 1, 2$. Ahora en estas condiciones tenemos que $\{a'X = 0\} \cap \{|X_i - \theta_i| \leq t_i, i = 1, 2\} = \{\overline{P_1 P_2}\}$ donde $P_1 = (\theta_1 + \varepsilon_1, \theta_2 - t_2)$ y $P_2 = (\theta_1 + t_1, \theta_2 - \varepsilon_2)$ con $\varepsilon_1 \in [-t_1 t_1], \varepsilon_2 \in [-t_2, t_2]$. El conjunto de los puntos del rectángulo que quedan fuera del cono puede representarse entonces como:

$$\wp = \{\gamma_1 P_1 + \gamma_2 P_2 + (1 - \gamma_1 - \gamma_2) V_2 / 0 \leq \gamma_1, \gamma_2 \leq 1\} \quad (2.17)$$

Si $x' = (x_1, x_2) \in \wp$ entonces su proyección sobre el cono según la métrica definida por la matriz Σ^{-1} , es decir el valor del EMV con restricciones en ese punto, es (vease (3.1), pag. 104):

$$x^* = \begin{pmatrix} \frac{1}{\delta} (\alpha^2 s_{22} - \alpha s_{12}) x_1 + \frac{1}{\delta} (\alpha s_{11} - \alpha^2 s_{12}) x_2 \\ \frac{1}{\delta} (-s_{12} + \alpha s_{22}) x_1 + \frac{1}{\delta} (s_{11} - \alpha s_{12}) x_2 \end{pmatrix} \quad (2.18)$$

donde $\delta = a' \Sigma a = s_{11} + \alpha^2 s_{22} - 2\alpha s_{12}$

Por tanto de (2.17) y (2.18):

$$\begin{aligned} x_1^* &= \frac{1}{\delta} (\alpha^2 s_{22} - \alpha s_{12}) (\gamma_1 (\theta_1 + \varepsilon_1) + \gamma_2 (\theta_1 + t_1) + (1 - \gamma_1 - \gamma_2) (\theta_1 + t_1)) + \\ &+ \frac{1}{\delta} (\alpha s_{11} - \alpha^2 s_{12}) (\gamma_1 (\theta_2 - t_2) + \gamma_2 (\theta_2 - \varepsilon_2) + (1 - \gamma_1 - \gamma_2) (\theta_2 - t_2)) \\ &= \frac{1}{\delta} (\alpha^2 s_{22} - \alpha s_{12}) (\theta_1 + \gamma_1 \varepsilon_1 + (1 - \gamma_1) t_1) \\ &+ \frac{1}{\delta} (\alpha s_{11} - \alpha^2 s_{12}) (\theta_2 - t_2 - \gamma_2 (\varepsilon_2 - t_2)) \\ &\geq \frac{1}{\delta} (\alpha^2 s_{22} - \alpha s_{12}) (\theta_1 + \varepsilon_1) + \frac{1}{\delta} (\alpha s_{11} - \alpha^2 s_{12}) (\theta_2 - t_2) = \theta_1 + \varepsilon_1 \end{aligned} \quad (2.19)$$

puesto que $\varepsilon_i \leq t_i$ y como $a'P_1 = 0$ tenemos $\theta_1 + \varepsilon_1 = \alpha (\theta_2 - t_2)$. De (2.19) deducimos entonces que $x_1^* - \theta_1 \geq \varepsilon_1 \geq -t_1$.

Ahora bien como x verifica $a'x < 0$, es decir $-x_1 + \alpha x_2 < 0$, tenemos además que:

$$\begin{aligned} x_1^* &= \frac{1}{\delta} (\alpha^2 s_{22} - \alpha s_{12}) x_1 + \frac{1}{\delta} (\alpha s_{11} - \alpha^2 s_{12}) x_2 \\ &< \frac{1}{\delta} (\alpha^2 s_{22} - \alpha s_{12}) x_1 + \frac{1}{\delta} (s_{11} - \alpha s_{12}) x_1 = x_1 \end{aligned}$$

con lo cual $x_1^* - \theta_1 < x_1 - \theta_1 \leq t_1$.

Por tanto $|x_1^* - \theta_1| \leq t_1, \forall x \in \wp$.

Además con argumentos similares

$$\begin{aligned} x_2^* &= \frac{1}{\delta} (-s_{12} + \alpha s_{22}) x_1 + \frac{1}{\delta} (s_{11} - \alpha s_{12}) x_2 \\ &> \frac{1}{\delta} (-s_{12} + \alpha s_{22}) \alpha x_2 + \frac{1}{\delta} (s_{11} - \alpha s_{12}) x_2 = x_2 \end{aligned}$$

luego $x_2^* - \theta_2 > x_2 - \theta_2 \geq -t_2$.

Y si $x \in \wp$:

$$\begin{aligned} x_2^* &= \frac{1}{\delta} (-s_{12} + \alpha s_{22}) (\theta_1 + \gamma_1 \varepsilon_1 + (1 - \gamma_1) t_1) \\ &\quad + \frac{1}{\delta} (s_{11} - \alpha s_{12}) (\theta_2 - (1 - \gamma_2) t_2 - \gamma_2 \varepsilon_2) \\ &\leq \frac{1}{\delta} (-s_{12} + \alpha s_{22}) (\theta_1 + t_1) + \frac{1}{\delta} (s_{11} - \alpha s_{12}) (\theta_2 - \varepsilon_2) \quad (2.20) \\ &= \frac{1}{\delta} (-s_{12} + \alpha s_{22}) \alpha (\theta_2 - \varepsilon_2) + \frac{1}{\delta} (s_{11} - \alpha s_{12}) (\theta_2 - \varepsilon_2) \\ &= \theta_2 - \varepsilon_2 \end{aligned}$$

puesto que $a'P_2 = 0$, con lo que $\theta_1 + t_1 = \alpha (\theta_2 - \varepsilon_2)$. Entonces de (2.20), $x_2^* - \theta_2 \leq -\varepsilon_2 \leq t_2$, ya que $\varepsilon_2 \in [-t_2, t_2]$.

Tenemos entonces la otra desigualdad y consecuentemente $|x_2^* - \theta_2| \leq t_2, \forall x \in \wp$.

■

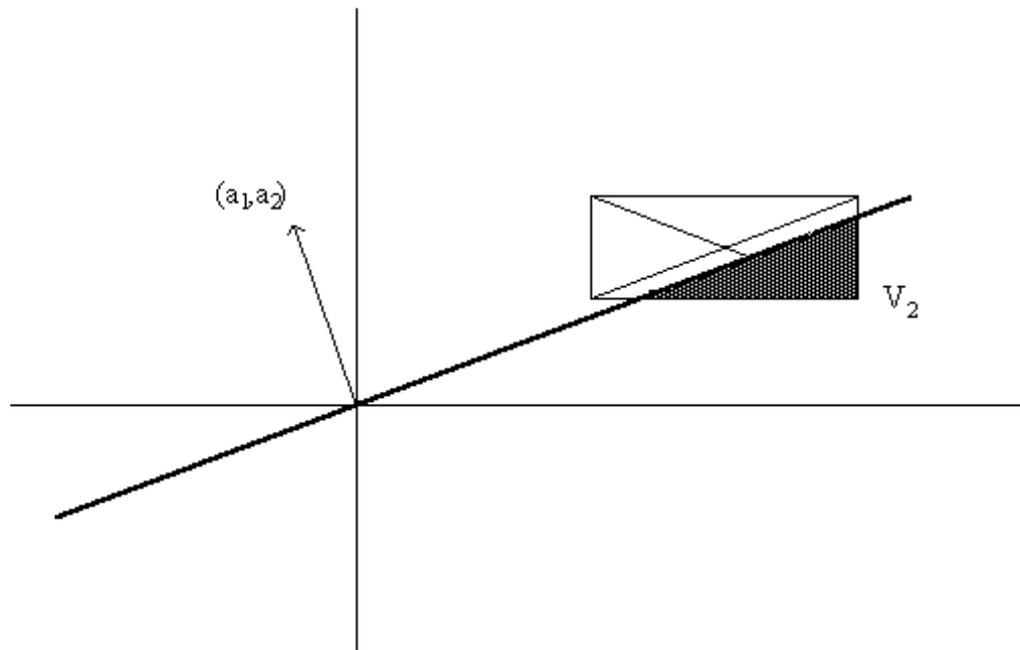


Figura 1: Esquema gráfico de la demostración

Nota 2.5 Hay que tener en cuenta que el teorema 2.11 que acabamos de probar es válido también para las situaciones en las que el EMV se calcula como proyección bajo una métrica sobre el cono de restricciones. Concretamente este resultado es válido para distribuciones normales bidimensionales verificando $Cov(X_1, X_2) \leq 0$ ya que esta condición es equivalente en este caso a $s_{12} \leq 0$.

Capítulo 3

Estimadores Alternativos

En capítulos anteriores hemos comparado el EMV restringido con el no restringido y hemos obtenido situaciones donde el primero siempre es mejor, como por ejemplo cuando se cuenta con la información dada por una única restricción, y otras en las que eso no ocurre como el octante positivo o algunos conos de orden. Dedicamos este capítulo a comparar el EMV con otros estimadores alternativos.

En la primera sección, bajo hipótesis de normalidad, estudiamos hasta que punto el EMV es el estimador óptimo bajo una única restricción comparando este estimador con otros obtenidos a través de proyecciones no ortogonales sobre el subespacio definido por la restricción. Como ya hemos señalado anteriormente este procedimiento de proyección no ortogonal es similar al de cambio de pesos empleado por Lee (88) en la estimación de coordenadas en el tree order.

Tenemos entonces que $X \rightsquigarrow N(\theta, \Sigma)$ donde Σ es una matriz definida positiva, $\theta \in C = \{x \in \mathbb{R}^k : a'x \geq 0\}$ con $a'\Sigma a = 1$. Como en capítulos anteriores X^* es el estimador máximo verosímil. En este caso $X^* = P_{\Sigma^{-1}}(X/C)$. Denotamos

además por W la matriz asociada a una métrica cualquiera y por $X^W = P_W(X/C)$.

En esta primera sección probamos primero que ningún estimador $c'X^W$ U-domina a $c'X^*$ en la estimación de $c'\theta$. Seguidamente damos una condición sobre W bajo la cual $c'X^*$ U-domina a $c'X^W$, condición que posteriormente vemos que es suficiente pero no necesaria. Un ejemplo nos muestra además que no siempre $c'X^*$ U-domina a $c'X^W$.

En la segunda sección tratamos el caso del octante positivo O_k^+ ya definido en el capítulo I. Como en ese capítulo consideramos $X = (X_1, \dots, X_k)$ donde X_i son variables aleatorias independientes con momento de orden 2 finito y provenientes de un modelo de localización $F(x, \theta) = F(x - \theta)$ (F función de distribución) donde el parámetro es la media de la distribución y tal que el modelo verifica la condición $\lim_{x \rightarrow \infty} x^2 F(-x) = 0$ y sabemos además que $\theta \in O_k^+$.

Comparamos en esta sección el estimador estandar X y estimadores intermedios entre este y el EMV X^* . Estos estimadores intermedios, siguiendo una tendencia general en la literatura, vamos a denominarlos con el nombre genérico de mixed. Trataremos con dos clases de estimadores de este tipo.

Primeramente consideramos los que vamos a denotar por X^α y que definimos como $X_i^\alpha = \max(X_i, \alpha X_i)$ donde $\alpha \in [0, 1]$ y $1 \leq i \leq k$. En lo relativo a la estimación global el parámetro θ probamos que el error cuadrático medio global de X^α es intermedio entre el de X y el de X^* y que X^α es minimax para cualquier $\alpha \in [0, 1]$ con lo que este criterio resulta poco útil a la hora de

discriminar entre estos estimadores. Seguidamente tratamos la estimación de funciones lineales y probamos que para cada $\alpha \in [0, 1)$ existe una dimensión k a partir de la cual no todas las funciones lineales de θ , $c'\theta$, se estiman mejor según el criterio del error cuadrático medio, utilizando $c'X^\alpha$ en lugar de $c'X$. Sin embargo esta dimensión es superior a la obtenida cuando tratábamos con X^* con lo que estos estimadores podrían tener alguna utilidad en ese sentido. Probamos además que fijada una dimensión k existe α_0 tal que $\forall \alpha \geq \alpha_0$ los estimadores mixed X^α verifican

$$E (c'X^\alpha - c'\theta)^2 \leq E (c'X - c'\theta)^2 \quad \forall c \in \mathfrak{R}^k$$

En la subsección siguiente tratamos la admisibilidad de estos estimadores dentro de la clase que hemos definido y encontramos condiciones suficientes, que involucran a α y c para afirmar cuando $c'X^\alpha$ es admisible para estimar $c'\theta$ en la clase de los estimadores mixed. Caracterizamos además las direcciones c para las que el único estimador admisible es $c'X^*$ (direcciones que incluyen a las coordenadas y a las diferencias de parámetros) y probamos que para el resto de las direcciones cualquier mixed X^α es admisible si la dimensión k es suficientemente grande. Particularizamos todos los resultados obtenidos a los modelos normal y doble exponencial.

Por otra parte el hecho de que los estimadores mixed anteriores sean inadmisibles frente al EMV para estimar las coordenadas nos llevan a definir otros estimadores mixed II que denotamos por X^β que tienen la particularidad de que el estimador del parámetro siempre está en la frontera del cono. De-

mostramos que, dentro de la clase que definen, y en el modelo normal, estos estimadores si son admisibles para estimar las coordenadas y las diferencias entre las coordenadas del parámetro θ .

3.1 Comparando el EMV con estimadores basados en otras métricas

En esta sección suponemos que $X \sim N(\theta, \Sigma)$ donde Σ es una matriz definida positiva, $\theta \in C = \{x \in \mathfrak{R}^k : a'x \geq 0\}$ con $a'\Sigma a = 1$. Como en capítulos anteriores denotamos por X^* el estimador máximo verosímil. Es decir $X^* = P_{\Sigma^{-1}}(X/C)$. Denotamos además por W una matriz simétrica, definida positiva cualquiera y por $X^W = P_W(X/C)$.

Tenemos entonces que

$$X^W = X I_{\{a'X \geq 0\}} + \left(X - \frac{a'X}{a'W^{-1}a} W^{-1}a \right) I_{\{a'X < 0\}} \quad (3.1)$$

y $X^* = X^{\Sigma^{-1}}$.

Teorema 3.1 *Supongamos que $\theta \in Fr(C)$. Entonces para cualquier $c \in \mathfrak{R}^k$ y cualquier $t > 0$*

$$\Pr(|c'X^W - c'\theta| \leq t) \leq \Pr(|c'X^* - c'\theta| \leq t)$$

Demostración. Supongamos, sin perdida de generalidad, que c es también un vector unitario según la métrica definida por Σ^{-1} . Es suficiente ver que

$$\Pr(|c'X^W - c'\theta| \leq t, a'X < 0) \leq \Pr(|c'X^* - c'\theta| \leq t, a'X < 0)$$

puesto que si $a'X \geq 0$ se tiene que $X^W = X^*$.

Sean

$$u = c - (c'\Sigma a) a \quad (3.2)$$

$$u_* = c - \frac{c'W^{-1}a}{a'W^{-1}a} a \quad (3.3)$$

Entonces $a'\Sigma u = 0$ y además teniendo en cuenta (3.1) y que por hipótesis $a'\theta = 0$ obtenemos

$$\{|c'X^* - c'\theta| \leq t, a'X < 0\} = \{|u'X - u'\theta| \leq t, a'X < 0\} \quad (3.4)$$

$$\{|c'X^W - c'\theta| \leq t, a'X < 0\} = \{|u_*'X - u_*'\theta| \leq t, a'X < 0\}$$

Ahora bien :

$$\begin{aligned} \Pr\left(|c'X^W - c'\theta| \leq t, a'X < 0\right) &= \Pr\left(|u_*'X - u_*'\theta| \leq t, a'X < 0\right) \\ &\leq \Pr\left(|u_*'X^* - u_*'\theta| \leq t, a'X < 0\right) \end{aligned} \quad (3.5)$$

donde la desigualdad es consecuencia del teorema 2.3, pag. 74. Además de la definición de u y por ser a un vector unitario para la métrica dada por Σ tenemos que

$$\begin{aligned} u_* - (u_*'\Sigma a) a &= \left(c - \frac{c'W^{-1}a}{a'W^{-1}a} a\right) - \left(c'\Sigma a - \frac{c'W^{-1}a}{a'W^{-1}a} (a'\Sigma a)\right) a \\ &= c - (c'\Sigma a) a = u \end{aligned}$$

lo que implica

$$\begin{aligned} \Pr\left(|u_*'X^* - u_*'\theta| \leq t, a'X < 0\right) &= \Pr\left(|u'X - u'\theta| \leq t, a'X < 0\right) \\ &= \Pr\left(|c'X^* - c'\theta| \leq t, a'X < 0\right) \end{aligned} \quad (3.6)$$

El resultado se obtiene entonces de (3.4), (3.5) y (3.6). ■

Veamos ahora que ocurre cuando el parámetro no necesariamente está en la frontera. Para que el resultado se mantenga vamos a tener que imponer cierta condición sobre la métrica W , condición que depende de la dirección c . Como las direcciones c y $-c$ son equivalentes en nuestro análisis vamos a suponer, sin pérdida de generalidad, que $a'\Sigma c \geq 0$. Obviamente, en la otra situación la condición sobre W sería la simétrica a la aquí dada.

Teorema 3.2 *Supongamos que $\theta \in C$ y que $u'W^{-1}a \leq 0$ donde $u = c - (c'\Sigma a)a$. Entonces, para cualquier $c \in \mathfrak{R}^k$ y cualquier $t > 0$*

$$\Pr\left(|c'X^W - c'\theta| \leq t\right) \leq \Pr\left(|c'X^* - c'\theta| \leq t\right) \quad (3.7)$$

Demostración. Suponemos sin pérdida de generalidad que $u'\theta = 0$ con lo cual de (3.2)

$$c'\theta = (u + (a'\Sigma c)a)'\theta = (a'\Sigma c)a'\theta \geq 0 \quad (3.8)$$

Además como $a'\Sigma u = 0$ se tiene que $a'X$ y $u'X$ son independientes. Para demostrar (3.7) basta como en ocasiones anteriores probar que

$$\Pr\left(|c'X^W - c'\theta| \leq t, a'X < 0\right) \leq \Pr\left(|c'X^* - c'\theta| \leq t, a'X < 0\right)$$

Ahora bien si denotamos por F la función de distribución de $a'X$ tenemos que $\Pr\left(|c'X^W - c'\theta| \leq t, a'X < 0\right)$ puede escribirse como

$$\int_{-\infty}^0 \Pr\left(|c'X^W - c'\theta| \leq t, a'X < 0 / a'X = z\right) dF(z) \quad (3.9)$$

Por tanto de (3.1), la independencia de $a'X$ y $u'X$ y (3.8), para cualquier $z < 0$

tenemos

$$\begin{aligned}
& \Pr\left(\left|c'X^W - c'\theta\right| \leq t, a'X < 0 / a'X = z\right) \\
&= \Pr\left(\left|c'\left(X - \frac{a'X}{a'W^{-1}a}W^{-1}a\right) - c'\theta\right| \leq t, a'X < 0 / a'X = z\right) \\
&= \Pr\left(\left|u'X + (a'\Sigma c)z - \frac{c'W^{-1}a}{a'W^{-1}a}z - c'\theta\right| \leq t, a'X < 0 / a'X = z\right) \\
&= \Pr\left(c'\theta - (a'\Sigma c)z + \left(\frac{c'W^{-1}a}{a'W^{-1}a}\right)z - t \leq u'X \leq c'\theta - (a'\Sigma c)z + \left(\frac{c'W^{-1}a}{a'W^{-1}a}\right)z + t\right) \\
&= \Pr\left(c'\theta + \left(\frac{u'W^{-1}a}{a'W^{-1}a}\right)z - t \leq u'X \leq c'\theta + \left(\frac{u'W^{-1}a}{a'W^{-1}a}\right)z + t\right) \\
&\leq \Pr(c'\theta - t \leq u'X \leq c'\theta + t) = \Pr(|u'X - c'\theta| \leq t)
\end{aligned} \tag{3.10}$$

donde la desigualdad proviene de una propiedad estandar de la distribución normal unidimensional, ya que por hipótesis se tiene que $u'W^{-1}a \leq 0$.

Entonces

$$\begin{aligned}
P\left(\left|c'X^W - c'\theta\right| \leq t, a'X < 0\right) &\leq P(|u'X - c'\theta| \leq t) \cdot P(a'X < 0) \\
&= P(|c'X^* - c'\theta| \leq t, a'X < 0)
\end{aligned}$$

donde la desigualdad es consecuencia de (3.9) y (3.10) y la igualdad se obtiene de la independencia de $u'X$ y $a'X$ y de (3.1). ■

Esta proposición no es cierta en general si no se verifica la condición $u'Wa \leq 0$ como queda de manifiesto en el siguiente ejemplo.

Ejemplo 3.1 Supongamos que $X \rightsquigarrow N_2(\theta, I_2)$ con $\theta \in C = \{x \in \mathfrak{R}^2 / x_2 \geq 0\}$.

Sean $c = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$, $W^{-1} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$ y $\theta = \begin{pmatrix} 0 \\ k \end{pmatrix}$ con $k \geq 0$. En estas condiciones para ciertos valores de t va a verificarse

$$\Pr\left(\left|c'X^W - c'\theta\right| \leq t\right) > \Pr(|c'X^* - c'\theta| \leq t)$$

Con las elecciones anteriores tenemos que $u = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$, $W^{-1}a = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ y por tanto $u'W^{-1}a = 1 > 0$.

Además de (3.1):

$$X^* = XI_{\{x_2 \geq 0\}} + (X - a'Xa)I_{\{x_2 < 0\}} = XI_{\{x_2 \geq 0\}} + \begin{pmatrix} x_1 \\ 0 \end{pmatrix} I_{\{x_2 < 0\}}$$

$$X^W = XI_{\{x_2 \geq 0\}} + \left(X - \frac{a'X}{a'Wa}Wa\right)I_{\{x_2 < 0\}} = XI_{\{x_2 \geq 0\}} + \begin{pmatrix} x_1 + x_2 \\ 0 \end{pmatrix} I_{\{x_2 < 0\}}$$

luego basta comparar, tomando $t = k$

$$P_2 = P_\theta \left(|c'X^W - c'\theta| \leq k, X \notin C \right)$$

con

$$P_1 = P_\theta \left(|c'X^* - c'\theta| \leq k, X \notin C \right)$$

Haciendo los cálculos pertinentes se obtiene:

$$P_1 = P_{(0,k)}(X_1 \in [-2k, 0], X_2 < 0) = \Phi(-k) \cdot \left(\frac{1}{2} - \Phi(-2k)\right)$$

y

$$\begin{aligned} P_2 &= P_{(0,k)}(X_1 + X_2 \in [-2k, 0], X_2 < 0) \\ &= \int_{-\infty}^0 \int_{-x_2-2k}^{-x_2} \frac{1}{2\pi} \exp\left\{-\frac{1}{2}(x_1^2 + (x_2 - k)^2)\right\} dx_1 dx_2 \\ &= \int_{-\infty}^0 \varphi(x_2 - k) (\Phi(-x_2) - \Phi(-x_2 - 2k)) dx_2 \end{aligned}$$

efectuando los cálculos con $k = 1$ tenemos:

$$P_1 = 0.0757 \quad \text{y} \quad P_2 = 0.09374$$

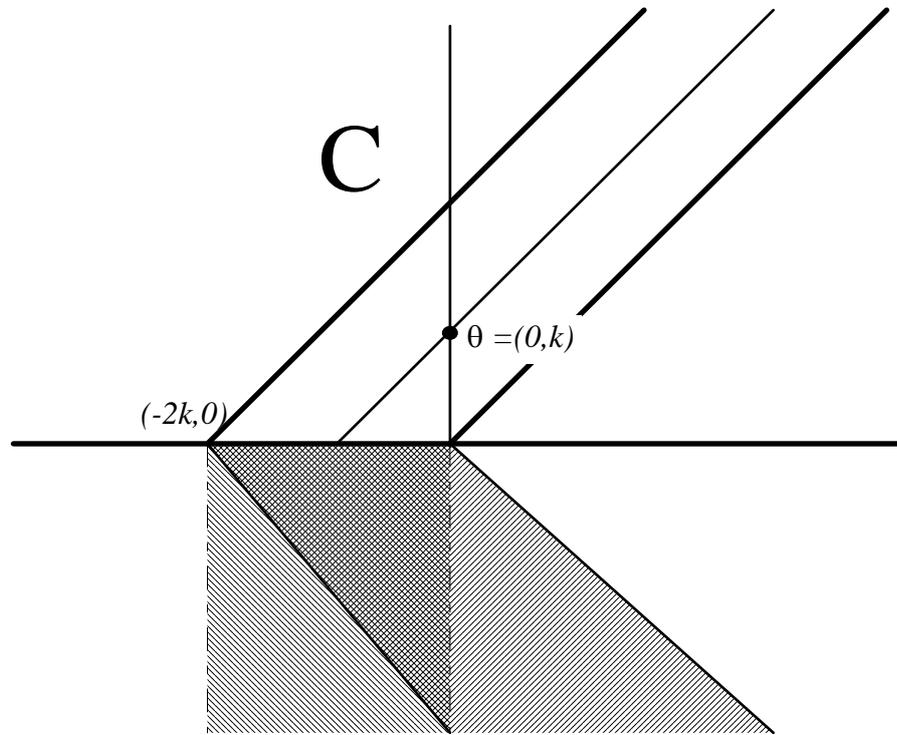


Figura 2: Esquema gráfico del ejemplo. P_1 es la probabilidad de la zona rayada hacia la derecha y P_2 la de la zona rayada hacia la izquierda es decir la desigualdad contraria a la del resultado anterior.

Nota 3.1 *A partir de este ejemplo es inmediato, sin más que considerar el giro apropiado, que este tipo de situación se tiene también cuando c es, por ejemplo, una de las aplicaciones coordenadas.*

Sin embargo el ejemplo anterior no es suficiente para establecer que el EMV restringido no U-domina a aquellos obtenidos por proyecciones que no verifican la desigualdad anterior. Es decir que a la vista de este ejemplo podría pensarse en un cierto recíproco del tipo:

Sea $X \rightsquigarrow N_k(\theta, \Sigma)$ con Σ definida positiva y $\theta \in C = \{x \in \mathfrak{R}^k / a'x \geq 0\}$. Entonces dados $a, c \in \mathfrak{R}^k$ con $a'c \geq 0$ y dada una métrica W verificando $u'W^{-1}a > 0$ existen $\theta \in C$ y $t > 0$ tales que

$$\Pr(|c'X^W - c'\theta| \leq t) > \Pr(|c'X^* - c'\theta| \leq t)$$

Sin embargo vamos a ver ahora que la afirmación anterior no es cierta.

Teorema 3.3 *Sea $X \rightsquigarrow N_k(\theta, I_k)$ tal que $\theta \in C = \{x \in \mathfrak{R}^k / a'x \geq 0\}$ con a unitario. Sea $c \in \mathfrak{R}^k$ verificando $a'c = 0$. Entonces para cualquier matriz W definida positiva y simétrica y para cualquier $t > 0$*

$$\Pr(|c'X^W - c'\theta| \leq t) \leq \Pr(|c'X^* - c'\theta| \leq t)$$

Demostración. Supongamos sin pérdida de generalidad que $a' = (1, 0, \dots, 0)$ con lo que $c' = (0, c_2, \dots, c_k)$, y también que $\theta' = (\theta_1, 0, \dots, 0)$. Denotemos además $W^{-1} = (\omega_{ij})_{(i,j) \in \{1, \dots, k\} \times \{1, \dots, k\}}$. Entonces tenemos que $u = c - (c'a)a = c$ y de (3.1)

$$X^* = XI_{\{a'X \geq 0\}} + \begin{pmatrix} 0 \\ X_2 \\ \vdots \\ X_k \end{pmatrix} I_{\{a'X < 0\}}$$

$$X^W = XI_{\{a'X \geq 0\}} + \begin{pmatrix} 0 \\ X_2 - \frac{\omega_{12}}{\omega_{11}} X_1 \\ \vdots \\ X_k - \frac{\omega_{1k}}{\omega_{11}} X_1 \end{pmatrix} I_{\{a'X < 0\}}$$

Por tanto

$$\Pr(|c'X^* - c'\theta| \leq t, X \notin C) = \Pr\left(\left|\sum_{i=2}^k c_i X_i\right| \leq t, X_1 < 0\right)$$

y por otra parte

$$\begin{aligned} & \Pr\left(|c'X^W - c'\theta| \leq t, X \notin C\right) \tag{3.11} \\ &= \Pr\left(\left|\sum_{i=2}^k c_i X_i - \frac{X_1}{\omega_{11}} \left(\sum_{i=2}^k c_i \omega_{1i}\right)\right| \leq t, X_1 < 0\right) \\ &= \int_{-\infty}^0 \Pr\left(\left|\sum_{i=2}^k c_i X_i - \frac{X_1}{\omega_{11}} \left(\sum_{i=2}^k c_i \omega_{1i}\right)\right| \leq t/X_1 = z\right) dF(z) \end{aligned}$$

donde $F(z)$ es la función de distribución de X_1 .

Ahora para cualquier $z < 0$

$$\begin{aligned} & \Pr\left(\left|\sum_{i=2}^k c_i X_i - \frac{X_1}{\omega_{11}} \left(\sum_{i=2}^k c_i \omega_{1i}\right)\right| \leq t/X_1 = z\right) \\ &= \Pr\left(-t + \frac{z}{\omega_{11}} \left(\sum_{i=2}^k c_i \omega_{1i}\right) \leq \sum_{i=2}^k c_i X_i \leq \frac{z}{\omega_{11}} \left(\sum_{i=2}^k c_i \omega_{1i}\right) + t\right) \\ &\leq \Pr\left(-t \leq \sum_{i=2}^k c_i X_i \leq t\right) = \Pr\left(\left|\sum_{i=2}^k c_i X_i\right| \leq t\right) \end{aligned}$$

Y entonces de (11) y $c'a = 0$

$$\begin{aligned} \Pr(|c'X^W - c'\theta| \leq t, X \notin C) &\leq \int_{-\infty}^0 \Pr\left(\left|\sum_{i=2}^k c_i X_i\right| \leq t\right) dF(z) \\ &= \Pr\left(\left|\sum_{i=2}^k c_i X_i\right| \leq t\right) \Pr(X_1 < 0) \end{aligned}$$

$$\begin{aligned}
&= \Pr \left(\left| \sum_{i=2}^k c_i X_i \right| \leq t, X_1 < 0 \right) \\
&= \Pr (|c'X^* - c'\theta| \leq t, X \notin C)
\end{aligned}$$

■

3.2 Estimadores mixed para el Octante Positivo

Consideramos $X = (X_1, \dots, X_k)$ donde X_i son variables aleatorias independientes con momento de orden 2 finito y provenientes de un modelo de localización $F(x, \theta) = F(x - \theta)$ (F función de distribución) donde el parámetro es la media de la distribución y tal que el modelo verifica $\lim_{x \rightarrow \infty} x^2 F(-x) = 0$. Suponemos que disponemos además de información adicional que nos dice que $\theta \in O_k^+$ (octante positivo de \mathfrak{R}^k).

Vamos a considerar primeramente el EMV X^* y los estimadores que denominamos mixed I y denotamos por X^α que se definen como sigue.

Definición 3.1 Dado $\alpha \in [0, 1]$, $X_i^\alpha = \max(X_i, \alpha X_i)$ para $1 \leq i \leq k$.

Los resultados que obtenemos para este modelo los particularizamos a los modelos normal y doble exponencial.

3.2.1 Estimación global del parámetro

El error cuadrático medio global de δ como estimador del parámetro $\theta \in \mathfrak{R}^k$ puede escribirse, en el caso de muestras independientes, como

$$E(\|\delta - \theta\|^2) = E\left(\sum_{i=1}^k (\delta_i - \theta_i)^2\right) = \sum_{i=1}^k E((\delta_i - \theta_i)^2) \quad (3.12)$$

Entonces podemos probar el siguiente teorema.

Teorema 3.4 *Para cualquier $\theta \in O^+$ se tiene que*

$$E(\|X^* - \theta\|^2) \leq E(\|X^\alpha - \theta\|^2) \leq E(\|X - \theta\|^2)$$

Demostración. Tenemos que $X = X^1$ y que $X^* = X^0$ y que además

$$E(X_i^\alpha - \theta_i)^2 = \int_0^\infty (x - \theta_i)^2 f(x) dx + \int_{-\infty}^0 (\alpha x - \theta_i)^2 f(x) dx$$

Podemos derivar bajo el signo integral y obtenemos para $\alpha > 0$

$$\begin{aligned} \frac{\partial}{\partial \alpha} E(X_i^\alpha - \theta_i)^2 &= \frac{\partial}{\partial \alpha} \int_{-\infty}^0 (\alpha x - \theta_i)^2 f(x) dx \\ &= \int_{-\infty}^0 2(\alpha x - \theta_i) x f(x) dx > 0 \end{aligned}$$

con lo cual $E(X_i^\alpha - \theta_i)^2$ es una función creciente de α si $\alpha > 0$. Entonces teniendo en cuenta (3.12) obtenemos el resultado. ■

Vamos a probar además que todos estos estimadores mixed son minimax. Para demostrar esto comenzamos recopilando un resultado de Kumar y Sharma (88). Seguidamente se da una lectura de este resultado bajo el punto de vista de la inferencia con restricciones más general que la del artículo antes mencionado y después ya más concretamente pasa a tratarse el caso de los estimadores mixed en el octante positivo.

Teorema 3.5 (Kumar y Sharma (88)) *Sea Σ_0 un subconjunto de \mathfrak{R}^k para el que existe una sucesión $\{(a_n^1, \dots, a_n^k) / n \geq 1\}$ tal que*

$$\lim_{n \rightarrow \infty} \{(\theta_1, \dots, \theta_k) / (\theta_1 + a_n^1, \dots, \theta_k + a_n^k) \in \Sigma_0\} = \mathfrak{R}^k$$

Si δ es un estimador verificando

$$R(\theta, \delta) \leq R < \infty \text{ siempre que } \theta \in \Sigma_0$$

donde R es el riesgo constante del estimador X , entonces δ es un estimador minimax de θ para $\theta \in \Sigma_0$.

Proposición 3.6 Sea C un cono de \mathfrak{R}^k de interior no vacío. Sea δ un estimador de $\theta \in C$ verificando

$$R(\theta, \delta) \leq R < \infty \text{ para } \theta \in C$$

donde R es el riesgo constante del estimador X . Entonces δ es minimax para θ cuando $\theta \in C$.

Demostración. Sea $a = (a_1, \dots, a_k)$ un elemento del interior del cono. Basta entonces tomar $\Sigma_0 = C$ y $a_n^i = a_i \cdot n$ en el teorema anterior. ■

Teorema 3.7 X^α es minimax para cualquier $\alpha \in [0, 1]$.

Demostración. Según el teorema de Kumar y Sharma (88) basta probar que el riesgo de estos estimadores es menor que k , que es el riesgo constante de estimador X , para todo $\theta \in O^+$ puesto que se verifican las condiciones del teorema tomando la sucesión

$$\left\{ (a_n^1, \dots, a_n^k) / n \geq 1 \right\} \text{ con } a_n^1 = \dots = a_n^k = n$$

con lo cual $\underline{\lim}_{n \rightarrow \infty} \left\{ (\theta_1, \dots, \theta_n) / (\theta_1 + a_n^1, \dots, \theta_n + a_n^k) \in C \right\} = \mathfrak{R}^k$.

Debe suceder entonces que:

$$R(X^\alpha, \theta) = \sum_{i=1}^k E((X_i^\alpha - \theta_i)^2) \leq R(X, \theta) = \sum_{i=1}^k E((X_i - \theta_i)^2)$$

luego basta que

$$E((X_i - \theta_i)^2) \geq E((X_i^\alpha - \theta_i)^2), \text{ para cualquier } \theta_i \geq 0 \text{ y para } 1 \leq i \leq k$$

y esto se sigue de la demostración del teorema 3.4. ■

Se puede afirmar entonces que este criterio no parece muy adecuado a la hora de valorar situaciones de este tipo puesto que no es capaz de discriminar entre los estimadores mixed.

Se puede dar una explicación del porque de estos resultados y que constituye el fundamento del teorema de Kumar y Sharma. El criterio trata de encontrar estimadores cuyo riesgo máximo sea lo menor posible con lo cual debe poner a cada estimador en la peor posible de las situaciones. En el caso no restringido el estimador minimax X tiene riesgo constante R . Por otra parte un estimador minimax para el problema no restringido δ no puede tener un riesgo máximo inferior a R ya que cuando el parámetro tiende a ∞ en todas sus componentes el estimador se comporta aproximadamente como si la información adicional no existiera y el problema fuera no restringido. Basta ver entonces que en ningún elemento de O^+ el estimador δ supera el riesgo uniforme de un estimador minimax X para el problema no restringido, puesto que ese riesgo máximo no puede reducirse. En este caso eso sucede porque los riesgos de los estimadores mixed para las coordenadas son siempre menores que los

del estimador minimax del problema no restringido bajo la función de pérdida cuadrática.

3.2.2 Estimación de funciones lineales del parámetro

En esta sección obtenemos primeramente resultados para la estimación de funciones lineales del parámetro θ a través de los estimadores mixed. Seguidamente tratamos la cuestión de la admisibilidad de estos estimadores mixed dentro de la clase que ellos mismos definen y que incluye al EMV. El modelo sigue siendo el descrito al comienzo de esta sección, es decir que suponemos que $X = (X_1, \dots, X_k)$ donde X_i son variables aleatorias independientes con momento de orden 2 finito y provenientes de un modelo de localización $F(x, \theta) = F(x - \theta)$ (F función de distribución) donde el parámetro θ es la media de la distribución y tal que el modelo verifica la condición $\lim_{x \rightarrow \infty} x^2 F(-x) = 0$. Suponemos que disponemos además de información adicional que nos dice que $\theta \in O_k^+$ (octante positivo de \mathfrak{R}^k).

Error cuadrático medio de los estimadores mixed en la estimación de funciones lineales

Como en el capítulo primero vamos a necesitar para el teorema posterior un lema previo. Sea Y una variable aleatoria de media θ proveniente de dicho modelo de localización verificando $\lim_{y \rightarrow \infty} y^2 F(-y) = 0$ y sean

$$\begin{aligned} h(\alpha, \theta) &= (1 + \alpha) E_\theta \left((Y^-)^2 \right) + 2\theta E_\theta (Y^-) \\ s(\theta) &= E_\theta (Y^-) \end{aligned}$$

entonces se tiene el siguiente lema.

Lema 3.8 *La función*

$$f_k(\alpha, \theta) = \frac{\sum_{i=1}^k c_i^2 h(\alpha, \theta_i)}{(1 - \alpha) \sum_{i=1}^k c_i^2 s^2(\theta_i)}$$

alcanza su mínimo en $\theta \in O_k^+$ cuando $\theta = 0$ para cualesquiera $\alpha \in [0, 1]$ y $c = (c_1, \dots, c_k) \in \mathfrak{R}^k$.

La demostración de este lema es similar a la del lema 1.3, pag. 29 y por esta razón no se incluye aquí. A partir de este lema probamos el resultado principal de esta subsección válido para cualquier modelo de localización que verifique las condiciones impuestas al comienzo de esta sección.

Teorema 3.9

$$E\left((c'(X^\alpha - \theta))^2\right) \leq E\left((c'(X - \theta))^2\right) \quad \forall c \in \mathfrak{R}^k, \forall \theta \in O_k^+$$

si y solamente si

$$k \leq 1 + \frac{1 + \alpha}{1 - \alpha} \cdot \frac{E_0\left(\left(X_i^-\right)^2\right)}{\left(E_0\left(X_i^-\right)\right)^2} \quad (3.13)$$

Demostración.

$$E\left(c'(X - \theta)\right)^2 = \sum_{i=1}^k c_i^2 E\left(X_i - \theta_i\right)^2 \quad (3.14)$$

mientras que

$$E\left(c'(X^\alpha - \theta)\right)^2 = \sum_{i=1}^k c_i^2 E\left(X_i^\alpha - \theta_i\right)^2 + 2 \sum_{i < j} c_i c_j E\left(X_i^\alpha - \theta_i\right) E\left(X_j^\alpha - \theta_j\right) \quad (3.15)$$

Calculamos entonces $E\left(X_i^\alpha - \theta_i\right)^2$ y $E\left(X_i^\alpha - \theta_i\right)$. Ahora bien:

$$X^\alpha - \theta = (X - \theta) I_{\{X \geq 0\}} + (\alpha X - \theta) I_{\{X < 0\}}$$

entonces

$$\begin{aligned} E(X^\alpha - \theta) &= E(X - \theta) + (\alpha - 1) E_\theta(X I_{\{X < 0\}}) \\ &= (1 - \alpha) E_\theta(X^-) = (1 - \alpha) s(\theta) \end{aligned} \quad (3.16)$$

Por otra parte:

$$\begin{aligned} E(X^\alpha - \theta)^2 &= E((X - \theta)^2 I_{\{X \geq 0\}}) + E((\alpha X - \theta)^2 I_{\{X < 0\}}) \\ &= E((X^+ - \theta)^2 - \theta^2 I_{\{X < 0\}}) + E((\alpha X^- + \theta)^2 - \theta^2 I_{\{X \geq 0\}}) \\ &= E(X - \theta + X^-)^2 - \theta^2 I_{\{X < 0\}} + E((\alpha X^-)^2 + 2\alpha\theta X^- + \theta^2 I_{\{X < 0\}}) \\ &= E((X - \theta)^2) - (1 - \alpha) \left((1 + \alpha) E((X^-)^2) + 2\theta E(X^-) \right) \\ &= E((X - \theta)^2) - (1 - \alpha) h(\alpha, \theta) \end{aligned} \quad (3.17)$$

Finalmente restando (3.15) de (3.14) y teniendo en cuenta (16) y (3.17) tenemos denotando $d_i = c_i s(\theta_i)$

$$\begin{aligned} &E(c'(X - \theta))^2 - E(c'(X^\alpha - \theta))^2 \\ &= (1 - \alpha) \sum_{i=1}^k c_i^2 h(\alpha, \theta_i) - 2(1 - \alpha)^2 \sum_{i < j} c_i c_j s(\theta_i) s(\theta_j) \\ &= (1 - \alpha) \sum_{i=1}^k c_i^2 h(\alpha, \theta_i) - (1 - \alpha)^2 \left((d' \cdot 1)^2 - (d' \cdot d) \right) \\ &= (1 - \alpha) \sum_{i=1}^k c_i^2 h(\alpha, \theta_i) - (1 - \alpha)^2 (d' \cdot d) - k(1 - \alpha)^2 (d' \cdot d) \cos^2(d, 1) \end{aligned}$$

de donde se obtiene

$$E(c'(X - \theta))^2 - E(c'(X^\alpha - \theta))^2 \geq 0 \quad \forall c \in \mathfrak{R}^k, \forall \theta \in O_k^+$$

si y solamente si $\forall c \in \mathfrak{R}^k, \forall \theta \in O_k^+$

$$k \leq \frac{1}{\cos^2(d, 1)} \left(1 + \frac{\sum_{i=1}^k c_i^2 h(\alpha, \theta_i)}{(1 - \alpha)(d' \cdot d)} \right)$$

Entonces utilizando el lema previo obtenemos inmediatamente la desigualdad (3.13). ■

Nota 3.2 Como el lado derecho de la desigualdad (3.13) es creciente en α tenemos como consecuencia que para un k dado podemos obtener el α mínimo para el que se produce reducción de error cuadrático en la estimación de cualquier función lineal del parámetro. De hecho obtendríamos puesto que

$$f_k(0,0) = \frac{E_0((X_i^-)^2)}{(E_0(X_i^-))^2}$$

$$\alpha \geq 1 - \frac{2f_k(0,0)}{k-1+f_k(0,0)}$$

Esto puede tener interés por ejemplo en el caso de un modelo normal $Y = X\beta + \varepsilon$ en el que el experimentador controle la matriz de diseño de tal forma que pueda conseguir que $(X'X) = I$. Entonces dado el número de regresores puede calcularse el α mínimo correspondiente a ese k dado.

Nota 3.3 El teorema anterior no prueba sin embargo que el vértice del cono sea el punto más desfavorable en la estimación de una función lineal fija cualquiera $c'\theta$. De hecho esto no es cierto como prueba el Ejemplo 1.5, pag. 40.

Tratamos ahora dos situaciones particulares habituales que consideramos de interés.

Ejemplo 3.2 Distribución Normal

Sea $X_i \rightsquigarrow N(\theta_i, \sigma^2)$ entonces teniendo en cuenta los cálculos hechos en el Ejemplo 2 del Capítulo 1.2 obtenemos que X^α es mejor que el estimador es-

tandar X , para estimar cualquier función lineal del parámetro cuando se sabe que este pertenece al octante positivo cuando

$$k \leq 1 + \frac{1 + \alpha}{1 - \alpha} \pi$$

El α mínimo que puede utilizarse en dimensión k para asegurarse una estimación con menor error cuadrático medio en la estimación de cualquier función lineal del parámetro sería entonces en este caso, cuando $k \geq 5$

$$\alpha \geq \frac{k - 1 - \pi}{k - 1 + \pi}$$

Ejemplo 3.3 Distribución Doble Exponencial

Análogamente cuando $X_i \rightsquigarrow DExp(\theta_i, \lambda)$ según los cálculos del Ejemplo 3 del Capítulo 1.2 y los resultados aquí obtenidos el estimador mixed X^α tiene menor error cuadrático medio que el estándar en la estimación de cualquier función lineal de θ si y solamente si

$$k \leq 1 + 2 \frac{1 + \alpha}{1 - \alpha}$$

Admisibilidad local de los estimadores mixed en la estimación de

funciones lineales Se pasa ahora a estudiar la admisibilidad de estos estimadores dentro de la familia de los mixed. Se considera concretamente la admisibilidad frente al estimador máximo verosímil de los estimadores mixed para funciones lineales del parámetro. El estudio se hace originalmente para modelos de localización bastante generales. Los resultados se particularizan después a los modelos normal y doble exponencial.

La función $E(c'(X^\alpha - \theta))^2$ es una función cuadrática en α relativamente controlable cuando $\theta_i = \theta$, $1 \leq i \leq k$. Estudiando entonces el vértice de esta parábola (que es siempre un mínimo) se puede deducir la admisibilidad para algunos de los estimadores de tal forma que si un punto α_0 es vértice de una de esas parábolas para un valor fijo de c , no solo $c'X^{\alpha_0}$ sino todos $c'X^\alpha$ con $\alpha \leq \alpha_0$ resultan ser estimadores admisibles (frente al EMV) para ese c dentro de la clase de los estimadores mixed.

Para las direcciones $c \in \mathfrak{R}^k$ tales que c ó $-c$ pertenecen al interior del cono del octante positivo podemos probar el siguiente resultado.

Teorema 3.10 *Sea $c \in \mathfrak{R}^k$ verificando $\cos^2(c, 1) > \frac{1}{k}$. Entonces $c'X^\alpha$ es admisible dentro de los mixed para estimar $c'\theta$ si*

$$0 \leq \alpha \leq \left(1 + \frac{E_0 \left((X_i^-)^2 \right)}{(k \cos^2(c, 1) - 1) \left(E_0 (X_i^-) \right)^2} \right)^{-1} \quad (3.18)$$

Demostración. Suponemos entonces que $\theta_1 = \dots = \theta_k = \theta = E(X_i)$.

Utilizando ahora (3.15), (16) y (3.17) tenemos:

$$\begin{aligned} E(c'(X^\alpha - \theta))^2 &= ((c'1)^2 - (c'c)) (1 - \alpha)^2 E_\theta^2(X_i^-) + \\ &+ (c'c) \left(E(X_i - \theta)^2 - (1 - \alpha) \left((1 + \alpha) E_\theta(X_i^-)^2 + 2\theta E_\theta(X_i^-) \right) \right) \\ &= \alpha^2 (c'c) \left((k \cos^2(c, 1) - 1) E_\theta^2(X_i^-) + E_\theta(X_i^-)^2 \right) + \\ &+ 2\alpha (c'c) \left(- (k \cos^2(c, 1) - 1) E_\theta^2(X_i^-) + \theta E_\theta(X_i^-) \right) + \Delta(\theta) \end{aligned}$$

y el coeficiente de α^2 es claramente positivo con lo cual esta función cuadrática en α tiene un mínimo.

Como el vértice de la parábola $a\alpha^2 + b\alpha + c$ está en el punto $\frac{-b}{2a}$ obtenemos como valor de ese vértice

$$V(\theta) = \frac{(k \cos^2(c, 1) - 1) E_\theta^2(X_i^-) - \theta E_\theta(X_i^-)}{(k \cos^2(c, 1) - 1) E_\theta^2(X_i^-) + E_\theta(X_i^-)^2}$$

Hay que tener en cuenta además que si esa función de θ toma un valor α_0 , no solo ese valor de α sino todos aquellos $\alpha \in [0, \alpha_0]$ nos dan estimadores mixed admisibles dentro de la clase. Esto es consecuencia de que el numerador de esa función es negativo cuando θ es suficientemente grande. Notar que $E_\theta(X_i^-) = s(\theta)$ es decreciente en θ puesto que $s'(\theta) = -F(-\theta)$ y por tanto $(k \cos^2(c, 1) - 1) E_\theta(X_i^-) - \theta \rightarrow -\infty$ cuando $\theta \rightarrow \infty$. Un argumento de continuidad prueba entonces la afirmación.

Para finalizar la prueba veamos que $V(\theta)$ alcanza su máximo cuando $\theta = 0$ para $\theta \geq 0$. Basta probar que la función

$$\begin{aligned} V_2(\theta) &= \frac{(k \cos^2(c, 1) - 1) E_\theta^2(X_i^-)}{(k \cos^2(c, 1) - 1) E_\theta^2(X_i^-) + E_\theta(X_i^-)^2} \\ &= \left(1 + \frac{E_\theta\left(\left(X_i^-\right)^2\right)}{(k \cos^2(c, 1) - 1) \left(E_\theta(X_i^-)\right)^2} \right)^{-1} \end{aligned}$$

alcanza su máximo en 0 cuando $\theta \geq 0$, o de forma equivalente que el mínimo de $\frac{E_\theta\left(\left(X_i^-\right)^2\right)}{\left(E_\theta(X_i^-)\right)^2}$ para $\theta \geq 0$ está en $\theta = 0$.

$$\frac{\partial}{\partial \theta} \frac{E_\theta\left(\left(X_i^-\right)^2\right)}{\left(E_\theta(X_i^-)\right)^2} = \frac{-2E_\theta^2(X_i^-) + 2E_\theta\left(\left(X_i^-\right)^2\right) F(-\theta)}{E_\theta^3(X_i^-)}$$

Esta derivada es positiva como consecuencia de la desigualdad de Schwarz ya que

$$E_\theta^2(X_i^-) = \left(\int_{-\infty}^0 y f(y - \theta) dy \right)^2$$

$$\leq \left(\int_{-\infty}^0 y^2 f(y - \theta) dy \right) \left(\int_{-\infty}^0 f(y - \theta) dy \right) = E_{\theta} \left((X^-)^2 \right) F(-\theta)$$

Por tanto $\frac{E_{\theta} \left((X_i^-)^2 \right)}{\left(E_{\theta} (X_i^-) \right)^2}$ es creciente en θ y su mínimo bajo la condición $\theta \geq 0$ se alcanza en $\theta = 0$. ■

Podemos dar fácilmente una cota superior para el lado derecho de (3.18) sin más que tener en cuenta que $\frac{E_0 \left((X_i^-)^2 \right)}{\left(E_0 (X_i^-) \right)^2} \geq 1$. Obtenemos entonces inmediatamente que

$$\left(1 + \frac{E_0 \left((X_i^-)^2 \right)}{\left(k \cos^2(c, 1) - 1 \right) \left(E_0 (X_i^-) \right)^2} \right)^{-1} \leq 1 - \frac{1}{k} \left(1 + \tan^2(c, 1) \right) \quad (3.19)$$

Además de (3.18) es claro que la dirección más favorable para obtener estimadores mixed admisibles es la diagonal. Obtenemos entonces el siguiente corolario.

Corolario 3.11 *Existe $c \in \mathfrak{R}^k$ tal que $c'X^{\alpha}$ es admisible para estimar $c'\theta$ al menos para los valores α verificando*

$$0 \leq \alpha \leq \left(1 + \frac{E_0 \left((X_i^-)^2 \right)}{\left(k - 1 \right) \left(E_0 (X_i^-) \right)^2} \right)^{-1} \quad (3.20)$$

Además según (3.19) el extremo superior de este intervalo está acotado superiormente por $1 - \frac{1}{k}$.

Por otra parte, cuando la dirección c es la diagonal es claro de (3.20) que el extremo superior del intervalo de admisibilidad tiende a 1 cuando $k \rightarrow \infty$ con lo cual todos los estimadores son admisibles siempre que k sea suficientemente grande. También se obtiene de la expresión (3.18) que si al aumentar

k tomamos un c tal que $\cos^2(c, 1) = cte < \frac{1}{k}$ sucede lo mismo; es decir que el intervalo de admisibilidad crece con k hasta incluir cualquier estimador de la clase de los mixed X^α . Esto lo resumimos en el siguiente teorema.

Teorema 3.12 *Sea $c \in \mathfrak{R}^k$ tal que $\cos^2(c, 1) = p < \frac{1}{k}$. Entonces*

$$\forall \alpha_0 < 1, \exists k_0 \in N \text{ tal que } \forall k > k_0, \exists \theta \in O_k^+$$

verificando

$$\arg \min_{\alpha} E(c'(X^\alpha - \theta))^2 \geq \alpha_0$$

Podemos ahora particularizar los resultados obtenidos a los modelos Normal y Doble Exponencial.

Ejemplo 3.4 Distribución Normal

En el caso normal con los cálculos efectuados anteriormente tenemos que utilizando (3.18) los estimadores mixed son admisibles para estimar $c'\theta$ con $\cos^2(c, 1) > \frac{1}{k}$ al menos cuando

$$0 \leq \alpha \leq \left(1 + \frac{\pi}{k \cos^2(c, 1) - 1}\right)^{-1}$$

y el extremo superior del intervalo en la dirección diagonal del cono es $\frac{k-1}{k-1+\pi}$.

Ejemplo 3.5 Distribución Doble Exponencial

Para este otro caso particular la expresión (3.18) se transforma en

$$0 \leq \alpha \leq \left(1 + \frac{2}{k \cos^2(c, 1) - 1}\right)^{-1}$$

y el extremo para la dirección diagonal nos queda $\frac{k-1}{k+1}$.

Sin embargo la técnica empleada hasta ahora no nos permite encontrar estimadores admisibles dentro de los mixed cuando $\cos^2(c, 1) \leq \frac{1}{k}$, es decir cuando se estima una dirección en la frontera o en el exterior del cono. Según el Teorema 3.9 de Moors (85) en algunas situaciones como en aquella en la que c es una aplicación coordinada no solo no podemos obtenerla mediante esta técnica sino que el EMV domina a los estimadores mixed. Esto es claro sin más que considerar la demostración que Moors hace de su teorema aunque para el caso particular del octante positivo ya lo hemos probado en el teorema 3.10. Quizá sea conveniente recordar en este momento que en este capítulo no se pretenden obtener estimadores admisibles para los parámetros sino comparar los estimadores que hemos definido con el EMV para conocer mejor su comportamiento.

Además hay otras direcciones para las que también el Estimador Máximo Verosímil domina a todos los estimadores mixed. Para las direcciones c tales que $c_i = 0$ para cualquier $i \neq i_1, i_2$ y además $c_{i_1} \cdot c_{i_2} \leq 0$ obtenemos el siguiente resultado.

Teorema 3.13 *Sea $c \in \mathfrak{R}^k$ verificando las condiciones anteriores. Entonces para cualquier $\alpha \in [0, 1]$ y $\theta \in O_k^+$*

$$E(c'(X^* - \theta))^2 \leq E(c'(X^\alpha - \theta))^2$$

Demostración. Si escribimos de nuevo $E(c'(X^\alpha - \theta))^2 = a\alpha^2 + b\alpha + d$ como función de α pero ahora sin suponer los parámetros iguales en cada

coordenada de las ecuaciones (3.15), (16) y (3.17) obtenemos que el coeficiente de grado 2 de esta función cuadrática de α es de nuevo positivo y que además

$$-\frac{b}{2} = -\sum c_i^2 \theta_i E(X_i^-) + 2 \sum_{i < j} c_i c_j E(X_i^-) E(X_j^-) \quad (3.21)$$

En este caso particular la ecuación anterior se transforma en

$$-\frac{b}{2} = -c_{i_1}^2 \theta_{i_1} E(X_{i_1}^-) - c_{i_2}^2 \theta_{i_2} E(X_{i_2}^-) - 2c_{i_1} c_{i_2} E(X_{i_1}^-) E(X_{i_2}^-) < 0$$

Con lo cual ninguno de los vértices de esas parábolas está en el octante positivo y los estimadores son todos inadmisibles frente al Estimador Máximo Verosímil ya que por la forma de estas funciones el mínimo de $a\alpha^2 + b\alpha + c$ en $\alpha \in [0, 1]$ si el vértice de la parábola es negativo se alcanza en $\alpha = 0$ que es el valor correspondiente al EMV. ■

En dimensión 2 puede resolverse completamente el problema en el sentido de que los valores de c exteriores al cono y fuera del octante negativo tienen al EMV como único estimador admisible dentro de los mixed y el interior del cono tiene un límite superior α_0 para cada c que determina el intervalo de admisibilidad de los estimadores. Podemos entonces enunciar el siguiente resultado.

Teorema 3.14 *Si $k = 2$ existen estimadores admisibles dentro de la clase de los mixed para estimar $c'\theta$ si y solamente si c ó $-c$ pertenecen al interior de O_2^+ .*

Los gráficos ilustran la situación que se tiene en \mathfrak{R}^2 para un modelo normal con varianza 1. En la figura 3 aparece el error cuadrático medio de los esti-

madores mixed $c'X^\alpha$ como función de θ cuando el parámetro es un elemento de la diagonal del cono; el valor de este parámetro aparece por tanto en el eje de abscisas. En esta figura, c es la diagonal del cono y α toma los valores 0, 0.1, 0.2, 0.3. La curva correspondiente a $\alpha = 0$ (el EMV) es la más gruesa, la de trazos corresponde a $\alpha = 0.3$, la intermedia a 0.2 y la restante a 0.1. Notar que para esta situación habíamos obtenido como extremo superior de nuestro intervalo de admisibilidad $\frac{1}{\pi+1}$ con lo cual los estimadores correspondientes a $\alpha = 0.1, 0.2$ son también admisibles entre los mixed como se observa en el gráfico.

En cuanto a la figura 4 tiene la misma estructura que la anterior para $c = (1, -1)$. Habíamos obtenido que todos los estimadores mixed en esta situación son inadmisibles frente al EMV como refleja este gráfico.

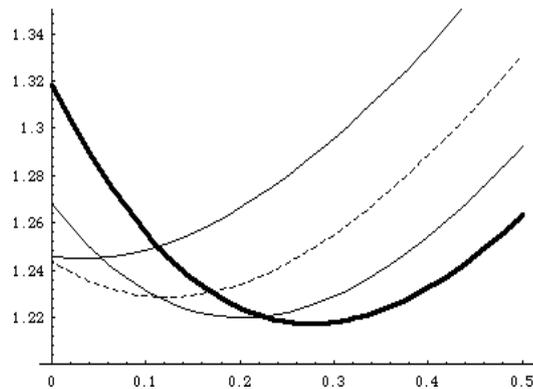


Figura 3: ECM para estimadores mixed en modelo normal cuando c es la diagonal de O_2^+

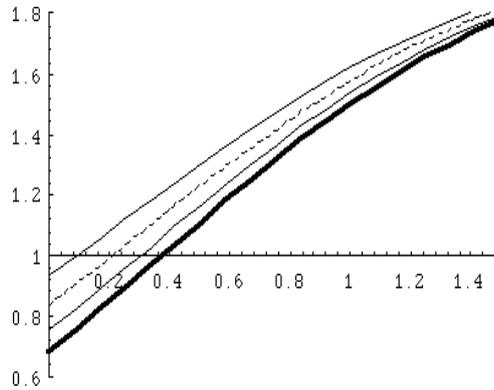


Figura 4: ECM para estimadores mixed en modelo normal cuando $c = (1, -1)$ en \mathfrak{R}^2

Consideramos ahora la situación en la que el vector c y el vector $-c$ están en el exterior del cono del octante positivo y no verifican la condición anterior en la que probamos la no existencia de estimadores admisibles en la clase diferentes del EMV. Es decir cuando $\cos^2(c, 1) < \frac{1}{k}$. De hecho la situación que falta por estudiar es aquella en la que existen más de dos coeficientes distintos de 0 en c y no todos ellos tienen el mismo signo. Para obtener estimadores admisibles en esta situación vamos a imponer la condición adicional $\lim_{\theta \rightarrow \infty} \theta E(X^-) = 0$ (donde $\theta = E(X)$).

Teorema 3.15 *Sea un modelo verificando $\lim_{\theta \rightarrow \infty} \theta E(X^-) = 0$. Sea $c \in \mathfrak{R}^k$ tal que existen i_1, i_2, i_3 con $c_{i_1}, c_{i_2}, c_{i_3} \neq 0$ y $c_{i_1} \cdot c_{i_2} < 0$. Entonces existe $\alpha > 0$ tal que $c'X^\alpha$ es admisible para estimar $c'\theta$ dentro de la clase de los mixed.*

Demostración. Basta, como en teoremas anteriores, probar que el mínimo de $E\left(c'(X^\alpha - \theta)^2\right)$ como función de α se encuentra en algún $\alpha > 0$ para algún valor de θ dentro del octante positivo. Como hemos visto antes esta función resulta ser convexa con lo que el coeficiente de α^2 es positivo en cualquier caso. Analizamos por tanto el vértice de la parábola $a\alpha^2 + b\alpha + c$, es decir el punto $-\frac{b}{2a}$ que en estas condiciones resulta ser menor que 1 y como $a > 0$ basta encontrar para cada c un valor de θ tal que en ese punto $-\frac{b}{2} > 0$. La ecuación (3.21) ya obtenida nos da el valor de $-\frac{b}{2}$.

$$-\frac{b}{2} = -\sum c_i^2 \theta_i E\left(X_i^-\right) + 2 \sum_{i < j} c_i c_j E\left(X_i^-\right) E\left(X_j^-\right)$$

El vector c puede escribirse como $\left(c_1, \frac{a_1}{\dots}, c_p, -c_{p+1}, \frac{a_2}{\dots}, -c_{p+q}, 0, \dots, 0\right)$ donde $c_i > 0$. Además podemos suponer que $p \geq 2$ considerando el vector $-c$ si es necesario. Se consideran entonces los puntos θ de tal forma que: $\theta_1 = \dots = \theta_p = 0$ y $\theta_{p+1} = \dots = \theta_{p+q} = \theta^0$ con lo cual:

$$-\frac{b}{2} = \epsilon + 2 \sum_{\substack{i < j \\ c_i, c_j \in a_1}} c_i c_j E_0\left(X_i^-\right) E_0\left(X_j^-\right)$$

y además $|\epsilon| \rightarrow 0$ cuando $\theta^0 \rightarrow \infty$ por hipótesis.

Por tanto $-\frac{b}{2} > 0$ para algún θ del octante positivo. ■

Si queremos delimitar los intervalos de admisibilidad que obtenemos con esta técnica, haciendo uso de los resultados anteriores obtenemos los siguientes valores.

Corolario 3.16 *Sea $c \in \mathfrak{R}^k$ tal que existen i_1, i_2, i_3 con $c_{i_1}, c_{i_2}, c_{i_3} \neq 0$ y $c_{i_1} \cdot c_{i_2} < 0$. Entonces $c'X^\alpha$ es admisible dentro de los mixed como estimador de*

$c'\theta$ al menos cuando α es mayor que 0 y menor o igual que

$$\max \left(\left(1 + \frac{E_0((X_i^-)^2)/(E_0(X_i^-))^2}{(p \cos^2(c_{a_1}, 1) - 1)} \right)^{-1}, \left(1 + \frac{E_0((X_i^-)^2)/(E_0(X_i^-))^2}{(q \cos^2(c_{a_2}, 1) - 1)} \right)^{-1} \right)$$

si $c = (c_{a_1}, c_{a_2}, 0)$ con $c_{a_1} = (c_1, \dots, c_p)$, $c_{a_2} = (-c_{p+1}, \dots, -c_{p+q})$ y $c_i > 0$ para

$$1 \leq i \leq p + q.$$

Ejemplo 3.6 Distribución Normal

En el caso normal tenemos que $E(X^-) = \varphi(\theta) + \theta\Phi(-\theta)$ y se puede probar, mediante la regla de l'Hopital, que se verifica la condición anterior con lo que el resultado es valido y además pueden acotarse inferiormente los vértices máximos; es decir el valor máximo de α , α_0 , tal que $c'X^\alpha$ es admisible como estimador de $c'\theta$ dentro de los estimadores mixed.

Haciendo los cálculos y con las notaciones anteriores se puede probar que en el caso normal:

$$\alpha_0 \geq \max \left(\left(1 + \frac{\pi}{p \cos^2(c_{a_1}, 1) - 1} \right)^{-1}, \left(1 + \frac{\pi}{q \cos^2(c_{a_2}, 1) - 1} \right)^{-1} \right)$$

si $c = (c_{a_1}, c_{a_2}, 0)$ con $c_{a_1} = (c_1, \dots, c_p)$, $c_{a_2} = (-c_{p+1}, \dots, -c_{p+q})$ y $c_i > 0$ para

$$1 \leq i \leq p + q.$$

Lo que hace básicamente el procedimiento anterior es reducir el caso $\cos^2(c, 1) < \frac{1}{k}$ al caso en el que c está en la frontera del cono llevando las variables cuyos coeficientes en c son negativos al infinito de forma que su influencia se haga mínima y prevalezca entonces la de las variables cuyo coeficiente es positivo transformando en cierto modo el c original en uno de la frontera del cono que tiene todos sus elementos positivos o cero y al menos dos de ellos positivos

que es la situación para la que ya se había probado la existencia de otros estimadores admisibles dentro de los mixed; evitándose precisamente la situación en la que uno solo de los elementos es positivo que es la que se transformaría en las aplicaciones coordenadas para las que no existe ningún estimador admisible distinto del EMV dentro de los mixed.

3.2.3 Otro tipo de estimadores mixed en O^+

En vista de que los estimadores mixed definidos anteriormente resultan inadmisibles frente al EMV en dos casos particularmente interesantes como son la estimación de coordenadas y las diferencias de medias, se definen en esta sección, para el caso particular del modelo normal, otro tipo de estimadores mixed siguiendo el modelo de Vijayasree y Singh (93). Vamos a denotar estos estimadores mixed II por X^β donde $\beta = (\beta_1, \dots, \beta_k)$ y los definimos como

Definición 3.2 Si $X \in C$ entonces $X^\beta = X$. Y cuando $X \notin C$:

$$X_i^\beta = \begin{cases} 0 & \text{si } X_i \leq 0 \\ \beta_i X_i & \text{si } X_i > 0 \end{cases} \quad \text{para } i = 1, 2, \dots, k$$

El valor del estimador está, en cualquier caso en la clausura del cono del octante positivo. Es decir que se analizan proyecciones sobre C con una métrica no unidad. En concreto dentro de esta clase de estimadores mixed se estudian la admisibilidad de X_i^β como estimador de θ_i y de $X_i^\beta - X_j^\beta$ como estimador de $\theta_i - \theta_j$ a través del error cuadrático medio de estos estimadores. En esta sección vamos a considerar solamente el caso normal que como ya hemos dicho es el más interesante; suponemos entonces que $X \rightsquigarrow N(\theta, 1)$ donde $\theta \geq 0$.

Damos primero un lema que contiene los cálculos necesarios. Utilizando la siguiente notación

$$\begin{aligned} s(\theta) &= \varphi(\theta) - \theta\Phi(-\theta) \\ t(\theta) &= \varphi(\theta) + \theta\Phi(\theta) \end{aligned} \quad (3.22)$$

$$f(\beta, \theta) = -(1 + \beta)(\Phi(\theta) - \theta\varphi(\theta)) - 2\beta\theta\varphi(\theta) + (1 - \beta)\theta^2\Phi(\theta) \quad (3.23)$$

Y denotando además mediante $g(\beta_i, \beta_j, \theta_i, \theta_j)$ la función

$$(\beta_i\varphi(\theta_i) - (1 - \beta_i)\theta_i\Phi(\theta_i))(\beta_j\varphi(\theta_j) - (1 - \beta_j)\theta_j\Phi(\theta_j)) - \varphi(\theta_i)\varphi(\theta_j) \quad (3.24)$$

podemos probar que

Lema 3.17

$$\begin{aligned} E(X_i^\beta - \theta_i)^2 &= (\Phi(\theta_i) - \theta_i\varphi(\theta_i) + \theta_i^2\Phi(-\theta_i)) + \\ &+ \left(1 - \prod_{j \neq i} \Phi(\theta_j)\right) (1 - \beta_i) f(\beta_i, \theta_i) \end{aligned} \quad (3.25)$$

Y además

$$\begin{aligned} &E\left(\left(X_i^\beta - \theta_i\right)\left(X_j^\beta - \theta_j\right)\right) \\ &= (1 - \beta_j)\theta_i\Phi(-\theta_i)t(\theta_j) + (1 - \beta_i)\theta_j\Phi(-\theta_j)t(\theta_i) \\ &+ s(\theta_i)s(\theta_j) + \left(1 - \prod_{k \neq i, j} \Phi(\theta_k)\right) g(\beta_i, \beta_j, \theta_i, \theta_j) \end{aligned} \quad (3.26)$$

Demostración.

$$\begin{aligned} E(X_i^\beta - \theta_i)^2 &= \Pr(X_j > 0, j \neq i) E(X_i^\beta - \theta_i)^2 + \\ &+ \Pr(X_j < 0, \text{ para algún } j \neq i) E(X_i^\beta - \theta_i)^2 \end{aligned}$$

Teniendo en cuenta la definición de los estimadores y utilizando para la primera esperanza (3.17) con el parámetro α igual a 0 tenemos que la expresión anterior se transforma en

$$\begin{aligned} & \left(1 - \prod_{j \neq i} \Phi(\theta_j) \right) \left(\Phi(\theta_i) - \theta_i \varphi(\theta_i) + \theta_i^2 \Phi(-\theta_i) \right) + \\ & + \left(\prod_{j \neq i} \Phi(\theta_j) \right) \left(\int_{-\infty}^0 \theta_i^2 \varphi(x - \theta_i) dx + \int_0^{\infty} (\beta_i x - \theta_i)^2 \varphi(x - \theta_i) dx \right) \end{aligned}$$

por otra parte la última integral vale

$$\begin{aligned} \int_0^{\infty} (\beta_i x - \theta_i)^2 \varphi(x - \theta_i) dx &= \int_0^{\infty} (\beta_i (x - \theta_i) - (1 - \alpha_{\beta_i}) \theta_i)^2 \varphi(x - \theta_i) dx \\ &= \beta_i^2 (\Phi(\theta_i) - \theta_i \varphi(\theta_i)) + (1 - \beta_i)^2 \theta_i^2 \Phi(\theta_i) \\ &\quad - 2\beta_i (1 - \beta_i) \theta_i \varphi(\theta_i) \end{aligned}$$

desarrollando el cuadrado y calculando las tres intregales resultantes. Tenemos entonces (25) tras sustituir y efectuar algunas operaciones. Para (26) :

$$\begin{aligned} & E \left((X_i^\beta - \theta_i) (X_j^\beta - \theta_j) \right) \\ &= \iint_C (x_i - \theta_i) (x_j - \theta_j) \varphi(x_i - \theta_i) \varphi(x_j - \theta_j) dx_i dx_j + \\ &\quad + \theta_i \theta_j P(X_i < 0, X_j < 0) - \theta_i P(X_i < 0) \int_0^{\infty} (\beta_j x_j - \theta_j) \varphi(x_j - \theta_j) dx_j - \\ &\quad - \theta_j P(X_j < 0) \int_0^{\infty} (\beta_i x_i - \theta_i) \varphi(x_i - \theta_i) dx_i + \\ &\quad + \left(1 - \prod_{k \neq i, j} \Phi(\theta_k) \right) \int_0^{\infty} \int_0^{\infty} (\beta_i x_i - \theta_i) (\beta_j x_j - \theta_j) \varphi(x_i - \theta_i) \varphi(x_j - \theta_j) dx_i dx_j \end{aligned}$$

descomponiendo la situación en las regiones adecuadas. La primera cuando el valor observado está en C , la segunda, tercera y cuarta cuando X_i, X_j o ambos valores son negativos y la última cuando a pesar de ser los dos positivos

alguno de los otros valores es negativo. Resolviendo estas integrales obtenemos (26). ■

Podemos probar entonces el resultado principal.

Teorema 3.18 *Para cualesquiera $\beta_i, \beta_j \in [0, 1]$, X_i^β es un estimador admisible para el parámetro θ_i y $X_i^\beta - X_j^\beta$ es admisible para estimar $\theta_i - \theta_j$ dentro de la clase de estimadores mixed II. Además si $\beta_i > 1$, entonces X_i^β es inadmisibile.*

Demostración. Utilizando (25) y denotando por X^* el EMV, que es también X^1 tenemos:

$$E(X_i^\beta - \theta_i)^2 - E(X_i^* - \theta_i)^2 = \left(1 - \prod_{j \neq i} \Phi(\theta_j)\right) (1 - \beta_i) f(\beta_i, \theta_i)$$

donde $f(\beta_i, \theta_i)$ está definido en (23).

Ahora si $\beta_i > 1$ se verifica que $E(X_i^\beta - \theta_i)^2 > E(X_i^* - \theta_i)^2$ puesto que en esas condiciones $f(\beta_i, \theta_i) < 0$ con lo cual todos los estimadores mixed de este tipo son inadmisibles frente al EMV.

Por otra parte la función $(1 - \beta) f(\beta, \theta)$ es una función cuadrática de β con coeficiente de $\beta^2 > 0$ (con un mínimo por tanto) y además cuando $\theta = 0$ se verifica

$$f(\beta, 0) = -\frac{1 + \beta}{2} \leq 0 \quad \text{si } \beta \geq -1 \quad (3.27)$$

con lo que la función que consideramos es negativa en $\theta = 0$ si $\beta \in [0, 1]$ y por tanto el valor de esta función parabólica en su vértice es también negativo.

Este vértice está, si tomamos en consideración (23) en el punto:

$$\beta^*(\theta) = \frac{\theta t(\theta)}{\Phi(\theta) + \theta \varphi(\theta) + \theta^2 \Phi(\theta)} \quad (3.28)$$

donde $t(\theta)$ está definido en (22).

Ahora bien como $\beta^*(0) = 0$ y además es claro de (3.28) que $\beta^*(\theta) \xrightarrow{\theta \rightarrow \infty} 1$ tenemos la parte del teorema referida a la estimación de coordenadas, si notamos que la aplicación inversa de la definida en (3.28) es claramente inyectiva; con lo cual para cada β en $[0, 1)$ podemos encontrar un valor de $\theta \geq 0$ en el que se alcanza el mínimo y además valores diferentes para valores de β diferentes. En cuanto al extremo superior del intervalo su condición de límite de los vértices garantiza la admisibilidad de este valor frente a toda la clase por la simetría de las parábolas respecto a su vértice.

En cuanto a la comparación de dos poblaciones, esto es el análisis de la admisibilidad de $X_i^\beta - X_j^\beta$ como estimador de $\theta_i - \theta_j$ en la clase pertinente se tiene de (25) y (26):

$$\begin{aligned}
& E \left((X_i^\beta - X_j^\beta - (\theta_i - \theta_j))^2 \right) - E \left((X_i^* - X_j^* - (\theta_i - \theta_j))^2 \right) \\
&= E \left(X_i^\beta - \theta_i \right)^2 + E \left(X_j^\beta - \theta_j \right)^2 - 2E \left((X_i^\beta - \theta_i) (X_j^\beta - \theta_j) \right) - \\
&- \left(E \left(X_i^* - \theta_i \right)^2 + E \left(X_j^* - \theta_j \right)^2 - 2E \left((X_i^* - \theta_i) (X_j^* - \theta_j) \right) \right) \\
&= \left(1 - \prod_{k \neq i} \Phi(\theta_k) \right) (1 - \beta_i) [f(\beta_i, \theta_i)] + \left(1 - \prod_{k \neq j} \Phi(\theta_k) \right) (1 - \beta_j) [f(\beta_j, \theta_j)] - \\
&- 2 \left[(1 - \beta_j) \theta_i \Phi(-\theta_i) t(\theta_j) + (1 - \beta_i) \theta_j \Phi(-\theta_j) t(\theta_i) \right] - \\
&- 2 \left(1 - \prod_{k \neq i, j} \Phi(\theta_k) \right) g(\beta_i, \beta_j, \theta_i, \theta_j)
\end{aligned} \tag{3.29}$$

donde las funciones $f(\beta, \theta)$, $t(\theta)$ y $g(\beta_i, \beta_j, \theta_i, \theta_j)$ están definidas en (23), (22) y (3.24).

Si ahora hacemos $\theta_k \xrightarrow{k \neq i, j} \infty$ entonces el último sumando tiende a 0 y además

$1 - \prod_{k \neq i} \Phi(\theta_k) \simeq \Phi(-\theta_j)$. Entonces la expresión (3.29) tiende a:

$$(1 - \beta_i) \Phi(-\theta_j) [f(\beta_i, \theta_i) - 2\theta_j t(\theta_i)] + (1 - \beta_j) \Phi(-\theta_i) [f(\beta_j, \theta_j) - 2\theta_i t(\theta_j)] \quad (3.30)$$

Si volvemos a (3.27) vemos que la expresión (3.30) es menor o igual que 0 cuando $\theta_i = \theta_j = 0$ para cualesquiera β_i, β_j entre 0 y 1 con lo cual de nuevo las parábolas que consideremos toman algún valor negativo y como su vértice corresponde a un mínimo el valor en este punto es también negativo.

Fijamos ahora un valor de β_j y calculamos el vértice de la parábola correspondiente cuando $\theta_k \xrightarrow{k \neq i, j} \infty$. Según la expresión (3.30) tenemos:

$$\beta_i^{**}(\theta) \xrightarrow{\theta_k \xrightarrow{k \neq i, j} \infty} \frac{(\theta_i - \theta_j) t(\theta_i)}{\Phi(\theta_i) + \theta_i \varphi(\theta_i) + \theta_i^2 \Phi(\theta_i)}$$

Entonces si

$$\theta_i = \theta_j = \beta_j \text{ tenemos } \beta_i^{**}(\theta) \xrightarrow{\theta_k \xrightarrow{k \neq i, j} \infty} 0$$

y también si en estas condiciones tomamos $\theta_j = \beta_j$ y $\theta_i \rightarrow \infty$ entonces

$$\frac{(\theta_i - \theta_j) t(\theta_i)}{\Phi(\theta_i) + \theta_i \varphi(\theta_i) + \theta_i^2 \Phi(\theta_i)} \rightarrow 1$$

Entonces utilizando argumentos similares al del caso de las coordenadas puede concluirse que $\forall \beta_i, \beta_j \in [0, 1]$ el estimador $X_i^\beta - X_j^\beta$ es admisible dentro de la clase que se está considerando ya que para cada pareja de valores hemos encontrado un elemento del octante positivo en el que las parábolas alcanzan su mínimo. Como antes los extremos se solventan a través de pasos al límite.

■

Capítulo 4

Un Modelo Uniforme de Escala

En este capítulo estudiamos cuestiones relativas a la estimación con restricciones pero en un modelo que, a diferencia de los considerados anteriormente, es un modelo de escala. Suponemos entonces que:

X_i es una variable aleatoria $U(0, \theta_i)$ para $1 \leq i \leq k$ que define la i -ésima población

n_i es el tamaño de la muestra extraída en la población i -ésima

$\hat{\theta}_i = \max(X_{i1}, \dots, X_{in_i})$ es el EMV no restringido de θ_i

Los modelos restringidos de escala están recibiendo cierta atención en los últimos años. Estos modelos han sido tratados recientemente en, por ejemplo, Kushary y Cohen (89) y Hwang y Peddada (94). El modelo que aquí estudiamos aparece en Effessi y Pal (92) y Misra y Dhariyal (95) si bien el tratamiento de los primeros se refiere exclusivamente al caso de dos poblaciones y los segundos estudian la admisibilidad de cierto tipo de estimadores bajo diferentes condiciones.

Nosotros comparamos los estimadores máximo verosímiles con restricciones y sin restricciones utilizando el criterio del error cuadrático medio cuando se

estiman funciones lineales de los parámetros y para dos tipos importantes de restricciones: tree order y orden total.

Los resultados obtenidos demuestran que, tanto en el caso de restricciones del tipo del orden total como del tree order, existen determinadas combinaciones lineales de θ , c para las que el EMV es mejor pero también se encuentran combinaciones lineales para las que es mejor el EMV sin restricciones.

Misra y Dhariyal (95) proponen, en el caso de restricciones de orden total, estimadores restringidos frente a los cuales el EMV, por ejemplo, es inadmisibles. Nosotros demostraremos que estimadores propuestos en este artículo tampoco superan, en todos los casos a otros estimadores similares que no tienen en cuenta la información adicional suministrada por las restricciones de orden.

Por otra parte los resultados aquí obtenidos demuestran que la parte (ii) del teorema 4.6 y el teorema 4.7 de Hwang y Peddada (94), que tratan de la estimación de coordenadas en modelos de escala, no se pueden generalizar a todos los modelos de escala, como sus autores afirman.

4.1 Tree Order

Vamos a suponer que $\theta_0 \geq \theta_i$ para $1 \leq i \leq k$; es decir la desigualdad inversa a la usual. Utilizamos esta restricción puesto que va a ser mucho más cómoda a la hora de calcular los EMV y manejar las diferencias, en las combinaciones lineales de los parámetros entre los estimadores restringidos y no restringidos. Además los EMV que se obtienen en este tipo de restricción guardan mayor

similitud con los que se obtienen en el orden total que los obtenidos con la desigualdad contraria. Sea entonces $C_k = \{x \in \mathfrak{R}^k : x_0 \geq x_i \text{ para } 1 \leq i \leq k-1\}$. Cuando $\theta \in C_k$ es inmediato comprobar que el estimador máximo verosímil θ^* de dicho vector viene dado por

$$\begin{aligned}\theta_i^* &= \hat{\theta}_i \text{ para } 1 \leq i \leq k-1 \\ \theta_0^* &= \max(\hat{\theta}_0, \dots, \hat{\theta}_{k-1})\end{aligned}$$

Como $\theta_i^* = \hat{\theta}_i$ para $1 \leq i \leq k-1$ la diferencia de los estimadores con y sin restricciones viene dada por el estimador para la coordenada raíz para la que, como veremos en la proposición siguiente " θ_0^* está más concentrado en torno a θ_0 que $\hat{\theta}_0$ ".

Proposición 4.1 *Para cualquier $t > 0$*

$$\Pr(|\theta_0^* - \theta_0| < t) \geq \Pr(|\hat{\theta}_0 - \theta_0| < t)$$

Demostración. Esto es evidente puesto que de $\hat{\theta}_i < \theta_i \leq \theta_0$ para $0 \leq i \leq k-1$ deducimos $\hat{\theta}_0 \leq \theta_0^* < \theta_0$. ■

Con los mismos argumentos obtenemos otras direcciones, $c'\theta$, en las que se obtiene un resultado del mismo tipo.

Proposición 4.2 *Si $c_i \geq 0$ para $0 \leq i \leq k-1$ entonces para cualquier $t > 0$*

$$\Pr(|c'\theta^* - c'\theta| < t) \geq \Pr(|c'\hat{\theta} - c'\theta| < t)$$

Nota 4.1 *Obviamente el resultado también es cierto si $c_i \leq 0$ para $0 \leq i \leq k-1$ y la prueba es la misma.*

Nota 4.2 Hay que notar que no solo tenemos reducción estocástica de error sino que además para cualquier $t > 0$ si $c_i \geq 0$ para $0 \leq i \leq k-1$

$$\{x \in \mathfrak{R}^k : |c'\hat{\theta} - c'\theta| < t\} \subset \{x \in \mathfrak{R}^k : |c'\theta^* - c'\theta| < t\}$$

Análogamente en la estimación simultánea de coordenadas, si $\tau = (t_0, \dots, t_{k-1})$

y

$$A_\tau^* = \{x \in \mathfrak{R}^k : |\theta_i^* - \theta_i| < t_i \text{ para } 0 \leq i \leq k-1\}$$

$$\hat{A}_\tau = \{x \in \mathfrak{R}^k : |\hat{\theta}_i - \theta_i| < t_i \text{ para } 0 \leq i \leq k-1\}$$

tenemos que $\hat{A}_\tau \subset A_\tau^*$ para cualquier τ con componentes positivas. En consecuencia la probabilidad de cubrimiento de intervalos de confianza simultáneos de longitud fija para las coordenadas es menor cuando dichos intervalos están centrados en el EMV restringido que cuando lo están en el EMV sin restricciones.

Por otro lado en el próximo teorema se demuestra que en el caso particular de que $n_i = n$ para $0 \leq i \leq k-1$ existen direcciones para las que el error cuadrático medio es menor utilizando $\hat{\theta}$ en lugar de θ^* a partir de un k .

Teorema 4.3 Existe k_0 tal que si $(\sum_{i=1}^{k-1} c_i)/c_0 < -(n+1)/(n+2)$ y $k \geq k_0$, existe $\theta \in C_k$ verificando

$$E((c'\theta^* - c'\theta)^2) > E((c'\hat{\theta} - c'\theta)^2)$$

Demostración. De $\theta_i^* = \hat{\theta}_i$ para $1 \leq i \leq k-1$ se tiene que

$$E((c'\theta^* - c'\theta)^2) - E((c'\hat{\theta} - c'\theta)^2) = c_0^2 \left(E(\theta_0^* - \theta_0)^2 - E(\hat{\theta}_0 - \theta_0)^2 \right) +$$

$$+2c_0 \left(\sum_{i=1}^{k-1} c_i \left(E \left((\theta_0^* - \theta_0) (\hat{\theta}_i - \theta_i) \right) - E \left((\hat{\theta}_0 - \theta_0) (\hat{\theta}_i - \theta_i) \right) \right) \right) \quad (4.1)$$

Sea $\theta_i = \alpha$ para $0 \leq i \leq k-1$. Entonces tenemos que cuando k tiende a ∞ , $\theta_0^* \xrightarrow{c.s.} \theta_0$ y

$$\begin{aligned} E(\theta_0^* - \theta_0)^2 &\longrightarrow 0 \\ E\left((\theta_0^* - \theta_0)(\hat{\theta}_i - \theta_i)\right) &\longrightarrow 0 \text{ para } 1 \leq i \leq k-1 \end{aligned}$$

Como además para el θ elegido

$$E(\hat{\theta}_0 - \theta_0) = E(\hat{\theta}_i - \theta_i) = \frac{-1}{n+1}\alpha$$

y

$$E(\hat{\theta}_0 - \theta_0)^2 = \frac{2}{(n+1)(n+2)}\alpha^2$$

tenemos que de la expresión (1)

$$\begin{aligned} &\lim_{k \rightarrow \infty} E\left((c'\theta^* - c'\theta)^2\right) - E\left((c'\hat{\theta} - c'\theta)^2\right) \\ &= \lim_{k \rightarrow \infty} -\frac{2c_0^2}{(n+1)(n+2)}\alpha^2 - 2c_0 \left(\sum_{i=1}^{k-1} c_i \right) \left(\frac{-1}{n+1}\alpha \right)^2 \end{aligned}$$

y esta última expresión es positiva si y solo si

$$\frac{\sum_{i=1}^{k-1} c_i}{c_0} < \frac{n+1}{n+2}$$

como queríamos probar. ■

Ejemplo 4.1 Un ejemplo de dirección del tipo de las estudiadas en el teorema anterior es para $c'\theta = \theta_0 - \theta_i$ con lo cual la estimación de la diferencia entre

el parámetro raíz y cualquiera de los otros utilizando el EMV $\theta_0^* - \theta_i^*$ no es siempre mejor que la obtenida a través del estimador no restringido $\hat{\theta}_0 - \hat{\theta}_i$.

Por otro lado en el teorema siguiente demostramos que se puede obtener un error cuadrático medio menor para el EMV restringido aumentando los tamaños muestrales para $1 \leq i \leq k - 1$.

Teorema 4.4 *Para cualquier $c \in \mathfrak{R}^k$*

$$\lim_{\substack{n_i \rightarrow \infty \\ 1 \leq i \leq k-1}} E \left((c'\theta^* - c'\theta)^2 \right) - E \left((c'\hat{\theta} - c'\theta)^2 \right) \leq 0$$

Demostración. Cuando $n_i \rightarrow \infty$ para $1 \leq i \leq k - 1$ tenemos que $\theta_i^* \xrightarrow{c.s.} \theta_i$ con lo que

$$\begin{aligned} E(\theta_i^* - \theta_i)^2 &\longrightarrow 0 \\ E((\theta_0^* - \theta_0)(\hat{\theta}_i - \theta_i)) &\longrightarrow 0 \end{aligned}$$

y de (1) tenemos

$$\begin{aligned} &\lim_{\substack{n_i \rightarrow \infty \\ 1 \leq i \leq k-1}} E \left((c'\theta^* - c'\theta)^2 \right) - E \left((c'\hat{\theta} - c'\theta)^2 \right) \\ &= c_0^2 \left(E(\theta_0^* - \theta_0)^2 - E(\hat{\theta}_0 - \theta_0)^2 \right) < 0 \end{aligned}$$

■

4.2 Orden Total

En esta sección suponemos que el parámetro verifica las restricciones correspondientes a un orden total, es decir que $\theta \in C'_k = \{x \in \mathfrak{R}^k : x_1 \leq \dots \leq x_k\}$.

El estimador máximo verosímil θ^* es

$$\theta_i^* = \max(\hat{\theta}_1, \dots, \hat{\theta}_i) \text{ para } 1 \leq i \leq k$$

Como en el caso del tree order y utilizando el mismo tipo de argumentos podemos probar que el estimador restringido es mejor para la estimación de las coordenadas.

Proposición 4.5 *Para $1 \leq i \leq k$ y para cualquier $t > 0$*

$$\Pr(|\theta_i^* - \theta_i| < t) \geq \Pr(|\hat{\theta}_i - \theta_i| < t)$$

También tenemos la siguiente proposición para las direcciones con coordenadas de igual signo

Proposición 4.6 *Si $\text{signo}(c_i) = \text{signo}(c_j)$ para $1 \leq i, j \leq k$ entonces para cualquier $t > 0$*

$$\Pr(|c'\theta^* - c'\theta| < t) \geq \Pr(|c'\hat{\theta} - c'\theta| < t)$$

Y en lo referente a la estimación simultánea de coordenadas si denotamos $\tau = (t_1, \dots, t_k)$ donde $t_i \geq 0$ para $1 \leq i \leq k$ tenemos que

Proposición 4.7 *Para cualquier $\tau \in O_k^+$, $\hat{A}_\tau \subset A_\tau^*$ donde*

$$A_\tau^* = \{x \in \mathfrak{R}^k : |\theta_i^* - \theta_i| < t_i \text{ para } 1 \leq i \leq k\}$$

$$\hat{A}_\tau = \{x \in \mathfrak{R}^k : |\hat{\theta}_i - \theta_i| < t_i \text{ para } 1 \leq i \leq k\}$$

Respecto a la estimación de otras funciones lineales generales de θ , en esta situación se tiene

$$E\left((c'\theta^* - c'\theta)^2\right) - E\left((c'\hat{\theta} - c'\theta)^2\right) = \sum_{i=1}^k c_i^2 \left(E(\theta_i^* - \theta_i)^2 - E(\hat{\theta}_i - \theta_i)^2\right) + 2 \sum_{i < j} c_i c_j \left(E\left((\theta_i^* - \theta_i)(\theta_j^* - \theta_j)\right) - E\left((\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)\right)\right) \quad (4.2)$$

La complejidad de (2) no nos permite encontrar tan específicamente como en el tree order situaciones en las que no hay reducción de error aunque si algunas como la estimación de la diferencia entre los parámetros extremos cuando el número de poblaciones es suficientemente grande. Supongamos también por comodidad que $n_1 = \dots = n_k = n$.

Teorema 4.8 *Existe k_0 tal que para cualquier $k \geq k_0$, $\exists \theta \in C'_k$ verificando*

$$E\left((\theta_1^* - \theta_k^*) - (\theta_1 - \theta_k)\right)^2 > E\left((\hat{\theta}_1 - \hat{\theta}_k) - (\theta_1 - \theta_k)\right)^2$$

Demostración. Como $\hat{\theta}_1 = \theta_1^*$ la expresión (2) se escribe como

$$\begin{aligned} & \left(E(\theta_k^* - \theta_k)^2 - E(\hat{\theta}_k - \theta_k)^2\right) \\ & - 2 \left(E\left((\theta_k^* - \theta_k)(\hat{\theta}_1 - \theta_1)\right) - E\left((\hat{\theta}_k - \theta_k)(\hat{\theta}_1 - \theta_1)\right)\right) \end{aligned} \quad (4.3)$$

si ahora elegimos θ tal que $\theta_i = \alpha$ para $1 \leq i \leq k$ y tomamos $k \rightarrow \infty$ tenemos que $\theta_k^* \xrightarrow{c.s.} \theta_k$ y (3) tiende a

$$-\frac{2}{(n+1)(n+2)}\alpha^2 + 2\left(\frac{1}{n+1}\alpha\right)^2 > 0$$

■

Misra y Dhariyal (95) dan condiciones sobre la no admisibilidad de estimadores de los parámetros, en este mismo tipo de modelos uniformes, cuando se sabe que estos verifican un orden total y proponen estimadores que mejoran al EMV en esta situación. Sin embargo en el ejemplo siguiente vemos que dichos estimadores también pueden verse superados por estimadores no restringidos.

Ejemplo 4.2 Uno de los estimadores que aparecen en Misra y Dhariyal (95) viene definido por $\delta = (\delta_1 \dots, \delta_k)$ tal que

$$\delta_1 = \frac{n_1 + 2}{n_1 + 1} \theta_1^*$$

$$\delta_i = \frac{N + 2}{N + 1} \theta_i^* \text{ para } 2 \leq i \leq k$$

donde $N = \sum_{i=1}^k n_i$. Sea $\gamma = \left(\frac{n_1+2}{n_1+1} \hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k \right)$. Entonces si $n_1 = \dots = n_k = n$ y $\theta_i = \alpha$ para $1 \leq i \leq k$ y consideramos la función lineal $c_1 \theta_1 + c_k \theta_k$ obtenemos que si hacemos $k \rightarrow \infty$

$$E((c_1 \delta_1 + c_k \delta_k) - (c_1 \theta_1 + c_k \theta_k))^2 - E\left(\left(c_1 \frac{n+2}{n+1} \hat{\theta}_1 + c_k \hat{\theta}_k\right) - (c_1 \theta_1 + c_k \theta_k)\right)^2$$

se aproxima a

$$-\frac{2c_k^2}{(n+1)(n+2)} \alpha^2 - 2\frac{c_1 c_k}{(n+1)^3} \alpha^2$$

y esta diferencia es positiva cuando

$$\frac{c_1}{c_k} < -\frac{(n+1)^2}{n+2}$$

Nota 4.3 Hwang y Peddada (94) tratan modelos de escala en la sección 4 de su artículo. Prueban que el EMV del primer parámetro de un orden total domina al estimador no restringido y afirman además que para cualquier modelo de

escala, verificando un orden total el estimador máximo verosímil no domina universalmente al estimador no restringido sino que para cada valor de $\theta \in C'_k$ existe una constante positiva c tal que

$$\Pr(|\theta_k^* - \theta_k| < c) < \Pr(|\hat{\theta}_k - \theta_k| < c) \quad (4.4)$$

donde θ_k^* y $\hat{\theta}_k$ son los EMV restringido y no restringido para la última coordenada respectivamente.

También aseguran que en cualquier modelo de escala, bajo las restricciones de un tree order $\theta_1 \leq \theta_i$ el EMV de la primera coordenada θ_1^* no domina universalmente a $\hat{\theta}_1$. Sin embargo como se verá estas dos últimas afirmaciones no son válidas en nuestro modelo de escala.

Hwang y Peddada basan estas afirmaciones en que si $\hat{\theta}_k \leq \theta_k^*$ y la desigualdad es estricta con probabilidad positiva entonces

$$\Pr(0 < \theta_k^*/\theta_k < c_+) < \Pr(0 < \hat{\theta}_k/\theta_k < c_+) \quad (4.5)$$

cuando se elige $c_+ = 1 + c/\theta_k$ y $c > \theta_k$. De aquí se deduce inmediatamente (4.4) puesto que $\theta_k > 0$. Sin embargo y a pesar de que en nuestro modelo bajo el orden total es cierto que $\hat{\theta}_k \leq \theta_k^*$ y esta desigualdad es estricta con probabilidad positiva, las probabilidades que aparecen en (4.5) valen 1 para esa elección de constantes.

Para el caso del tree order usual $\theta_1 \leq \theta_i$, $1 \leq i \leq k$ se tiene que $\theta_i^* = \max(\hat{\theta}_1, \hat{\theta}_i)$ para $1 \leq i \leq k$ con lo que también $\hat{\theta}_i \leq \theta_i^* < \theta_i$ y

$$\Pr(|\theta_i^* - \theta_i| < c) \geq \Pr(|\hat{\theta}_i - \theta_i| < c)$$

contradiciendo el Teorema 4.7 de Hwang y Peddada.

Parece por tanto que los Teoremas 4.6 (ii) y 4.7 de dicha referencia no son tan generales como sus autores afirman, sino que para que se verifiquen hay que añadir alguna condición que involucre al soporte de las variables, por ejemplo que dicho soporte sea \mathbb{R}^+ .

Bibliografía

- [1] Abelson, R. P. and Tukey, J. W. (1963). Efficient utilization of non-numerical information in quantitative analysis: general theory and the case of the simple order. *Ann. Math. Statist.* **34**, 1347-1369.
- [2] Akkerboom, J. C. (1990). *Testing Problems with Linear or Angular Inequality Constraints*. Springer-Verlag, New York.
- [3] Anderson, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* **6**, 170-176.
- [4] Bartholomew, D. J. (1961). A test for homogeneity of means under restricted alternatives (with discussion). *J. Roy. Statist. Soc. B* **23**, 239-381.
- [5] Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk H. D. (1972). *Statistical Inference under Order Restrictions*. Wiley, New York.
- [6] Bickel, P. J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Ann. Statist.* **9**, 1301-1309.
- [7] Blumenthal, S. and Cohen, A. (1968). Estimation of two ordered translation parameters. *Ann. Math. Statist.* **39**, 517-530.
- [8] Bohrer, R. (1973). An optimality property of Scheffé bounds. *Ann. Statist.* **1**, 766-772.
- [9] Bowden, D. C. (1970). Simultaneous confidence bands for linear regression models. *J. Amer. Statist. Assoc.* **65**, 413-421.
- [10] Calvin, J. (1990). Minimax estimation of a bounded normal vector. *J. Multiv. Analys.* **35**, 130-139.

- [11] Casella, G. and Strawderman, W. E. (1981). Estimating a bounded normal mean. *Ann. Statist.* **9**, 870-878.
- [12] Charras, A. and Van Eeden, C. (1991a). Bayes and admissibility properties of estimators in truncated parameter spaces. *Canadian J. Statist.* **19**, 121-134.
- [13] Charras, A. and Van Eeden, C. (1991b). Limits of Bayes estimators in convex, truncated parameter spaces. *Stat. Prob. Let.* **11**, 479-483.
- [14] Charras, A. and Van Eeden, C. (1992). Bayes properties of estimators of location parameters in truncated parameter spaces. *Stat. and Dec.* **10**, 81-86.
- [15] Cohen, A. and Sackrowitz, H. (1970). Estimation of the last mean of a monotone sequence. *Ann. Math. Statist.* **41**, 2021-2034.
- [16] Dasgupta, A. (1985). Bayes minimax estimation in multiparameter families when the parameter space is restricted to a bounded convex set. *Sankhyā, Series A* **47**, 326-332.
- [17] Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *J. Amer. Statist. Assoc.* **50**, 1096-1121.
- [18] Dykstra, R. L. and Robertson, T. (1982). Order restricted statistical tests on multinomial and Poisson parameters: the starshaped restriction. *Ann. Statist.* **10**, 1246-1252.
- [19] Dykstra, R. L. and Robertson, T. (1983). On testing monotone tendencies. *J. Amer. Statist. Assoc.* **78**, 342-350.
- [20] Elfessi, A. and Pal, N. (1992). Estimation of the smaller and the larger of two uniform scale parameters. *Comm. Stat.-Theor. Meth.* **21**, 2997-3015.
- [21] Hwang, J. T. (1985). Universal domination and stochastic domination: estimation simultaneously under a broad class of loss functions. *Ann. Statist.* **13**, 295-314.
- [22] Hwang, J. T. and Peddada, S. D. (1994). Confidence interval estimation subject to order restrictions. *Ann. Statist.* **22** 67-93.
- [23] Hochberg, Y. and Tamhane, A. C. (1987). *Multiple Comparison Proce-*

dures. Wiley, New York.

- [24] Katz, M. W. (1961). Admissible and minimax estimates of parameters in truncated spaces. *Ann. Math. Statist.* **32**,136-142.
- [25] Katz, M. W. (1963). Estimating ordered probabilities. *Ann. Math. Statist.* **34**, 967-972.
- [26] Kelly, R.E. (1989). Stochastic reduction of loss in estimating normal means by isotonic regression. *Ann. Statist.* **17** 937-940.
- [27] Kumar, S. and Sharma, D. (1988). Simultaneous estimation of ordered parameters. *Comm. Stat.-Theor. Meth.* **17**, 4315-4336.
- [28] Kushary, D. and Cohen, A. (1989). Estimating ordered location and scale parameters. *Stat. Dec.* **7**, 201-213.
- [29] Kushary, D. and Cohen, A. (1991). Estimation of ordered Poisson parameters. *Sankhyā, Series A* **53**, 334-356.
- [30] Lee, C. C. (1981). The quadratic loss of isotonic regression under normality. *Ann. Statist.* **9**, 686-688.
- [31] Lee, C.C. (1988). The quadratic loss of order restricted estimators for treatment means with a control. *Ann. Statist.* **16** 751-758.
- [32] Lehmann, E.L. (1983). *Theory of Point Estimation*. Wiley, New York.
- [33] Menéndez, J. A and Salvador, B. (1991). Anomalies of the likelihood ratio test for testing restricted hypothesis. *Ann. Statist.* **19**, 889-898.
- [34] Menéndez, J. A and Salvador, B. (1992). Equivalence of likelihood ratio tests and obliquity. *Stat. Prob. Let.* **14**, 223-229.
- [35] Menéndez, J. A., Rueda, C. and Salvador, B. (1991). Conditional test for testing a face of the tree order cone. *Comm. Statist.-Sim. Comp.* **20**, 751-762.
- [36] Menéndez, J. A., Rueda, C. and Salvador, B. (1992a). Testing non-oblique hypotheses. *Comm. Stat.-Theor. Meth.* **21**, 471-484.
- [37] Menéndez, J. A., Rueda, C. and Salvador, B. (1992b). Dominance of like-

- likelihood ratio tests under order constraints. *Ann. Statist.* **20**, 2087-2099.
- [38] Misra, N. and Dhariyal, I. D. (1995). Some inadmissibility results for estimating ordered uniform scale parameters. *Comm. Stat.-Theor. Meth.* **24**, 675-685.
- [39] Moors, J. J. A. (1985). *Estimation in Truncated Parameter Spaces*. Ph.D. thesis. Tilburg University.
- [40] Mukerjee, H., Robertson, T. and Wright, F. T. (1986). A probability inequality for elliptically contoured densities with applications in order restricted inference. *Ann. Statist.* **14**, 1544-1554.
- [41] Pal, N. and Kushary, D. (1992). On order restricted location parameters of two exponential distributions. *Stat. Dec.* **10**, 133-152.
- [42] Perlman, M. D. (1969). One-sided problems in multivariate analysis. *Ann. Math. Statist.* **40**, 549-567. (Correcciones en *Ann. Math. Statist.* **42**, 1777).
- [43] Rao, C. R. and Shinozaki, N. (1978). Precision of individual estimators in simultaneous estimation of parameters. *Biom.* **65**, 23-30.
- [44] Raubertas, R. F., Nordheim, E.V. and Lee, C. C. (1986). Hypothesis test for normal means constrained by linear inequalities. *Commun, Statist.-Theor. Meth.* **15** 2809-2833.
- [45] Robertson, T., Wright, F. T., and Dykstra, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, New York.
- [46] Rueda, C. (1989). *Contrates de Hipótesis con Restricciones bajo Condiciones de Oblicuidad*. Tesis Doctoral. Universidad de Valladolid.
- [47] Rueda, C. and Salvador, B. (1995). Reduction of risk using restricted estimators. *Comm. Stat.-Theor. Meth.* **24**, 1011-1023.
- [48] Scheffé, H. (1959). *The Analysis of Variance*. Wiley, New York.
- [49] Seber, G. A. F. (1977). *Linear Regression Analysis*. Wiley, New York.
- [50] Sengupta, D. and Sen, P. K. (1991). Shrinkage estimation in a restricted parameter space. *Sankhyā, Series A* **53**, 389-411.

- [51] Shaked, M. (1979). Estimation of starshaped sequences of Poisson and normal means. *Ann. Statist.* **7**, 729-741.
- [52] Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistical Review* **56**, **1**, 49-62.
- [53] Singh, H., Gupta, R. D. and Misra N. (1993). Estimation of parameters of an exponential distribution when the parameter space is restricted with an application to two-sample problem. *Comm. Stat.-Theor. Meth.* **22**, 461-477.
- [54] Tong, Y. L. (1990). *The Multivariate Normal Distribution*. Springer-Verlag, New York.
- [55] Vijayasree, G. and Singh, H. (1993). Mixed estimators of two ordered exponential means. *J. Stat. Plan. Inf.* **35**, 47-53.