

Proy. T-13/156

P.N

14  
67

UNIVERSIDAD DE GRANADA  
FACULTAD DE CIENCIAS



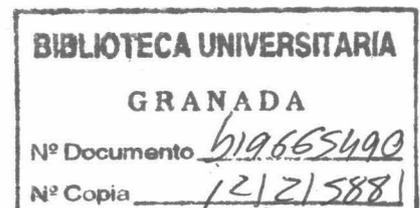
Departamento de Estadística e Investigación Operativa

# INFERENCIA EN MDS Y SU TRATAMIENTO COMPUTACIONAL

TESIS DOCTORAL

José Fernando Vera Vera.

GRANADA, 10 de Mayo de 1993



UNIVERSIDAD DE GRANADA  
FACULTAD DE CIENCIAS

INFERENCIA EN MDS Y SU  
TRATAMIENTO COMPUTACIONAL

JOSÉ FERNANDO VERA VERA

TESIS DOCTORAL

Sección de Matemáticas

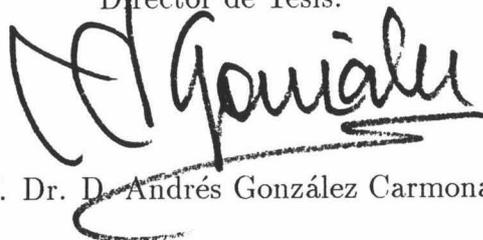
INFERENCIA EN MDS Y SU  
TRATAMIENTO COMPUTACIONAL

Memoria presentada para  
optar al grado de Doctor en  
Ciencias (Matemáticas) por  
José Fernando Vera Vera.



Vº Bº

Director de Tesis:



Prof. Dr. D. Andrés González Carmona.

DEPARTAMENTO DE ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

FACULTAD DE CIENCIAS

UNIVERSIDAD DE GRANADA

Deseo manifestar mi agradecimiento a mis compañeros y amigos del Departamento de Estadística e Investigación Operativa por el apoyo y ayuda recibidos. De manera muy especial quiero expresar mi más sincero agradecimiento, admiración y respeto al director de esta memoria, Catedrático de Universidad, Prof. Dr. D. Andrés González Carmona, a quien debo mi interés y dedicación por las matemáticas y sin cuya ayuda intelectual y humana nunca hubiese realizado esta memoria.

# Indice

<b>1</b>	<b>Introducción al problema del MDS</b>	<b>3</b>
1.1	Nociones generales de MDS. . . . .	5
1.2	Distancias y disimilaridades. . . . .	9
1.2.1	Medidas de disimilaridad. . . . .	11
1.2.2	Clasificación de los datos segun Combs. . . . .	13
1.3	Una solución gráfica al problema del MDS. . . . .	16
<b>2</b>	<b>Modelos de MDS y Medidas de ajuste.</b>	<b>23</b>
2.1	Datos, distancias y modelos bilineales. . . . .	23
2.2	El procedimiento métrico clásico. . . . .	27
2.3	Modelos y medidas de ajuste. . . . .	29
2.3.1	Tipos de funciones $f$ y modelos asociados. . . . .	29
2.3.2	Errores y medidas de ajuste. . . . .	32
2.3.3	Minimización del STRESS y de la ALINEACION. . . . .	35
2.3.4	Problemas que plantea el método del Gradiente. . . . .	42
2.4	Diferencias idividuales en MDS. . . . .	43
2.4.1	El método ALSCAL. . . . .	45
2.4.2	El método INDSCAL. . . . .	68
2.5	La interpretación final en MDS. . . . .	74
2.5.1	La Configuración. . . . .	75
2.5.2	La dimensión del problema. . . . .	78
2.5.3	Algunas consideraciones complementarias. . . . .	80
<b>3</b>	<b>Inferencia en MDS.</b>	<b>83</b>
3.1	Desarrollo estadístico del MDS. . . . .	85
3.2	MDS sin diferencias individuales. . . . .	87
3.2.1	Modelización de la variabilidad residual. . . . .	87

3.2.2	Estimación en el modelo Euclídeo homocedástico biparamétrico. . . . .	94
3.2.3	Estimación en el modelo Euclídeo homocedástico triparamétrico. . . . .	99
3.2.4	Estimación en el modelo Euclídeo heteroscedástico triparamétrico. . . . .	112
3.3	MDS con diferencias individuales. . . . .	128
3.3.1	Métricas individualizadas. . . . .	129
3.3.2	Modelización de la variabilidad residual. . . . .	133
3.3.3	Estimación de las componentes de la varianza. . . . .	138
3.3.4	Estimación de la Configuración. . . . .	149
3.3.5	Estudio de la resolución numérica de las ecuaciones de verosimilitud. . . . .	159
3.3.6	Desarrollo computacional del modelo. . . . .	165
3.3.7	Comparación de modelos y contraste de hipótesis. . . . .	170

# Capítulo 1

## Introducción al problema del MDS

El análisis exploratorio multidimensional de estructuras mediante proximidades o "multidimensional scaling" (MDS), es un método que en su origen fue concebido como un procedimiento muy general de representación de datos para la construcción de una configuración de puntos conocida una determinada información acerca de las proximidades entre los elementos del estudio. Hoy día, este método está constituido por un conjunto de procedimientos y técnicas que tratan de representar coeficientes de proximidad entre conjuntos de objetos mediante distancias en un entorno multidimensional. Así por ejemplo, dada una matriz de correlación entre diversos elementos, el MDS representa dichos elementos como puntos en un plano euclídeo de forma que dos puntos se encontrarán tan próximos entre sí como estén de correlados los elementos a los que representan. Si esta relación entre correlaciones y distancias es lo suficientemente precisa, conseguiremos una representación que pondrá de manifiesto las relaciones intrínsecas existentes, hecho que de otro modo permanecería oculto al investigador ya que siempre es más fácil observar una gráfica que una tabla de coeficientes de correlación.

En una visión histórica del MDS, aunque las primeras ideas surgen de algunos trabajos publicados en los años 20 y 30, puede situarse el origen de esta técnica en el trabajo publicado por Torgerson en el año 1952 titulado *Multidimensional Scaling: I Theory and Method, Psychometrika, 17, 401-419* en donde se introduce el término MDS y se apuntan las primeras ideas de éste. No obstante, el MDS sería formulado de forma más precisa por Shepard en

el año 1962 en el trabajo *MDS With an Unknown Distance Function. I. II. Psychometrika*, 27, 125-140 & 219-246. En este trabajo, Shepard demostró de forma empírica que conocida una ordenación entre las distancias  $d(i, j)$ ,

$$d(i_1, i_2) \leq d(i_3, i_4) \leq \dots \leq d(i_{p-1}, i_p)$$

es posible encontrar una configuración de puntos en un espacio euclídeo, de baja dimensión, de forma que las interdistancias euclídeas reproduzcan de forma bastante exacta la ordenación original.

Esta idea de Shepard fue desarrollada posteriormente por Kruskal, Guttman, Lingoes, Young, Benzecri, Carroll y otros. Dichos métodos reciben el nombre de MDS no métrico ya que centran el estudio en una representación que conserve la ordenación presentada por los datos originales sin que constituya un aspecto primordial el que además exista una interrelación exacta entre proximidad y distancia tal y como pretende el MDS métrico.

Paralelamente a estos desarrollos podemos decir que el primer procedimiento de MDS métrico fue dado por Richardson en 1938, *Multidimensional Psychophysics, Psychological Bulletin* 35, 659-660, en el que apuntó el modelo de *distancia = constante  $\times$  disimilaridad*. Este modelo está basado en un teorema de Young y Householder de 1938, *Discussion of a set of Points in Terms of their Mutual Distances, Psychometrika* 3, 19-22, que ofrece un método de construcción de configuración para distancias euclídeas dadas entre puntos. Torgerson extendió el método al caso de *distancia = constante + constante  $\times$  disimilaridad* mediante un procedimiento constituido por dos etapas: primero, una constante aditiva es añadida y determinada; segundo, transforma la construcción de Young-Householder.

En resumen, podríamos considerar los trabajos anteriores como los orígenes formales del desarrollo de las dos vertientes del MDS, tanto la métrica como la no métrica, cada una de las cuales trata diferentes tipos de problemas de la representación multidimensional relativos a la forma de obtención de las proximidades de partida. Numerosos autores han contribuido a su desarrollo actual tal y como citamos en la bibliografía.

El aspecto más innovador que se plantea hoy día en la investigación del MDS circula en torno a la introducción de inferencia estadística como herramienta fundamental de las técnicas confirmatorias en MDS. Es en este aspecto donde se desarrolla este trabajo tal y como después veremos, centrándonos en algunos aspectos de la inferencia en MDS métrico.

## 1.1 Nociones generales de MDS.

En este apartado vamos a comentar de forma general cual es la importancia del MDS dentro de las técnicas del análisis de datos además de exponer brevemente su filosofía de trabajo bajo un punto de vista clásico, sin hacer referencia al aspecto al que nos conduce este trabajo de la inferencia estadística, el cual será extensamente expuesto en capítulos posteriores.

Es conocido que uno de los problemas más frecuentes en muchas disciplinas se plantea cuando necesitamos medir y entender las relaciones entre objetos siendo desconocidas las dimensiones subyacentes de los mismos. Así MDS se presenta como una herramienta para la representación espacial de objetos en base a proximidades como si éstas fuesen distancias.

El MDS es una técnica que tiene la gran ventaja de no verse afectada por la naturaleza del experimento. No requiere un conocimiento a priori de los atributos mediante los cuales han sido clasificados los estímulos sino que parte únicamente de una tabla de proximidades sin que necesite conocer la procedencia de las mismas. Otro problema es la obtención de dicha tabla pero ése se considera un problema completamente al margen del desarrollo de la técnica.

En general, cuando se conocen las distancias reales entre los diferentes sujetos, su representación euclídea no constituye ningún problema. Sin embargo, cuando por la naturaleza de las relaciones entre dichos objetos éstas no pueden ser tratadas como distancias sino más bien como parecidos subjetivos de quien los observa, es cuando entra en juego la potencia del MDS para conseguir la representación de los mismos, de forma que partiendo simplemente de esa información de proximidad, sea posible representar e interpretar las relaciones en la estructura interna de dichos objetos.

El estudio de problemas del tipo antes descrito, permite deducir una serie de conclusiones respecto al uso del MDS.

En primer lugar, para un único sujeto,  $n$  estímulos pueden ser fijados en  $n - 1$  dimensiones de forma que las distancias entre las representaciones espaciales de éstos se correspondan con las similitudes obtenidas a partir de ellos. No obstante, aunque mediante una dimensión alta se obtiene una buena representación de los datos, no quiere esto decir que una solución en una dimensión elevada sea la más adecuada.

En segundo lugar, con un pequeño arreglo, un número de estímulos pueden forzarse desde una solución en  $n - 1$  dimensiones a una solución de dos o tres

dimensiones con un incremento del error mínimo, lo cual permite un estudio mucho más sencillo de las relaciones internas entre dichos estímulos. Además ofrece la ventaja de poder elegir el investigador el número de dimensiones en que quiere realizar la representación.

Finalmente hay que decir que si el criterio de cálculo relaciona las filas ordenadas de distancias con las de proximidades, los estímulos pueden representarse en  $n - 2$  ó menos dimensiones que cuando se usan las medidas entre ellos.

Para ofrecer una visión de trabajo de MDS hay que empezar haciendo hincapié en cómo se obtienen los datos de partida de un estudio, ya que en muchos casos ese detalle marca el tipo de análisis que puede realizarse. Hay que decir pues que los datos de partida de un estudio se toman en base a cuatro tipos de medidas:

- -NOMINAL.- Los objetos se clasifican sólomente en grupos, como masculino y femenino, por citar un ejemplo. Es el nivel menor y está referido como de tipo categórico.
- -ORDINAL.- Los objetos se organizan según un orden de magnitud. Por ejemplo, autobús-camioneta-coche sería el rango de esos vehículos en términos de tamaño. Al nivel normal de medidas no existe un indicador sobre si la diferencia de tamaño entre autobús y camioneta es mayor o menor que entre camioneta y coche.
- -INTERVALO.- Los objetos se sitúan en una escala de forma que el tamaño de las diferencias entre éstos, queda reflejado por la escala. De todas formas, hay que tener en cuenta que los intervalos escalares no tienen ceros reales. Así, la temperatura en grados Fahrenheit es una escala. La diferencia entre 20 grados y 50 grados es la misma que entre 50 grados y 80 grados. No obstante no es cierto el afirmar que 80 grados sea 4 veces más caliente que 20 grados, ya que no disponemos de un cero absoluto en cuanto a la escala, por lo que eso no será cierto en general a menos que se situara el origen de ésta en 0 grados absolutos como punto de referencia.
- -RAZON O PROPORCION.- Los objetos se sitúan en una escala de forma que su posición en ésta representa el valor exacto del atributo que se mide. Es el nivel de medida más riguroso. Así por ejemplo, masa y

velocidad son claros ejemplos representativos de esta idea. Un coche a 60 Km/h viaja 2 veces más rápido que otro a 30 Km/h.

Una vez obtenida la tabla de proximidades, se deduce un espacio en el que las distancias entre los puntos son similares a las proximidades de que se partía. Esto es lo que se conoce como un análisis ordinario, el cual se realiza en varias etapas.

En primer lugar se obtiene una configuración inicial sobre una recta. Se calcula el valor de la bondad del ajuste y se estudia si este puede considerarse adecuado o no. De no serlo, se realizaría una nueva iteración, obteniendo una nueva representación en un plano y así sucesivamente hasta encontrar un ajuste aceptable para lo que queremos estudiar.

En líneas generales, la forma de representar en cada dimensión, así como el método iterativo empleado y sobre todo las diferentes medidas de la bondad del ajuste, son los principales factores que determinan la existencia de los diferentes métodos de MDS que se presentan en la actualidad. Estos métodos aparecen hoy en día con una implementación computerizada ya que sin el uso del ordenador, el desarrollo del MDS no habría avanzado debido a que la mayoría de los cálculos necesitan de un procedimiento iterativo de optimización, como después veremos.

Hemos descrito en líneas generales como actúa MDS. A un nivel más particular, MDS parte normalmente de las diferencias entre cada par de conjuntos de estímulos. Estos datos suelen ser disimilaridades o disimilitudes o proximidades o incluso simplemente datos. Lo que queremos es representar esos estímulos en un espacio de coordenadas de forma que las distancias entre los puntos representados, sean lo más parecidas a las proximidades de que partíamos.

La posición de cada estímulo será dada mediante sus coordenadas en cada una de las dimensiones consideradas. MDS calcula unas coordenadas iniciales y compara las diferencias entre las proximidades y las distancias de forma que si no son lo suficientemente parecidas, realiza un nuevo cálculo moviendo las coordenadas y comparándolas nuevamente con los datos, hasta que las diferencias sean satisfactorias.

Otro aspecto importante a destacar es el sentido que recoge en el método el concepto de dimensión. Cuando nos referimos a dimensiones lo hacemos como referencia al número de ejes coordenados del espacio de estímulos. Para determinarlas, lo primero en que debemos fijarnos es en como varía la bondad

del ajuste al aumentar la dimensionalidad.

Existen diferentes formas de calcular la bondad del ajuste, tal y como antes apuntábamos. Los casos más utilizados en la práctica son los coeficientes de alineación, el STRESS, correlaciones, varianzas y desviaciones típicas. Por ejemplo, como relaciones entre estos índices de ajuste podemos decir que en general, el STRESS decrece y la correlación aumenta al aumentar la dimensionalidad de la representación MDS.

Otra línea de actuación en cuanto a la tasación de la bondad del ajuste en MDS es la que ofrecen diferentes métodos de simulación computerizada. Así pues, como claro exponente en este sentido hay que destacar el empleo del método de Monte Carlo para la determinación de la dimensionalidad. La idea fundamental de la aplicación de este método estriba en la sustitución del espacio de estímulos por un conjunto de números aleatorios. Varios investigadores han determinado los valores del STRESS obtenido para diferentes números de estímulos mediante conjuntos de datos derivados de tablas de números aleatorios. Se han creado pautas para la determinación del número máximo de dimensiones que resulta apropiado considerar para cada número de estímulos. No obstante el inconveniente de esta técnica es que no siempre es posible disponer del número necesario de estímulos para poder utilizar dichas pautas.

Existen diferentes formas de superar el problema anterior. Una que se suele utilizar frecuentemente consiste en repetir el experimento y comprobar si los nuevos datos ofrecen la misma solución que los antiguos. Otra forma consiste en subdividir en grupos el conjunto de datos y comprobar si coinciden las soluciones obtenidas parcialmente para cada conjunto. Un tercer método es el incluir unas pocas réplicas de estímulos y analizarlas como si fuesen distintos estímulos del problema. Lo próximos que estén estos estímulos replicados de ellos mismos en el espacio del cual provienen, nos da una medida del ruido en los datos y un indicador de la dimensionalidad del análisis. Por último, otro criterio de dimensionalidad es el grado de interpretabilidad de los resultados obtenidos en cada dimensión. Las dimensiones que no puedan interpretarse probablemente no existirán.

Otro aspecto a tener en cuenta es que cuando MDS ofrece los resultados del análisis (coordenadas de los puntos), éstos no se agrupan en torno a los ejes de referencia, o sea, los ejes de coordenadas no son representativos en cuanto a la interpretación de la configuración. Cualquier rotación de los puntos sería otra solución posible ya que conserva las distancias. No obstante, es posible representar a su vez ejes que si sean representativos para la interpretación de

relaciones entre la configuración.

Finalmente, vamos a poner de manifiesto las características que hacen al MDS diferente del Análisis Factorial (AF) como técnica más aproximada en el campo multivariante. Para ello, hay que decir que así como MDS se refiere a una colección de procedimientos de análisis de datos (no a un único procedimiento), el AF refiere a una familia de procedimientos utilizados para analizar un conjunto de datos multivariantes, así como varias clasificaciones de atributos para un número de estímulos.

El MDS está basado en distancias entre puntos, mientras que el AF está basado en los ángulos entre vectores. Ambos modelos utilizan generalmente el espacio euclídeo, aunque el MDS tiene la ventaja de que siempre es más fácil interpretar distancias entre puntos que ángulos entre vectores. Además, el AF resuelve el problema en un número de dimensiones relativamente grande, principalmente porque la mayoría de los procedimientos están basados en la suposición de relaciones lineales entre variables. Esta es una suposición muy severa con respecto a los datos. Las aproximaciones en MDS no contienen esa suposición y el resultado es que este normalmente proporciona soluciones más fáciles de interpretar en relación a la realidad del problema y en una dimensión más baja.

Aparte de esas consideraciones, existe una razón fundamental para preferir MDS al AF y es referente al tipo de datos analizados por cada procedimiento. Como antes apuntábamos, los datos para MDS, cuando están recogidos directamente de juicios ó disimilaridades, se encuentran menos expuestos a la contaminación experimental y más dispuestos a contener estructuras relevantes. Los datos del AF generalmente contienen puntuaciones para cada estímulo en una lista de atributos que pueden o no ser relevantes.

## 1.2 Distancias y disimilaridades.

MDS está basado en la relación entre el concepto de *disimilaridad* y el concepto de *distancia*. Existen ciertos aspectos en común con respecto a estas dos ideas. *Disimilaridad* es la desigualdad perceptible entre dos objetos bajo el punto de vista de un perceptor. El término *diferencia* es normalmente sinónimo pero debido a que también puede indicar algo similar a diferencias aritméticas entre dos números, es preferible el término *disimilaridad* aunque sea más largo, pues es menos ambiguo.

Por otro lado, tenemos el concepto de *similaridad* el cual produce una relación inversa a la de la *disimilaridad*. Las relaciones entre *disimilaridad* y *similaridad* o *proximidad* las veremos formalmente a continuación.

Consideremos un conjunto formado por un número determinado de objetos de los cuales  $i$  y  $j$  son dos exponentes. Llamamos *disimilaridad* entre  $i$  y  $j$  a una función que indicaremos por  $d(i, j)$  o  $d_{ij}$  que muestra el grado de semejanza entre ambos objetos en relación a un cierto número de características cualitativas y cuantitativas. Se llama *similaridad* o similitud entre  $i$  y  $j$  a la cantidad  $s_{ij}$  dada por la relación

$$s_{ij} = 1 - d_{ij}$$

Las propiedades que caracterizan una disimilaridad son:

1.  $d(i, j) \geq 0$
2.  $d(i, i) = 0 \quad \forall i, j$
3.  $d(i, j) = d(j, i)$

y como consecuencia, las propiedades que caracterizan una similaridad serán:

1.  $s(i, j) \leq 1$
2.  $s(i, i) = 1 \quad \forall i, j$
3.  $s(i, j) = s(j, i)$

La estrecha relación entre las características de las disimilaridades y las propiedades matemáticas de las distancias, constituyen la base lógica del MDS. A un nivel muy básico, ambas son propiedades de pares de objetos: las disimilaridades describen un par de estímulos y las distancias un par de puntos. Incluso cuando nos referimos a un único concepto como el de *muy diferentes*, normalmente se tiene presente un estímulo respecto al cual se ha comparado. Además de esto, los dos conceptos comparten las siguientes características:

1. **Origen.** Un juicio de no disimilaridad se aplica a dos estímulos idénticos. De igual forma, una distancia igual a cero se aplica a dos puntos coincidentes.

2. **Positividad.** Cuando dos estímulos no son idénticos en algún aspecto relevante, se supondrá que tienen un grado de disimilaridad positivo. De igual forma, la distancia entre dos puntos no coincidentes es siempre positiva.
3. **Simetría.** Normalmente, el orden en que dos estímulos han sido especificados tiene poco o ningún efecto sobre las disimilaridades percibidas, especialmente, cuando los estímulos tienen el mismo nivel de generalidad.
4. **Consistencia.** Esta propiedad es la más importante para las disimilaridades en el sentido de que tres o más juicios de disimilaridad deben ser consistentes con otro, o sea, si dos estímulos son cada uno de ellos respecto a un tercero, separadamente, muy similares, entonces éstos entre sí no pueden ser muy disimilares. Esta propiedad de las disimilaridades concuerda con la desigualdad triangular de las distancias.

Hay que añadir a lo antes expuesto que no existe una regla general que permita definir una disimilaridad, apta para cualquier tipo de análisis. Depende de varios factores: características de las variables, naturaleza de los objetos, finalidad del análisis, etc ... Es imprescindible que cualquiera que sea la definición, cumpla los tres axiomas antes expuestos así como las cuatro propiedades. A continuación vamos a dar algunos tipos de medidas de disimilaridad así como comentar brevemente los principales métodos de obtención de datos para un análisis con MDS.

### 1.2.1 Medidas de disimilaridad.

Se trata ahora de discutir los principales tipos de medidas de disimilaridad que se obtienen a nivel experimental. Hay que resaltar que, generalmente, es necesario realizar transformaciones previas de las medidas experimentales que suelen ser de similaridad o proximidad para que puedan ser tratadas como medidas de disimilaridad. Veamos los principales casos:

1. **Métodos de Calificación directa.** Esta es una de las formas más utilizadas debido a la facilidad de obtención de disimilaridades de forma directa a través de las opiniones de los sujetos. El método más importante dentro de este grupo puede considerarse el de la utilización de escalas ordenadas categóricamente para la asignación de valores. El

ejemplo más típico es el de la escala que recorre los rangos desde *muy similar*, hasta, *nada similar*. Así, si dividimos ese rango en  $n$  posibilidades podremos clasificar las opiniones según el grado de similaridad. Un factor que deberemos tener en cuenta para este tipo de escalas en lo que respecta a los desarrollos inferenciales posteriores, estriba en el número de posibilidades en que se dividan esas categorías. Para que la variable pueda considerarse como continua, Ramsay (1973) ha demostrado que como mínimo deben asignarse siete valores. Hay igualmente que tener en cuenta que a la hora de la clasificación existe una gran variabilidad en cuanto al número de clases utilizadas por los sujetos en sus asignaciones. Así pues, mientras algunos sólo utilizan cuatro o cinco clases otros utilizan un amplio abanico. Por ello, en estos casos, parece recomendable utilizar técnicas como la que desarrolló Takane (1981) para categorías con menos de seis cortes ya que en esos casos parece más indicado reducir el número de categorías en virtud del tipo de sujeto que clasifica.

2. **Clasificación de disimilaridades.** Este método, aunque resulta relativamente poco usado es de un gran potencial práctico y consiste en clasificar los estímulos en términos de sus disimilaridades con respecto a cada uno o a un determinado número de los restantes estímulos.
3. **Frecuencias de Transición.** Una forma de obtención de medidas de proximidad consiste en encontrar el número o proporción de veces que un estímulo se asocia con una respuesta particular por parte de un sujeto. Suele ser frecuente este método en experimentos en los que intervienen animales. Una vez obtenidos los resultados, es necesario transformarlos en índices de disimilaridad para lo cual se suele emplear el método del inverso de la raíz cuadrada de las frecuencias o bien considerar uno menos la frecuencia de confusión. Finalmente hay que tener en cuenta en este método que suele ser normal la aparición de asimetría en los datos y que cualquier índice estará correlado con otro.
4. **Frecuencias de ocurrencia incontrolada.** Intimamente relacionado con el tipo anterior se encuentra este tipo de frecuencias que pueden definirse como el número de asociaciones, donde el límite superior de dichas asociaciones no es conocido o controlado experimentalmente. Por ejemplo, se podría contar el número de veces que los habitantes de una ciudad se desplazan a otra en un período fijo del año. Nuevamente

aparece asimetría en los datos. Una variante de este método se utiliza cuando a cada sujeto se le pide que ordene el estímulo entre un número de grupos. La frecuencia con que dos estímulos son clasificados en el mismo grupo resulta una medida de similaridad. El número de grupos puede dejarse o no a la elección del sujeto. La ventaja de estos datos es que un gran número de estímulos puede ser juzgado de una sola vez.

### 1.2.2 Clasificación de los datos según Combs.

Merece una especial atención la clasificación de los datos a tratar con MDS que dió Coombs en su libro *A theory of Data*, John Wiley & Sons, Inc, New York. En él expone que las relaciones básicas que caracterizan a los datos de tipo psicológico son de dos tipos: Dominancia y consonancia.

1. **Dominancia.** En este caso se crea un orden de rangos estricto para cada conjunto de elementos, respecto a algún atributo con el que se quieran comparar. Ej.- La marca A es *mayor* que la B respecto a la cualidad X.
2. **Consonancia.** Se establece una comparación de igualdad entre los elementos. Ej.- A es *igual* que B respecto a X. O bien, A *no es igual* a B respecto a X.

Puede resultar sorprendente el hecho de que virtualmente todos los estudios de MDS conllevan implícitamente una de las dos clases anteriores. Además, si permitimos asociaciones en la ordenación de las preferencias se obtienen automáticamente relaciones de dominancia y consonancia dentro del mismo conjunto de datos. El concepto de dominancia u ordenación implica claramente un espacio continuo unidimensional para cualquier par de elementos para el que los términos *menor que* o *mayor que* tengan sentido. La idea de consonancia o igualdad puede construirse como un intervalo de un continuo, dentro del cual el sujeto supone que todos sus puntos son indistinguibles. De hecho, podría parecer que las relaciones de consonancia y dominancia producen representaciones unidimensionales en lugar de multidimensionales. No obstante puede ser muy útil el caracterizar estímulos o sujetos en varias dimensiones de algún espacio geométrico. La ordenación de distancias entre pares de puntos suministra el continuo unidimensional apropiado que conducirá a la construcción multidimensional.

Centremos ahora nuestra atención en los tipos de datos psicológicos que pueden ser tratados. Para ello Coombs consideró tres descriptores básicos:

- Cuando la relación es de dominancia o consonancia.
- Cuando los puntos son de un mismo conjunto o de dos conjuntos diferentes.
- Cuando la relación es entre un par de puntos o entre un par de pares de puntos.

Basándose en estos tres criterios, Coombs desarrolló la siguiente tabla de clasificación de datos.

	<i>Pares de puntos</i>	<i>Pares de pares</i>
<i>Puntos de conjuntos diferentes</i>	<b>Cuadrante II</b> Estímulos simples	<b>Cuadrante I</b> Elección preferencial
<i>Puntos del mismo conjunto</i>	<b>Cuadrante III</b> Comparación de estímulos	<b>Cuadrante IV</b> Datos de Similaridad

Veamos detenidamente cada una de estas clases.

- **Cuadrante I.- Elección preferencial de datos.** Los datos de este cuadrante se caracterizan por existir unas relaciones subyacentes entre pares de pares de puntos en dos conjuntos diferentes. Consideremos el siguiente ejemplo: Supongamos que a un individuo se le pide que realice ciertos juicios de preferencia sobre varios automóviles y que este sujeto percibe únicamente dos características de los coches: el grado de lujo y el grado de deportividad. Coombs propuso el concepto de punto ideal que sería el coche ideal bajo esos puntos de vista con el que deberían compararse las demás clasificaciones. En este ejemplo, los conjuntos son el punto ideal  $I$  y los modelos de coches. Los pares de puntos serán pues  $(marca1, I) \cdots (marcaN, I)$ . Se miden las distancias entre cada marca y el punto ideal  $d(marcaj, I)$ . Se comparan dichas distancias de forma que un modelo será más preferido que otro si su distancia al punto ideal

$I$  es menor. Hay que resaltar que esta clasificación es completamente dependiente del punto ideal  $I$ . Si hubiésemos tomado otro punto  $J$  como punto ideal, por ejemplo un promedio entre lujo y deportividad, las conclusiones serían diferentes. El tipo de representación con MDS que se obtiene para este tipo de datos suele ser multidimensional.

- **Cuadrante II.-** *Estímulos Simples.* Los datos de este cuadrante, son clasificados por Coombs como relaciones entre pares de puntos de dos conjuntos diferentes. Supongamos que tenemos un conjunto de atributos para describir a cada elemento de un conjunto de estímulos (Ej. marcas) y se pregunta a un gran número de personas sobre cada elemento, para que dé valores a todos los atributos respecto a dicho elemento.
- **Cuadrante III.-** *Comparación de estímulos.* Los datos de este cuadrante se caracterizan por las relaciones entre los pares de puntos del mismo conjunto. Este tipo de datos y el anterior suelen ofrecer representaciones en MDS de tipo unidimensional.
- **Cuadrante IV.-** *Datos de Similaridad.* Esta clase de datos envuelve una estructura de apariencia o percepción de como los sujetos perciben los estímulos: si de forma similar o no. Esto origina relaciones entre los pares de puntos del mismo conjunto. Este tipo de datos es el que se obtendría al preguntar a los individuos qué pares de automóviles son más similares. El proceso se repite hasta tener ordenados los datos. Los ejes de la representación mostrarán las dimensiones subyacentes de percepción con las que se han ordenado los pares. Son representaciones de tipo multidimensional.

Otra forma de presentar los tipos de datos utilizados en MDS es describiendo los diferentes tipos de matrices de datos. Coombs enuncia varios tipos de matrices de proximidades, donde por proximidad se refiere a distancia psicológica, tanto en el sentido de similaridad como en el disimilaridad. Veamos dichos tipos.

1. **Matriz de datos del tipo estímulo por estímulo.** Supongamos en primer lugar el caso en que los pares de estímulos son ordenados. Obtendremos una matriz en la que los  $n(n - 1)/2$  distintos pares de estímulos han sido ordenados y se tienen  $n(n - 1)/2$  puntuaciones que indican la

ordenación. Este tipo de proximidad se denomina proximidad incondicional o íntegra. Es íntegra porque cada casilla excepto la diagonal principal, tiene un número que es el grado de similaridad o disimilaridad del par. Es incondicional porque cada casilla puede compararse con otra cualquiera sin ningún problema. Otro tipo de matrices lo constituye el caso de condicional íntegra, el cual aparece cuando para cada fila de la matriz, el evaluador ordena los restantes  $n - 1$  estímulos en términos de similaridad decreciente, respecto de esa fila-estímulo.

2. **Matriz de datos del tipo individuos por estímulo.** Supongamos que  $N$  individuos ordenan  $n$  estímulos. En este caso resulta una matriz triangular inferior de tipo condicional en la que los valores solo son comparables en cada fila (cada fila es el resultado de la ordenación de una persona). Los pares de puntos son tomados de dos conjuntos diferentes: el de los individuos y el de los estímulos.

Finalmente hay que decir que las matrices íntegras se asocian a los cuadrantes III y IV mientras que las diagonales inferiores se asocian a los cuadrantes I y II.

### 1.3 Una solución gráfica al problema del MDS.

En este apartado se trata de dar una visión general de la resolución del problema del MDS comenzando con algunas ideas meramente intuitivas hasta lograr algunos desarrollos matemáticos.

Supongamos que se tiene una tabla de similitudes entre cuatro objetos. Para su representación en un plano, elegimos en primer lugar cuatro puntos totalmente aleatorios y medimos sus distancias. A continuación, a partir de la primera tabla, se construye una tabla de rangos, o sea, construimos una tabla de forma que tenga un valor 1 en la posición que ocupe en la tabla original la menor proximidad, un valor 2 la siguiente y así sucesivamente. Considerando la tabla de distancias, se reordena ésta por filas según el orden de rangos obtenido a partir de la tabla original de similitudes. Si calculamos ahora la diferencia entre la tabla de distancias originales y la nueva tabla de distancias ordenadas se obtiene una tabla denominada matriz de corrección de la tabla de distancias originales, ya que dicha tabla de distancias era totalmente aleatoria al serlo los puntos de los cuales dependía. Una vez obtenida la matriz de

corrección, sus valores serán las longitudes de los vectores que darán lugar a la nueva configuración de puntos. Midiendo las distancias entre los nuevos puntos, se obtiene una nueva tabla de distancias y de esa forma, una nueva matriz de corrección.

Todo el proceso descrito, se desarrolla considerando en cada movimiento de los puntos, un punto móvil y los demás fijos. Por esa razón y debido a que los cuatro puntos en principio son móviles, aparece la necesidad de crear un procedimiento iterativo. No existe ninguna garantía de que un determinado conjunto de movimientos nos lleve a una solución adecuada y además hay que decir que el procedimiento descrito compara únicamente las distancias *conjuntas* (aproximación condicional) pero no dice nada sobre las distancias *disjuntas* entre puntos no próximos (aproximación incondicional).

Veamos a continuación el desarrollo analítico que resuelve el problema anterior.

Consideremos una matriz de similaridades y tomemos un conjunto aleatorio de puntos del plano como primera aproximación. Calculemos la matriz de distancias de esa configuración. A partir de dicha matriz de distancias, se obtiene la matriz de distancias ordenadas totalmente puesto que la aproximación que se desea es de tipo incondicional. Considerando ahora ambas matrices, la matriz original de distancias, la cual notaremos por  $d_{ij}$  y la matriz ordenada de distancias que notaremos  $d_{ij}^*$ , la cantidad

$$\lambda_{ij} = d_{ij} - d_{ij}^*$$

dará una doble información: módulo y sentido del vector corrector desde el punto  $i$  al punto  $j$ . La proporción en que la distancia original debe ser corregida es

$$\frac{\lambda_{ij}}{d_{ij}} = \frac{d_{ij} - d_{ij}^*}{d_{ij}} = 1 - \frac{d_{ij} - d_{ij}^*}{d_{ij}} = f_{ij}$$

Se llama factor de corrección entre los puntos  $i$  y  $j$  a la cantidad  $f_{ij}$  si  $i \neq j$  y 0 si  $i = j$ .

Consideremos una configuración en un espacio  $m$ -dimensional. Sean los puntos  $i, j$  y sea  $f_{ij}$  el factor de corrección entre dichos puntos. En términos vectoriales, si sobre el eje  $a$  del sistema de referencia dado, la coordenada de  $i$  es  $x_{ia}$  y la de  $j$ ,  $x_{ja}$  tendremos que la corrección sobre el vector  $\vec{ij}$  será

$$f_{ij}\vec{i}j = \frac{f_{ij}}{(j-i)}$$

por lo que para el eje  $a$  quedará,

$$v_a = f_{ij}(x_{ja} - x_{ia}) = \frac{\left(1 - \frac{d_{ij}^*}{d_{ij}}\right)}{(x_{ja} - x_{ia})}$$

Así pues, el vector de traslación total  $f_{ij}\vec{i}j$  será la suma con respecto a las traslaciones de las componentes

$$f_{ij}\vec{i}j = \sum_{a=1}^m f_{ij}(x_{ja} - x_{ia})$$

Hasta ahora, hemos determinado únicamente el movimiento a lo largo del eje  $a$  del punto  $i$  respecto al punto  $j$ . No obstante, todos los demás puntos deben ser tenidos en cuenta. Para determinar el vector de traslación del punto  $i$  con respecto no al punto  $j$  fijo sino a todos los otros que constituyen la configuración, por ejemplo con respecto nuevamente al eje  $a$ , bastará con sumar las componentes  $v_a$  de cada uno de ellos y se obtendrá la componente de traslación del movimiento del punto  $i$  con respecto al eje  $a$ . Así pues:

$$\sum_{j=1}^n v_a(j) = \sum_{j=1}^n \frac{1 - \frac{d_{ij}^*}{d_{ij}}}{(x_{ja} - x_{ia})}$$

donde

$$\frac{d_{ii}^*}{d_{ii}} = 0 \text{ por definición.}$$

$n$  = Número de puntos de la configuración.

$v_a(j)$  = Vector de traslación de  $i$  a  $j$  en el eje  $a$ .

Luego la nueva coordenada del punto  $i$  respecto al eje  $a$  será:

$$x_{ia}^{(t+1)} = x_{ia}^{(t)} + \left[ \sum_{j=1}^n \left(1 - \frac{d_{ij}^*}{d_{ij}}\right) (x_{ja} - x_{ia}) \right]^{(t)}$$

Esta forma de traslación del punto  $i$  con respecto a los puntos de la configuración plantea problemas en la práctica ya que si, por ejemplo,  $v_a(j) = 2$ ,  $\forall j$  y  $n = 50$ , entonces  $x_{ia}^{(t+1)} = x_{ia}^{(t)} + (49 \times 2)$  y esa traslación es excesivamente grande. Resulta pues más sensato, sumar un promedio de las  $v_j^a$  y así:

$$x_{ia}^{(t+1)} = x_{ia}^{(t)} + \left[ \frac{1}{n} \sum_{j=1}^n \left( 1 - \frac{d_{ij}^*}{d_{ij}} \right) (x_{ja} - x_{ia}) \right]^{(t)}$$

Cuando realizamos la traza del movimiento de los puntos se observa que cuanto mayor es el número de iteraciones, los movimientos de los puntos son cada vez más cortos. Además se observa que las direcciones de movimiento son variables así como la longitud de las mismas. Por tanto parece razonable disponer de algún coeficiente que indique como cambian con cada iteración las relaciones entre las distancias de los puntos. Un posible candidato podría ser el coeficiente de correlación por rangos pero en los casos en que el número de puntos es pequeño no es aplicable. Por ello utilizamos una versión modificada del mismo. Sabemos que los factores de corrección  $\lambda_{ij}$  y  $f_{ij}$  valen cero si  $d_{ij}$  es igual a  $d_{ij}^*$ . Así pues, lo que se trata de hacer mediante el movimiento de los puntos es minimizar las diferencias  $(d_{ij} - d_{ij}^*)$ . Podríamos pues valorar la proximidad entre  $d_{ij}$  y  $d_{ij}^*$  calculando el coeficiente de correlación de Pearson  $r$  sobre dichos pares. Si la solución es perfecta se tiene que  $r = 1$ .

Si todo marcha correctamente, al representar el diagrama de Shepard para las iteraciones debe ocurrir que los puntos se vayan alineando con respecto a la diagonal conforme el número de iteración crece. De igual forma cabe esperar que el coeficiente de correlación  $r$  se aproxime más a 1 cuanto mayor sea dicho número. Si se dibuja una gráfica del valor de  $r$  en cada iteración, generalmente se observa que la correlación varía fuertemente en las primeras iteraciones, tendiendo a estabilizarse posteriormente.

Con lo dicho se obtienen dos claras conclusiones: Por un lado, el desarrollo del método depende de la configuración inicial elegida por lo que un aspecto a desarrollar es la forma de obtención de dicha configuración inicial. Por otro lado cabe esperar una mejora en el ajuste de una configuración dada que requiera el mínimo número de iteraciones y además asegure en cierto modo la obtención de una conclusión aceptable.

Tal y como se ha visto hasta ahora, la utilización de imágenes ordenadas como distancias objetivo puede ser una de las muchas posibilidades para resolver el problema. Otra posibilidad que consideraremos a continuación es

derivar dichas distancias objetivo vía regresión monótona.

En principio, como hacíamos antes, consideramos una tabla de similitudes y tomando una configuración de puntos totalmente arbitraria, medimos las distancias entre dichos puntos. Para ver como se ajusta dicha representación a los datos, calculamos las distancias y se comparan con los datos originales. Una forma de hacerlo, es representando los pares (*distancia, proximidad*), obteniéndose un diagrama de Shepard asociado a esa primera configuración. Si la representación fuese perfecta, los puntos deberían situarse en torno a una curva que debería ascender fuertemente y además inclinarse a la derecha. Como ese no es el caso, normalmente, vamos a intentar derivar a partir de esa idea una forma para la construcción de las distancias objetivo. Parece evidente que todos los puntos que se encuentren en cualquier función estrictamente creciente en el diagrama de Shepard serían candidatos a distancia objetivo. No obstante, existen infinitas curvas de ese tipo e infinitas formas de definir las distancias objetivo. De cualquier forma, parece razonable definir las distancias objetivo de forma que se eviten movimientos innecesarios en los puntos. Esto conlleva el hecho de que la curva sea dibujada a través de la representación de los puntos de forma que los puntos estén tan próximos a ella como sea posible. La proximidad de los puntos a la curva de distancias objetivo se mide horizontalmente ya que estamos interesados en valores objetivo que requieran una distancia de ajuste mínima.

Así pues, el problema que se nos plantea no es más que un problema clásico de regresión con la única variante de que la curva de ajuste no es lineal sino en general, estrictamente creciente. Para resolverlo empleamos el método de mínimos cuadrados. Así pues, si la distancia objetivo se nota por  $\hat{d}_{ij}$  el error de aproximación cometido será

$$e_{ij} = (d_{ij} - \hat{d}_{ij})$$

Minimizando la suma de los errores al cuadrado, encontramos la mejor secuencia estrictamente monótona de  $\hat{d}_{ij}$ .

Hay que resaltar el hecho de que en realidad, las *distancias objetivo* no son realmente distancias en el sentido estricto de la palabra. En particular, dada una configuración, sería imposible decir cuales son las distancias objetivo. Éstas realmente están constituidas por números calculados de alguna determinada forma que sirven como valores objetivo. Así en el ambiente de la regresión monótona dichas distancias objetivo son conocidas como disparidades.

El empleo de disparidades en lugar de imágenes ordenadas (como hacíamos anteriormente), no afecta al desarrollo de las fórmulas antes expuestas y basta con sustituir  $d_{ij}^*$  por  $\hat{d}_{ij}$ .



## Capítulo 2

# Modelos de MDS y Medidas de ajuste.

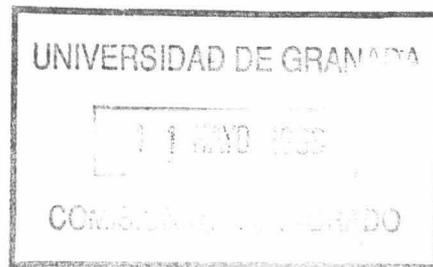
En el capítulo anterior se han desarrollado a nivel intuitivo las ideas principales que constituyen la técnica del MDS. En este capítulo trataremos de formalizar matemáticamente dichas ideas, así como dar una visión del desarrollo del MDS en el sentido del análisis exploratorio de datos, es decir, sin hacer aún uso de la inferencia estadística como herramienta fundamental. De este último aspecto nos ocuparemos en el siguiente capítulo el cual irá destinado a la inferencia en MDS.

Hagamos un planteamiento general de la situación en que nos encontramos.

### 2.1 Datos, distancias y modelos bilineales.

En primer lugar, consideremos los datos,  $d_{ijrt}$  como las respuestas de cada sujeto  $r$  ( $r = 1, \dots, R$ ) a un par de objetos, conceptos o estímulos cualesquiera  $i, j$  ( $i, j = 1, \dots, n$ ) en el ensayo o réplica  $t$  ( $t = 1, \dots, T$ ). Tal y como apuntábamos en el capítulo anterior, un dato debe llevar implícito en el proceso de su obtención el hecho de reflejar dos características esenciales: simetría con respecto a los estímulos y consistencia, de tal forma que pueda afirmarse que los datos, o alguna transformación monótona de los mismos, inducen una métrica en el conjunto de los estímulos. Dicha métrica es la que se intenta obtener.

El parecido entre el principio de consistencia y la desigualdad triangular



induce a representar las  $d_{ijrt}$  mediante el modelo de distancia

$$d_{ijr}^* = (X_i - X_j)W_r(X_i - X_j)^t$$

donde  $X_i = (x_{i1}, \dots, x_{iM})$ , son las coordenadas del punto  $i$  en dimensión  $M$  y  $W_r$  es una matriz simétrica definida positiva que varía de sujeto a sujeto. Esta matriz suele considerarse de tres tipos: *métrica identidad*, *métrica diagonal* o *métrica completa*, según que sea  $W_r = I$ , que sea  $W_r$  una matriz diagonal o bien en el caso más general, simplemente que  $W_r$  sea una matriz simétrica semidefinida positiva cualquiera.

Un modelo alternativo lo constituye la relación

$$S_{ijr}^* = X_i W_r X_j^t$$

que se conoce como *modelo bilineal*. Ambos modelos están relacionados mediante la transformación doblemente centrada:

$$S_{ijr}^* = \frac{1}{2} (d_{i.r}^{*2} + d_{.jr}^{*2} - d_{.r}^{*2} - d_{ijr}^{*2})$$

donde los promedios se toman respecto a distancias al cuadrado. Veamos a continuación la relación entre distancias estimadas y productos escalares.

Sean  $X_i = (x_{i1}, \dots, x_{iM})$  y  $X_j = (x_{j1}, \dots, x_{jM})$  dos vectores en un espacio de dimensión  $M$ . Su producto escalar será

$$S_{ij}^* = X_i W X_j^t = \sum_{k=1}^M \sum_{t=1}^M x_{it} w_{tk} x_{jk}$$

En la mayoría de las aplicaciones, el origen de coordenadas no es algo determinante y puede fijarse arbitrariamente, debido a que un movimiento rígido no altera la configuración final. Así pues, resulta conveniente desplazar el origen de coordenadas al *centroide* o media generalizada de la configuración. No obstante, mientras que las distancias no se alteran por una traslación, los productos escalares sí que lo hacen. Veamos pues como se resuelve este problema.

Supongamos que el origen es el centroide y notemos por  $S_{ij}^*$  el producto escalar.

**Proposición 2.1.1** *Las distancias Euclídeas pueden convertirse en productos escalares de vectores sobre un origen (arbitrario) en el centroide de todos los puntos mediante la ecuación:*

$$S_{ij}^* = \frac{1}{2} (d_{i.}^{*2} + d_{.j}^{*2} - d_{..}^{*2} - d_{ij}^{*2})$$

donde

$$1. d_{i.}^{*2} = \frac{1}{n} \sum_{k=1}^n d_{ik}^{*2}$$

$$2. d_{.j}^{*2} = \frac{1}{n} \sum_{k=1}^n d_{kj}^{*2}$$

$$3. d_{..}^{*2} = \frac{1}{n} \sum_{t=1}^n \sum_{k=1}^n d_{tk}^{*2}$$

DEMOSTRACIÓN.

Tenemos que

$$\begin{aligned} d_{ij}^{*2} &= \sum_{k=1}^M \sum_{t=1}^M (x_{it} - x_{jt}) w_{tk} (x_{ik} - x_{jk}) = \\ &= \sum_{k=1}^M \sum_{t=1}^M w_{tk} (x_{it} x_{ik} + x_{jt} x_{jk} - x_{it} x_{jk} - x_{jt} x_{ik}) = \\ &= \sum_{k=1}^M \sum_{t=1}^M w_{tk} x_{it} x_{ik} + \sum_{k=1}^M \sum_{t=1}^M w_{tk} x_{jt} x_{jk} - \\ &= \sum_{k=1}^M \sum_{t=1}^M w_{tk} x_{it} x_{jk} - \sum_{k=1}^M \sum_{t=1}^M w_{tk} x_{jt} x_{ik} = \\ &= S_{ii}^* + S_{jj}^* - S_{ij}^* - S_{ij}^* = \\ &= S_{ii}^* + S_{jj}^* - 2S_{ij}^* \end{aligned}$$

Puesto que el origen de coordenadas se ha situado en el centroide, se tiene que

$$\sum_{l=1}^n x_{lh} = 0 \quad \forall h = 1, \dots, M$$

por lo que se verifica que:

$$S_{.j}^* = \frac{1}{n} \sum_{l=1}^n \left( \sum_{k=1}^M \sum_{t=1}^M x_{lt} w_{tk} x_{jk} \right) =$$

$$\frac{1}{n} \sum_{k=1}^M \sum_{t=1}^M w_{tk} x_{jk} \left( \sum_{l=1}^n x_{lt} \right) = 0$$

$$S_{i.}^* = \frac{1}{n} \sum_{l=1}^n \left( \sum_{k=1}^M \sum_{t=1}^M x_{it} w_{tk} x_{lk} \right) =$$

$$\frac{1}{n} \sum_{k=1}^M \sum_{t=1}^M w_{tk} x_{it} \left( \sum_{l=1}^n x_{lk} \right) = 0$$

$$S_{..}^* = \frac{1}{n} \sum_{l=1}^n S_{l.}^* = \frac{1}{n} \sum_{l=1}^n S_{.l}^* = 0$$

Puesto que

$$S_{i.}^* = S_{.j}^* = 0$$

y siendo

$$S^* = \frac{1}{n} \sum_{l=1}^n S_{ll}^*$$

se llega a que

$$d_{.j}^{*2} = S^* + S_{jj}^* - 2S_{.j}^* = S^* + S_{jj}^*$$

$$d_{i.}^{*2} = S_{ii}^* + S^*$$

$$d_{..}^{*2} = \frac{1}{n} \sum_{i=1}^n d_{i.}^{*2} = S^* + S^* = 2S^*$$

luego

$$d_{ij}^{*2} - d_{i.}^{*2} - d_{.j}^{*2} + d_{..}^{*2} = S_{ii}^* + S_{jj}^* - 2S_{ij}^* -$$

$$S_{ii}^* - S_{.i}^* - S_{.j}^* - S_{jj}^* + 2S_{..}^* = -2S_{ij}^*$$

Así pues, multiplicando por  $-\frac{1}{2}$ , queda:

$$S_{ij}^* = \frac{1}{2} (d_{i.}^{*2} + d_{.j}^{*2} - d_{..}^{*2} - d_{ij}^{*2})$$

C.Q.D.

La transformación anterior fue descrita por Young & Houselholder (1938) y utilizada por Torgerson (1958) que estimó un índice promedio de disimilaridad para cada par de estímulos y, transformando dicho promedio mediante la proposición anterior, ajustó dichos valores según el método

$$S_{ijr}^* = X_i W_r X_j^t$$

para  $W_r = I$ . Este procedimiento es conocido como *procedimiento métrico clásico* y es utilizado hoy día por la mayoría de los métodos de MDS para la generación de los valores iniciales en los procedimientos iterativos de la matriz de configuración  $X$ . Por esa razón vamos a estudiar a continuación dicho procedimiento.

## 2.2 El procedimiento métrico clásico.

Como comentábamos en el apartado anterior, el método descrito por Torgerson es utilizado hoy día por la mayoría de los métodos de MDS para la obtención de una configuración inicial. Este método viene pues a resolver uno de los problemas planteados en el capítulo anterior, cuando hablábamos de la necesidad de partir en los métodos iterativos de una buena configuración inicial, en lugar de tener que hacerlo mediante una configuración totalmente aleatoria.

Supongamos que poseemos una matriz de disimilaridades como matriz de datos iniciales. En primer lugar, el método clásico transforma esas proximidades en distancias estimadas mediante un procedimiento conocido como *estimación de la constante aditiva*, el cual se basa en la siguiente relación entre distancias y disimilaridades:

$$d_{ijr}^* = d_{ijr} + c_r$$

Torgerson demostró que la menor constante  $c_r$  para cada sujeto fijo  $r$ , que garantiza la verificación de la desigualdad triangular para cualquier tripleta de estímulos  $(s, k, l)$ , viene dada por

$$c_r = \text{Max}_{(s,k,l)} \{d_{slr} - d_{skr} - d_{klr}\}$$

Así pues, considerando de esa forma la constante para cada sujeto  $r$ , las disimilaridades son transformadas en distancias estimadas  $d_{ijr}^*$ , bajo el punto de vista métrico del método de la constante aditiva descrito por Torgerson.

El siguiente paso consiste en transformar las distancias estimadas en productos escalares. Para ello se emplea la relación que demostrábamos en la proposición anterior, o sea,

$$S_{ij}^* = \frac{1}{2} (d_{i.}^{*2} + d_{.j}^{*2} - d_{..}^{*2} - d_{ij}^{*2})$$

Una vez obtenida la matriz de productos escalares estimados  $S^*$  el método factoriza dicha matriz en el producto de la configuración por su traspuesta, o sea

$$S^* = XX^t$$

Para obtener las coordenadas de la configuración estimadas por esa descomposición, se pueden utilizar diferentes métodos como, por ejemplo, los asociados al análisis de componentes principales. No obstante suele ser usual utilizar en MDS el método descrito por Eckart-Young para realizar dicha descomposición.

Para finalizar, indiquemos que la forma de actuar de este método genera numerosos problemas si se considera como solución al problema del MDS, en lugar de como procedimiento de obtención de la primera configuración. Uno de esos problemas y que puede considerarse como el más importante, es que resulta necesario medir en primer lugar las disimilaridades mediante algún factor de escala y, por otro lado, las perturbaciones aleatorias que aparecen al obtener cualquier medida de disimilaridad, se aumentan por considerar los cuadrados de los datos.

## 2.3 Modelos y medidas de ajuste.

Como ya se ha dicho anteriormente, la idea fundamental del MDS consiste en que las distancias entre los puntos se corresponden con las disimilaridades entre los objetos a los que representan. Una buena forma de poder observar dicha relación sería mediante el empleo del diagrama de Shepard o diagrama de dispersión. Así pues, si representamos los puntos definidos por los pares  $(d_{ijr}, d_{ijr}^*)$ , parece evidente que lo deseable resulta ser que pequeñas disimilaridades se correspondan con pequeñas distancias a la vez que grandes disimilaridades se correspondan con grandes distancias. Geométricamente esto quiere decir que la gráfica resultante de la representación debe tener un aspecto claramente creciente, de forma que su crecimiento sea lento a la izquierda y vaya acelerando rápidamente cuanto más nos aproximemos a la derecha. Lógicamente esta relación se invierte en el caso en que consideremos similaridades como datos iniciales en lugar de juicios de disimilaridad.

Como veíamos antes, las diferentes formas de relacionar las disimilaridades con las distancias, dan lugar a diferentes tipos de MDS. Aunque ésta no es la única causa que origina diferentes tipos de análisis ya que otros factores, como el método que determina la bondad del ajuste o el tipo de datos iniciales, constituyen importantes diferencias entre los métodos, sí que es uno de los principales factores para la clasificación de las técnicas del MDS.

La relación entre disimilaridades y distancias se formaliza matemáticamente a través de una aplicación  $f$  de forma que según sea dicha aplicación así se distinguirá el método. Así pues, sea

$$f : D \mapsto D^*$$

dicha aplicación, donde  $D$  será el espacio de las disimilaridades y  $D^*$  el de las distancias estimadas. Si la relación  $d^* = f(d)$  conserva las propiedades de las proximidades estaremos en MDS métrico. Por el contrario, si únicamente preserva el orden de las disimilaridades, estaremos ante aplicaciones cuya principal característica estriba en la monotonía y definirán un MDS no métrico.

### 2.3.1 Tipos de funciones $f$ y modelos asociados.

Veamos a continuación diferentes modelos en base a una clasificación según el tipo de función que se defina.

**Modelo absoluto.**

Se denomina modelo absoluto de MDS al modelo que queda determinado por la relación

$$d_{ij}^* = d_{ij} \quad \forall i, j$$

Este modelo es el más simple y pretende ajustar de forma exacta cada proximidad con distancia para cada par de estímulos  $i, j$  y para cada sujeto  $r$ . Si en esa representación no considerásemos todos los posibles pares  $(i, j)$  sino únicamente los pares observados, obtendríamos su forma natural ante observaciones faltantes. En realidad, es un modelo de ajuste bastante pobre y claramente métrico.

**Modelo proporcional.**

Este modelo se corresponde con el ajuste

$$d_{ij}^* = f(d_{ij}) = b \cdot d_{ij} \quad \forall i, j$$

el cual en forma matricial quedaría explícitamente como

$$X_i W X_j^t = b \cdot d_{ij}$$

siendo  $W$  la matriz de la métrica.

Esta función indica un cambio en la escala de obtención de las disimilaridades, como estimación de las distancias. De nuevo pertenece al grupo métrico del MDS. La característica fundamental de este modelo y de ahí su nombre, es el hecho de que conserva la proporcionalidad entre disimilaridades o sea

$$\frac{d_{ij}^*}{d_{kt}^*} = \frac{bd_{ij}}{bd_{kt}} = \frac{d_{ij}}{d_{kt}} = \frac{f(d_{ij})}{f(d_{kt})}$$

**Modelo de tipo intervalo.**

Este modelo resulta como una extensión natural del modelo anterior, de forma que ahora se calculan las distancias estimadas en virtud de una transformación lineal completa de las disimilaridades de partida. Su forma sería

$$d_{ij}^* = f(d_{ij}) = a + b \cdot d_{ij} \quad \forall i, j$$

En este modelo, la proporción entre las diferencias de distancias es idéntica a la proporción entre las correspondientes diferencias de disimilaridades, de ahí el nombre de intervalo.

Evidentemente, la función  $f$  no tiene por qué ser de tipo lineal. En principio podría definirse cualquier función. No obstante a un nivel experimental se han obtenido como principales funciones no lineales

- $f(d_{ij}) = \log(d_{ij})$
- $f(d_{ij}) = a + b \cdot \log(d_{ij})$
- $f(d_{ij}) = a + b \cdot \exp(d_{ij})$
- $f(d_{ij}) = a + b \cdot d_{ij} + c \cdot d_{ij}^2$

### MDS no métrico.

Como antes comentábamos, los modelos anteriores son todos métricos en el sentido de que conservan las propiedades de las disimilaridades. No obstante, puede que sólo estemos interesados en preservar el orden de éstas a través de las distancias estimadas. En estos casos, el único requerimiento que se le pide a la función  $f$  es que sea monótona. Así en estos casos, se define

$$d_{ij}^* = f(d_{ij}) \quad \forall i, j$$

de forma que

$$d_{ij} < d_{kl} \quad \rightarrow \quad d_{ij}^* < d_{kl}^*$$

### Modelo nominal.

Aún resulta posible concebir modelos de MDS más débiles que el anterior, como por ejemplo, modelos en los que  $f$  es menos restringida que en el caso de ser monótona. Consideremos por ejemplo el caso en que las proximidades son todas 0 ó 1. Podríamos desear que todas las distancias que representan 1 sean mayores que las que representan 0. De forma más general, si tenemos como datos  $a$ ,  $b$  ó  $c$ , éstos podrían tener asociados tres grupos de distancias disjuntos, uno para cada tipo de dato. Así pues, todo lo que las distancias

representan son un carácter cualitativo, lo cual podría enunciarse definiendo  $f$  de la forma

$$d_{ij}^* = f(d_{ij}) \quad \forall i, j$$

de manera que

1. Si  $d_{ij} \neq d_{kl} \rightarrow d_{ij}^* \neq d_{kl}^*$
2. Si  $d_{ij} = d_{kl} \rightarrow d_{ij}^* = d_{kl}^*$

### 2.3.2 Errores y medidas de ajuste.

Todos los modelos expuestos anteriormente ajustan proximidad a distancia mediante alguna ecuación de igualdad. Sin embargo, parece más razonable el hecho de no considerar la igualdad en dichas ecuaciones en virtud de los errores que se cometen cuando se obtiene cualquier dato de forma experimental. Bajo esa hipótesis, resulta más real en lugar de considerar que

$$d_{ij}^* = f(d_{ij})$$

suponer que la relación entre disimilaridades y distancias estimadas es de tipo aproximado o sea

$$d_{ij}^* \approx f(d_{ij})$$

dependiendo de un error asociado a cada par, que notaremos por  $e_{ij}$ , el cual puede en principio asociarse a la forma de obtención de la disimilaridad, o al procedimiento iterativo, etc... Así pues, se tiene que

$$e_{ij}^2 = (d_{ij}^* - f(d_{ij}))^2$$

La suma de los errores al cuadrado, nos dará una idea de como es el error asociado a la función  $f$ . Por tanto, consideremos para todos los pares  $(i, j)$  realmente observados

$$\sum_{i=1}^n \sum_{j=i+1}^n (d_{ij}^* - f(d_{ij}))^2$$

esta expresión se suele escribir de la forma

$$\sum_{i < j}^n (d_{ij}^* - f(d_{ij}))^2$$

No obstante, la expresión anterior puede resultar engañosa, ya que depende de las unidades de medida que hayan sido tomadas. Así pues, el problema se suele resolver normalizando dicha expresión, o sea, expresando dicha cantidad en relación con el tamaño de  $X$ . Concretamente, si dividimos por la cantidad

$$\sum_{i < j} d_{ij}^{*2}$$

obtendremos el valor que notaremos por  $S_f^2$  y cuya expresión es

$$S_f^2 = \frac{\sum_{i < j} e_{ij}^2}{\sum_{i < j} d_{ij}^{*2}} = \frac{\sum_{i < j} (d_{ij}^* - f(d_{ij}))^2}{\sum_{i < j} d_{ij}^{*2}}$$

Considerando la raíz cuadrada de esa cantidad, obtendremos el valor de la función STRESS asociada a la función  $f$ .

**Definición 2.3.2** Se llama  $f$ -STRESS o función STRESS asociada a la función  $f$  y se notará por  $S_f$  a la cantidad

$$S_f = \sqrt{\frac{\sum_{i < j} (d_{ij}^* - f(d_{ij}))^2}{\sum_{i < j} d_{ij}^{*2}}}$$

donde

$$S_f = 0 \quad \leftrightarrow \quad d_{ij}^* = f(d_{ij})$$

para todos los pares  $(i, j)$  observados.

Por último, hay que añadir que se ha tomado la raíz cuadrada porque generalmente, los valores que se obtienen experimentalmente, suelen ser demasiado pequeños si no se toma la raíz cuadrada y en consecuencia se tiende a despreciar cantidades significativas. Por otro lado, hay que decir también que esta

medida de la bondad del ajuste suele ir asociada generalmente con la regresión monótona.

Si en lugar de considerar los valores  $f(d_{ij})$  como distancias estimadas, consideramos las distancias ordenadas definidas en el capítulo anterior y que notamos  $r_{ij}$ , sin más que sustituir en la expresión del STRESS asociado a  $f$ , los valores  $f(d_{ij})$  por los  $r_{ij}$  tendremos el *coeficiente de alineación*. Veámoslo.

Llamemos  $2S_0$  a la cantidad  $S_f^2$ . Así pues tendremos

$$2S_0 = \frac{\sum_{i<j} (d_{ij}^* - r_{ij})^2}{\sum_{i<j} d_{ij}^{*2}}$$

Si desarrollamos la expresión anterior se tiene que

$$2S_0 = \frac{\sum_{i<j} d_{ij}^{*2} + \sum_{i<j} (r_{ij})^2 - 2 \sum_{i<j} d_{ij}^* r_{ij}}{\sum_{i<j} d_{ij}^{*2}}$$

y puesto que

$$\sum_{i<j} d_{ij}^{*2} = \sum_{i<j} (r_{ij})^2$$

ya que las  $r_{ij}$  no son más que una ordenación de las distancias, tendremos que

$$2S_0 = 1 + 1 - \frac{2 \sum_{i<j} d_{ij}^* r_{ij}}{\sum_{i<j} d_{ij}^{*2}}$$

de donde queda que

$$S_0 = 1 - \frac{\sum_{i<j} d_{ij}^* r_{ij}}{\sum_{i<j} d_{ij}^{*2}}$$

**Definición 2.3.3** Se llama *Coficiente de Monotonía de Guttman* a la cantidad



$$\mu = \frac{\sum_{i < j} d_{ij}^* r_{ij}}{\sum_{i < j} d_{ij}^{*2}}$$

o bien en términos de las distancias ordenadas,

$$\mu = \frac{\sum_{i < j} d_{ij}^* r_{ij}}{\sqrt{\left(\sum_{i < j} d_{ij}^{*2}\right) \left(\sum_{i < j} (r_{ij})^2\right)}}$$

La recta de regresión no pasa por el origen ni por el centro del diagrama imagen, a no ser que la solución sea perfecta, lo cual implicaría que el coeficiente  $\mu$  valga 1. Se suele utilizar el coeficiente de correlación de Pearson para medir la calidad del ajuste en el diagrama de dispersión, tal y como antes se comentaba, ya que se puede demostrar que ambos muestran el mismo significado.

A nivel práctico,  $\mu$  tiene la desventaja de que toma valores cercanos a 1, incluso cuando la solución no es muy buena. Así pues, lo que se suele hacer es sustituirlo por el coeficiente de alineación el cual varía mucho más a nivel numérico.

**Definición 2.3.4** Se llama *Coficiente de Alineación* a la cantidad

$$K = \sqrt{1 - \mu}$$

Este coeficiente nos una medida de la *variabilidad no explicada* de los puntos en el diagrama. Por otro lado,  $\mu^2$  ofrece una medida de la *variabilidad explicada* en dicho diagrama.

### 2.3.3 Minimización del STRESS y de la ALINEACION.

En este apartado, se trata de introducir un procedimiento general para resolver el problema del MDS en base a las medidas del STRESS y del Coeficiente de Alineación. Así pues, tendremos que encontrar una configuración  $X$  y una

transformación  $f$  (en el caso del STRESS) de forma que juntas optimicen la relación

$$f(d_{ij}) \simeq d_{ij}^* \quad \forall i, j \text{ observados.}$$

Considerando el modelo general descrito en la relación anterior, deberemos minimizar la cantidad

$$S = \sum_{i < j}^n (d_{ij}^* - f(d_{ij}))^2$$

donde supondremos que  $f$  debe ser conocida, tal y como ocurre en el modelo de MDS proporcional. Así pues, el problema consiste en encontrar una configuración  $X$  de forma que minimice la cantidad  $S$ . Bajo un punto de vista general, el problema del MDS puede dividirse en dos problemas de optimización de forma que la estrategia que se sigue para su resolución consiste en optimizar una parte suponiendo fija la otra, entrando en un procedimiento iterativo. Las dos partes a optimizar son

1. La función de transformación  $f$ .
2. La matriz de configuración  $X$ .

Explícitamente el proceso que se sigue consiste en:

- **Paso 1.**- Dada una configuración  $X$ , calcular la distancia entre los puntos y encontrar la función  $f$ , dentro de la familia de funciones deseada, que minimiza la cantidad  $S$ .
- **Paso 2.**- Dada la función  $f$ , calcular a través de ella la distancia objetivo entre cada par y obtener la configuración  $X$  asociada a dichas distancias.

Una vez completado el segundo paso se vuelve al primero y así sucesivamente, hasta completar un procedimiento iterativo, cuya finalización de parada viene determinada por diversos criterios como puede ser el número de iteraciones o la variación del valor que toma  $S$  de una iteración a otra.

La determinación de la función  $f$  es un aspecto, en general, difícil de resolver ya que depende fundamentalmente del tipo de análisis que se esté realizando. Veamos pues, como se determina el segundo paso, el cual es más genérico.

Para centrar el problema, si suponemos que tenemos  $n$  estímulos y la representación se desea realizar en dimensión  $m$  se trata de encontrar  $n \times m$  coordenadas  $x_{ka}$  de forma que la función  $S$  alcance un mínimo en dicha matriz  $X$ . Desde otro punto de vista, puede considerarse el encontrar un punto en dimensión  $n \times m$  que minimice la función  $S$ , lo cual es tanto como considerar en vez de la matriz  $X$  el  $\text{Vec}(X) \in \mathbb{R}^{n \cdot m}$ .

El procedimiento a seguir consistirá en que dada una componente de  $\text{Vec}(X)$ ,  $x_{ia}$ , trataremos de construir un proceso iterativo que paso a paso minimice el valor de  $S$  en la dimensión representada por  $x_{ia}$  suponiendo que las restantes componentes permanecen constantes. Así pues, ese procedimiento minimiza  $S$  componente a componente. Sea pues

$$\delta_{ij} = f(d_{ij})$$

el valor obtenido para todo par  $(i, j)$ , de la función  $f$  fijada por el primer paso. Así

$$S = \sum_{i < j}^n (d_{ij}^* - \delta_{ij})^2 = \sum_{i < j} \left[ \left( \sum_{a=1}^m (x_{ia} - x_{ja})^2 \right)^{1/2} - \delta_{ij} \right]^2$$

luego tenemos que

$$\begin{aligned} \frac{\partial S}{\partial x_{ia}} &= \frac{\partial}{\partial x_{ia}} \sum_{ij} (d_{ij}^* - \delta_{ij})^2 = \\ &= \frac{\partial}{\partial x_{ia}} \sum_j \sum_i (d_{ij}^* - \delta_{ij})^2 = \sum_j \frac{\partial}{\partial x_{ia}} \sum_i (d_{ij}^* - \delta_{ij})^2 = \\ &= \sum_j \frac{\partial}{\partial x_{ia}} (d_{ij}^* - \delta_{ij})^2 = \sum_j 2(d_{ij}^* - \delta_{ij}) \frac{\partial}{\partial x_{ia}} (d_{ij}^* - \delta_{ij}) = \\ &= \sum_j 2(d_{ij}^* - \delta_{ij}) \frac{\partial}{\partial x_{ia}} \left[ \left( \sum_{a=1}^m (x_{ia} - x_{ja})^2 \right)^{1/2} - \delta_{ij} \right] = \\ &= \sum_j 2(d_{ij}^* - \delta_{ij}) \frac{\partial}{\partial x_{ia}} \left( \sum_{a=1}^m (x_{ia} - x_{ja})^2 \right)^{1/2} - 0 = \\ &= \sum_j 2(d_{ij}^* - \delta_{ij}) \frac{1}{2} \left( \sum_{a=1}^m (x_{ia} - x_{ja})^2 \right)^{-1/2} \frac{\partial}{\partial x_{ia}} \sum_{a=1}^m (x_{ia} - x_{ja})^2 = \end{aligned}$$

$$\begin{aligned} & \sum_j (d_{ij}^* - \delta_{ij})(d_{ij}^{*2})^{-1/2} \frac{\partial}{\partial x_{ia}} (x_{ia} - x_{ja})^2 = \\ & \sum_j (d_{ij}^* - \delta_{ij}) d_{ij}^{*-1} 2(x_{ia} - x_{ja}) \frac{\partial}{\partial x_{ia}} (x_{ia} - x_{ja}) = \\ & 2 \sum_j \left(1 - \frac{\delta_{ij}}{d_{ij}^*}\right) (x_{ia} - x_{ja}) \end{aligned}$$

Haciendo  $\frac{\partial S}{\partial x_{ia}} = 0$ , tendremos que

$$\begin{aligned} & \sum_j \left(1 - \frac{\delta_{ij}}{d_{ij}^*}\right) (x_{ia} - x_{ja}) = 0 \Rightarrow \\ & \sum_j \left(1 - \frac{\delta_{ij}}{d_{ij}^*}\right) x_{ia} - \sum_j \left(1 - \frac{\delta_{ij}}{d_{ij}^*}\right) x_{ja} = 0 \Rightarrow \\ & \sum_j x_{ia} - \sum_j \frac{\delta_{ij}}{d_{ij}^*} x_{ia} - \sum_j x_{ja} + \sum_j \frac{\delta_{ij}}{d_{ij}^*} x_{ja} = 0 \Rightarrow \\ & nx_{ia} = \sum_j \frac{\delta_{ij}}{d_{ij}^*} x_{ia} + \sum_j x_{ja} - \sum_j \frac{\delta_{ij}}{d_{ij}^*} x_{ja} \Rightarrow \\ & x_{ia} = \frac{1}{n} \sum_j \frac{\delta_{ij}}{d_{ij}^*} (x_{ia} - x_{ja}) + \frac{1}{n} \sum_j x_{ja} \end{aligned}$$

Si consideramos que las coordenadas de los puntos recogidas en  $X$ , se encuentran dadas en un espacio centrado o sea, de forma que la suma de las coordenadas en cada uno de los  $m$  ejes de coordenadas vale 0, tendremos que  $\forall a = 1, \dots, m$

$$\sum_j x_{ja} = 0;$$

con lo que

$$x_{ia} = \frac{1}{n} \sum_j \frac{\delta_{ij}}{d_{ij}^*} (x_{ia} - x_{ja})$$

En el desarrollo anterior tendremos en cuenta que si  $i = j$  se define

$$\frac{\delta_{ii}}{d_{ii}^*} = 0$$

Para obtener con rigor una expresión mejor, consideremos

$$c_{ij} = \begin{cases} 1 + \sum_{k=1}^n \frac{\delta_{ik}}{d_{ik}^*} & \text{si } i = j \\ 1 - \frac{\delta_{ij}}{d_{ij}^*} & \text{si } i \neq j \end{cases}$$

**Proposición 2.3.5** *Las expresiones*

$$x_{ia} = \frac{1}{n} \sum_j c_{ij} x_{ja}$$

y

$$x_{ia} = \frac{1}{n} \sum_j \frac{\delta_{ij}}{d_{ij}^*} (x_{ia} - x_{ja})$$

son equivalentes.

DEMOSTRACIÓN.

Consideremos

$$\begin{aligned} x_{ia} &= \frac{1}{n} \sum_j c_{ij} x_{ja} = \\ &= \frac{1}{n} \left[ \sum_{j \neq i} c_{ij} x_{ja} + \sum_{j=i} c_{ij} x_{ia} \right] = \\ &= \frac{1}{n} \left[ \sum_{j \neq i} \left( 1 - \frac{\delta_{ij}}{d_{ij}^*} \right) x_{ja} + \left( 1 + \sum_{k=1}^n \frac{\delta_{ik}}{d_{ik}^*} \right) x_{ia} \right] = \\ &= \frac{1}{n} \left[ \sum_{j \neq i} x_{ja} - \sum_{j \neq i} \frac{\delta_{ij}}{d_{ij}^*} x_{ja} + x_{ia} + \sum_k \frac{\delta_{ik}}{d_{ik}^*} x_{ia} \right] = \end{aligned}$$

$$\frac{1}{n} \left[ \sum_{j \neq i} x_{ja} + x_{ia} - \sum_{j \neq i} x_{ja} \frac{\delta_{ij}}{d_{ij}^*} + \sum_k \frac{\delta_{ik}}{d_{ik}^*} x_{ia} \right] =$$

$$\frac{1}{n} \left[ \sum_j x_{ja} - \sum_{j \neq i} x_{ja} \frac{\delta_{ij}}{d_{ij}^*} + \sum_k \frac{\delta_{ik}}{d_{ik}^*} x_{ia} \right] =$$

cuando  $j = i$ ,  $\frac{\delta_{ij}}{d_{ij}^*} = 0$ , luego podemos escribir

$$\sum_{j \neq i} x_{ja} \frac{\delta_{ij}}{d_{ij}^*} = \sum_j x_{ja} \frac{\delta_{ij}}{d_{ij}^*} = (*)$$

como hemos supuesto  $X$  centrada tendremos tambien que

$$\sum_j x_{ja} = 0$$

luego

$$(*) = \frac{1}{n} \left[ 0 + \sum_j x_{ja} \frac{\delta_{ij}}{d_{ij}^*} + \sum_k \frac{\delta_{ik}}{d_{ik}^*} x_{ia} \right] = (**)$$

y finalmente, tomando  $k = j$  en el último sumatorio tendremos

$$(**) = \frac{1}{n} \sum_j \frac{\delta_{ij}}{d_{ij}^*} (x_{ia} - x_{ja})$$

C.Q.D.

Por tanto el procedimiento iterativo se puede escribir en virtud de la proposición anterior como

$$x_{ia}^{(t+1)} = \frac{1}{n} \sum_j c_{ij}^{(t)} x_{ja}^{(t)}$$

Hemos probado en la proposición anterior que

$$x_{ia} = \frac{1}{n} \sum_j c_{ij} x_{ja} = \frac{1}{n} \sum_j \frac{\delta_{ij}}{d_{ij}^*} (x_{ia} - x_{ja})$$

Como hemos supuesto  $X$  centrada, tenemos que

$$\sum_j x_{ja} = 0,$$

luego podemos escribir

$$\begin{aligned} x_{ia} &= \frac{1}{n} \left[ \sum_j x_{ja} - \sum_j \frac{\delta_{ij}}{d_{ij}^*} (x_{ja} - x_{ia}) \right] = \\ &= \frac{1}{n} \sum_j \left( x_{ja} - \frac{\delta_{ij}}{d_{ij}^*} (x_{ja} - x_{ia}) \right) = \\ &= \frac{1}{n} \sum_j \left( x_{ia} + \left(1 - \frac{\delta_{ij}}{d_{ij}^*}\right) (x_{ja} - x_{ia}) \right) \end{aligned}$$

y el procedimiento iterativo puede ponerse como

$$\begin{aligned} x_{ia}^{(t+1)} &= \frac{1}{n} \sum_j \left[ x_{ia}^{(t)} + \left(1 - \frac{\delta_{ij}}{d_{ij}^*}\right) (x_{ja}^{(t)} - x_{ia}^{(t)}) \right] = \\ &= x_{ia}^{(t)} - \frac{1}{n} \sum_j \left[ \left(1 - \frac{\delta_{ij}}{d_{ij}^*}\right) (x_{ia}^{(t)} - x_{ja}^{(t)}) \right]^{(t)} \end{aligned}$$

Así pues, el procedimiento iterativo para la obtención de la configuración minimizando el STRESS quedará de la forma

$$x_{ia}^{(t+1)} = x_{ia}^{(t)} - \frac{1}{n} \sum_j \left[ \left(1 - \frac{\delta_{ij}}{d_{ij}^*}\right) (x_{ia}^{(t)} - x_{ja}^{(t)}) \right]^{(t)}$$

o bien

$$x_{ia}^{(t+1)} = x_{ia}^{(t)} - \frac{1}{2n} \left[ \frac{\partial S}{\partial x_{ia}} \right]^{(t)}$$

### 2.3.4 Problemas que plantea el método del Gradiente.

Veremos en este apartado del método del gradiente, así como los problemas que acarrea su empleo.

Este método de minimización de una función en base a sus parámetros, moviéndolos en la dirección de la pendiente adecuada, plantea el problema de no garantizar que el mínimo obtenido sea el mínimo absoluto. No obstante, la mayoría de los procedimientos resuelven el problema de la convergencia hacia el valor deseado de forma satisfactoria. Normalmente, éstos ofrecen un historial de las iteraciones que muestra como el STRESS o cualquier otro criterio de convergencia utilizado, mejora en las iteraciones. Suele observarse que en las primeras iteraciones, el criterio mejora considerablemente y después tiende a estabilizarse en torno a un valor límite. Normalmente, las primeras tres o cuatro iteraciones consiguen casi todo el movimiento de los puntos. Después, la mayor parte de las restantes iteraciones mueven los puntos lentamente en torno al mínimo (local). En caso de que no sucedan las cosas de esa forma, resultará que la configuración inicial tomada no es buena, por lo que debe efectuarse un análisis utilizando otra.

Hay que resaltar que ninguno de los algoritmos existentes en MDS garantiza que la solución óptima final obtenida sea globalmente la mejor. Lo que sí se garantiza, es que sí es la mejor a nivel local. La razón de esto es clara y puede comprenderse fácilmente sin más que observar el movimiento de los puntos de la configuración  $X$  de una iteración a otra. El STRESS se busca minimizando las traslaciones para cada uno de los  $n \times m$  ejes. La resultante de esas traslaciones de base es lo que se denomina gradiente. El gradiente ofrece la dirección de movimiento de los puntos en la que el STRESS decrece de forma más rápida. Puesto que movemos los puntos en esa dirección, el método de optimización también es conocido como *método del paso descendente*. La longitud del vector gradiente queda determinada por el tamaño de paso determinado. Así pues, podemos pensar en el gradiente como un vector de  $X^{(t)}$  a  $X^{(t+1)}$ . No obstante, como el gradiente se obtiene mediante diferenciación, se estudia el cambio del STRESS en un entorno infinitesimal de  $X^{(t)}$ . Así pues, el proceso puede resumirse diciendo que se comienza en un punto dado; partiendo de ahí, se busca la dirección a seguir en un entorno muy reducido de dicho punto y una vez determinada, se llega hasta la nueva posición de dicho punto y así sucesivamente. Puesto que la visión que se posee en todo instante es local (en un entorno muy reducido del punto en el que nos encontremos),

sólo podemos afirmar que en cada momento se localice el camino más indicado dentro de ese entorno, pero no se asegura que en ese proceso lleguemos al mejor punto a nivel global.

Parece evidente que cuando se comienza el proceso de búsqueda es de crucial importancia partir del lugar adecuado para conseguir el mínimo final. Así pues, la elección de una configuración inicial  $X$  de forma aleatoria no es la mejor forma de llegar a obtener la solución del mínimo STRESS, mediante el método del gradiente. Por ello, la mayoría de los métodos modernos de MDS utilizan algún método, tal y como antes decíamos, para obtener de forma *racional* dicha configuración inicial.

## 2.4 Diferencias individuales en MDS.

A lo largo del desarrollo del MDS se han utilizado diferentes ajustes a modelos lineales. En este apartado vamos, sin embargo, a destacar unas técnicas que resaltan por el hecho de demostrar la importancia de permitir variaciones a nivel individual en la métrica euclídea. Estos desarrollos se deben fundamentalmente a los trabajos realizados por Takane, Young, De Leeuw, Carroll y Chang. Para ilustrar dicha variante en los métodos MDS vamos a exponer brevemente dos técnicas que pueden resumir las dos vertientes más importantes que existen en MDS tratado desde el punto de vista clásico del análisis de datos, sin la incorporación de la inferencia estadística. Estas técnicas son conocidas como ALSCAL e INDSCAL. La razón de exponerlas aquí tiene diversos fundamentos. Por un lado, puede decirse que ambas constituyen dos bloques en cuanto a la forma de desarrollar el MDS desde un punto de vista teórico. Por otro lado, puede considerarse que ambas separadamente constituyen los exponentes máximos del MDS no métrico y del MDS métrico, respectivamente. Cualquier otro método constituirá una variante de una u otra técnica.

La innovación que ambos métodos ofrecieron en el desarrollo del MDS se debe a la representación de las diferencias individuales mediante el modelo euclídeo ponderado. Este modelo se basa en la idea de realizar una única representación en un espacio euclídeo en la que las diferencias entre los individuos se pongan de manifiesto mediante la asignación de diferentes pesos a las dimensiones del espacio. Además, el hecho de que las dimensiones del espacio no puedan rotarse constituye uno de los principales atractivos de este método.

Aunque en los apartados anteriores se consideraban en algunos casos pon-

deraciones, merece la pena describir ambas técnicas por la riqueza de ideas nuevas que aportan en sus desarrollos, además de que sin ellas quedaría incompleta la visión general del desarrollo del MDS clásico que se pretende dar en este capítulo.

El método ALSCAL es un método que ha sido desarrollado en diferentes etapas por Takane, Young y De Leew. Es uno de los modelos de ajuste del modelo euclídeo ponderado a los datos, más general en cuanto al tipo de datos a ser tratados. Este procedimiento es apropiado para datos con posibles observaciones faltantes. Además pueden tratarse datos de tipo nominal, ordinal, de intervalo o de razón o proporción, con o sin simetría y con o sin replicaciones. Además este modelo incluye al modelo euclídeo no ponderado.

Para hacer un poco de historia, diremos que el modelo euclídeo ponderado y sus procedimientos de ajuste fueron propuestos por varios autores: Horan, Bloxom, Carroll y Chang. El más importante de todos fue el propuesto por Carroll y Chang denominado CANDECOMP en su primera etapa e INDSCAL en la segunda.

Brevemente podemos decir que INDSCAL es una generalización  $n$ -dimensional del trabajo en torno a la descomposición canónica a dos vías desarrollado por Eckart y Young en 1936. Este procedimiento se realiza después de una transformación inicial de las disimilaridades observadas en momentos producto, mediante un procedimiento alternante de obtención de los estimadores mínimo-cuadráticos de las ponderaciones de los individuos  $W$  (para estimaciones fijas de  $X$ ) obteniendo posteriormente estimadores mínimo cuadráticos de  $X$  (supuesta  $W$  fija). Este procedimiento posee dos aspectos que son relevantes. El criterio de ajuste de CANDECOMP se define en términos de los momentos producto calculados a partir de los datos *brutos* pero no en los términos de éstos. Así, INDSCAL no optimiza el ajuste entre el modelo euclídeo y los datos en el sentido estricto, sino que ajusta un vector producto y una transformación de los datos. En segundo lugar, debido a la forma de convertir disimilaridades en productos escalares, el procedimiento es métrico.

Todos los procedimientos anteriores, requieren de algunas hipótesis sobre los datos. Principalmente necesitan que los datos sean simétricos, que no tengan observaciones faltantes, que no tengan réplicas y que al menos sean definidos en el tipo intervalo. El primer procedimiento no métrico de Carroll y Chang se denominó NINDSCAL. Este consiste en dos fases siendo la primera la correspondiente a CANDECOMP y la segunda la incorporación de la regresión monótona de Kruskal sobre la que después comentaremos diversos aspectos,

sobre todo a la hora de la interpretación de un análisis con MDS. Ambas fases son utilizadas interactivamente. Mientras que la primera fase minimiza un criterio de ajuste denominado STRAIN (sobre el cual no nos detendremos aquí por no constituir un aspecto importante en nuestras pretensiones), la segunda minimiza el STRESS de Kruskal. Aunque el procedimiento lleva implícitas dos funciones diferentes, NINDSCAL no asegura la convergencia en un punto estable y a veces oscila o diverge después de algunas iteraciones.

Por esas razones, Carroll y Chang propusieron posteriormente otro procedimiento no métrico para minimizar el STRAIN, mejorando CANDECOMP. Esta aproximación que envuelve el STRAIN en todas las fases de cada iteración es el primer procedimiento no métrico de MDS estable para un modelo euclídeo ponderado. No obstante, por llevar implícito el STRAIN, no optimiza directamente entre el modelo de distancia y los datos brutos, sino que lo hace entre el producto escalar calculado a través de una función monótona adecuada de los datos y el producto escalar calculado a través de las coordenadas.

Veamos a continuación el desarrollo del método ALSCAL.

#### 2.4.1 El método ALSCAL.

Partimos de la base de que el método euclídeo ponderado utilizado es

$$d_{ijr}^{*2} = \sum_{a=1}^m w_{ra} (x_{ia} - x_{ja})^2$$

siendo  $w_{ia} \geq 0$ , aunque esta restricción de no negatividad es opcional. En cuanto al tipo de datos a ser tratados tenemos los cuatro tipos, nominal y ordinal como caso no métrico e intervalo y razón o proporción como caso métrico. Además, se distinguirá entre medidas discretas y continuas, así como tres tipos de condicionalidad, que serán incondicional, matriz condicional por filas y matriz condicional. Se notará por

- $O$  a las observaciones al cuadrado
- $D$  a las disparidades al cuadrado
- $D^*$  a las distancias al cuadrado

Cada uno de esos símbolos representan una colección de matrices.  $O_r$  para cada individuo  $r$ , con datos  $O_{ijr}$  para el par de estímulos  $(i, j)$ .  $D$  será la

colección de  $D_r$  con elementos  $d_{ijr}^2$  y  $D^*$  es la colección de las  $D_r^*$  con elementos  $d_{ijr}^{*2}$ .

Con esas definiciones podemos representar formalmente el problema del MDS como un problema de transformación. Se trata de obtener

$$t : O \longrightarrow D$$

de forma que

$$t(O_{ijr}) = d_{ijr}$$

donde  $t$  será una función que dependerá del tipo de los datos, del proceso y de la condicionalidad.

Para entender los diferentes niveles, procesos y restricciones de condicionalidad debe introducirse un concepto que es crucial para el desarrollo posterior. En este desarrollo se entenderán todos los datos como categóricos, o sea, se entenderá una variable de observación como formada por observaciones que pertenecen a una variedad de categorías de forma que todas las observaciones de una determinada categoría son empíricamente equivalentes. Esto es así ya que en el desarrollo de este método se supone que el punto de vista categórico de los datos ofrece una mayor precisión y veracidad en cuanto a la información que éstos suministran.

Tal y como después veremos, los tres tipos de restricciones de las medidas se refieren a tres aspectos de la observación categórica. Las restricciones del proceso se refieren a las relaciones entre las observaciones dentro de una misma categoría; las restricciones de nivel se refieren a las relaciones entre todas las observaciones entre diferentes categorías y por último, las restricciones de condicionalidad se refieren a la posibilidad de formar categorías. Los pasos a seguir son, en primer lugar, la fijación de las restricciones del proceso, después las de nivel y finalmente las condicionales.

Existen dos tipos de restricciones del proceso: una cuando se supone que la generación del proceso es discreta y otra cuando se supone continua. Así, si se supone el proceso discreto deberemos representar todas las observaciones de una categoría con el mismo número después de que se haya realizado la transformación  $t$ . Si por el contrario se supone continuo, cada observación dentro de una categoría determinada podría representarse como un número seleccionado dentro de un intervalo cerrado de  $\mathbb{R}$ . Formalmente podemos escribir

$$t^d : (O_{ijr} \sim O_{mno}) \Rightarrow (d_{ijr} = d_{mno})$$

donde  $t^d$  indica transformación discreta y  $\sim$  equivalente

$$t^c : (O_{ijr} \sim O_{mno}) \Rightarrow (d_{ijr}^- = d_{mno}^- \leq d_{ijr} \leq d_{ijr}^+ = d_{mno}^+)$$

y además

$$t^c : (O_{ijr} \sim O_{mno}) \Rightarrow (d_{ijr}^- = d_{mno}^- \leq d_{mno} \leq d_{ijr}^+ = d_{mno}^+)$$

donde  $t^c$  indica transformación continua,  $(d_{ijr}^-)$  es la cota inferior del intervalo y  $(d_{ijr}^+)$  es la cota superior.

En cuanto a esas relaciones, hay que destacar que una de las implicaciones de la equivalencia empírica (categórica) es que las cotas superiores e inferiores de todas las observaciones de una categoría particular son las mismas para todas las observaciones. También hay que resaltar que, para todas las observaciones de una categoría particular, las correspondientes observaciones de las nuevas escalas deben estar en el mismo intervalo pero no ser iguales.

Volvamos ahora al segundo conjunto de restricciones sobre las transformaciones de medida  $t$ ; las restricciones de nivel. Con estas restricciones se determina la naturaleza de la transformación  $t$  de forma que se corresponda con el nivel de medida supuesto de las variables de observación. Existen muchas restricciones interesantes pero nosotros nos centraremos en tres.

1. Para las variables nominales, no se introduce ninguna restricción ya que quedan determinadas con las restricciones del proceso.
2. Para las variables ordinales, se requiere además de las restricciones del proceso, que los números reales asignados a las observaciones en las diferentes categorías representen el orden de las observaciones empíricas o sea

$$t^o : (O_{ijr} < O_{mno}) \Rightarrow (d_{ijr} \leq d_{mno})$$

donde por  $t^o$  representamos las restricciones de orden.

3. Para las variables cuantitativas (intervalo o proporción) se necesita que los números reales asignados a las observaciones, estén relacionados linealmente con las observaciones o sea

$$t^{i,p} : d_{ijr} = \sum_{q=0,1}^p \delta_q O_{ijr}^q$$

donde el sumatorio empezará en 0 para las de intervalo y en 1 para las de proporción.

Hay que resaltar que se piensa en observaciones categóricas incluso si el nivel de medida es cuantitativo a pesar de que esto no resulte muy ilustrativo cuando cada categoría tenga una única observación. Así, la distinción entre discreta-continua es puramente formal.

Finalmente, vamos a detenernos en el último tipo de restricciones de medida, aquellas que se refieren a la condicionalidad de las observaciones. A veces, es posible que las características de las medidas de las observaciones sean condicionales en algunos aspectos de forma que no puedan compararse unas observaciones con otras. Esto suele ocurrir cuando en los juicios emitidos por diferentes sujetos resulte erróneo hacer comparaciones entre sujetos diferentes, debido a que las opiniones son condicionales a cada sujeto, según la naturaleza del experimento. Así pues, a esas observaciones suele referirse como *matrices de condicionabilidad* de forma que tiene sentido comparar dentro de una matriz pero no entre matrices. Es posible también tener *observaciones condicionadas por filas* así como *incondicionales*.

Formalmente diremos que el dominio de la transformación de medida  $t$  depende del tipo de condicionalidad. Para datos incondicionales, el dominio lo constituye el conjunto completo de observaciones y la transformación se notará por  $t$ . Para datos matriz-condicionales el dominio es una única matriz de datos y la transformación se notará por  $t_i$ . Por último, para los fila-condicional el dominio es una fila de una matriz y se notará la transformación por  $t_{ij}$ .

## EL SSTRESS

Para estudiar la bondad del ajuste, se define un criterio que entra dentro del entorno del STRESS. Así pues, se define un criterio de mínimos cuadrados sobre las distancias al cuadrado y que se notará por

$$\Phi^2(X, W, D) = \sum_{r=1}^R \sum_{i=1}^n \sum_{j=1}^{i-1} (d_{ijr}^{*2} - d_{ijr}^2)^2$$

Puesto que el criterio está dado en el entorno del STRESS pero con la diferencia de que en este caso está definido respecto a las distancias al cuadrado  $d_{ijr}^{*2}$  y a las distancias estimadas al cuadrado  $d_{ijr}^2$ , nos referiremos a él como SSTRESS (Nótese que  $d_{ijr}^2$  es el estimador mínimo cuadrático de  $d_{ijr}^{*2}$  y no el cuadrado del estimador mínimo cuadrático de  $d_{ijr}^*$ ).

Aunque el STRESS y el SSTRESS no son estrictamente equivalentes, la restricción monótona

$$t^o(O_{ijr}^2) = d_{ijr}^2$$

definida en base a  $O_{ijr}^2$  y a  $d_{ijr}^2$  sí que es equivalente a la misma restricción definida en base a  $O_{ijr}$  y a  $d_{ijr}$ . Esa equivalencia se mantiene claramente para los niveles de medida nominal y proporcional, aunque no ocurre lo mismo con el tipo intervalo ya que una relación lineal entre  $O_{ijr}$  y  $d_{ijr}$  implica una relación no lineal entre  $O_{ijr}^2$  y  $d_{ijr}^2$ . No obstante, esa dificultad será resuelta como después se verá. Así pues, puede decirse que para los cuatro tipos de medidas resultan equivalentes las restricciones

$$t(O_{ijr}) = d_{ijr} \Leftrightarrow t(O_{ijr}^2) = d_{ijr}^2$$

Con lo expuesto no quiere decirse que SSTRESS y STRESS sean equivalentes. Una de las diferencias fundamentales entre ambos estriba en que para grandes valores de  $d_{ijr}$  y  $d_{ijr}^*$  se obtiene un mayor énfasis con SSTRESS que con STRESS. Así pues, si consideramos

$$d_{ijr} = d_{ijr}^* + e_{ijr}$$

tendremos que con STRESS se obtiene una discrepancia total entre  $d_{ijr}$  y  $d_{ijr}^*$  de  $e_{ijr}^2$  mientras que con el SSTRESS se obtiene

$$\left[ d_{ijr}^{*2} - d_{ijr}^2 \right]^2 = \left[ d_{ijr}^{*2} - (d_{ijr}^* + e_{ijr})^2 \right]^2 = e_{ijr}^2 \left[ e_{ijr} + 2d_{ijr}^* \right]^2$$

luego en este caso influye la magnitud y el signo de  $e_{ijr}$  y la magnitud de  $d_{ijr}^*$ .

Aunque tanto STRESS como SSTRESS son perfectamente compatibles con las restricciones mencionadas anteriormente, para el desarrollo de ALSCAL es preferible el empleo del SSTRESS por razones de cálculo. Tal y como se mencionaba anteriormente, los pesos de las diferencias individuales  $W$  son lineales con respecto a las distancias al cuadrado pero no con respecto a las propias

distancias simples. Esto simplifica en gran medida el procedimiento de estimación ya que el estimador mínimo cuadrático de  $W$  se obtiene mediante operaciones matriciales elementales cuando se utiliza el SSTRESS como criterio de optimización.

### **El Algoritmo ALSCAL.**

El algoritmo ALSCAL lleva implícitas dos fases mayores y dos fases menores. La primera fase mayor contiene la estimación mínimo cuadrática de las observaciones  $D$ , bajo la hipótesis de que la configuración  $X$  y los pesos  $W$  son constantes. Esto equivale a resolver el problema de mínimos cuadrados condicional que minimiza el SSTRESS bajo la hipótesis de que  $X$  y  $W$  no son variables, lo que se puede escribir de la forma

$$\text{Min}_D [\Phi^2(D/X, W)]$$

La segunda fase mayor posee dos subfases separadas de minimización; la primera, resuelve el problema

$$\text{Min}_W [\Phi^2(W/X, D)]$$

y la segunda

$$\text{Min}_X [\Phi^2(X/W, D)]$$

Las dos fases menores son la de inicialización y la de terminación. Veamos detalladamente el proceso.

0.- **Fase de Inicialización.** Esta fase consiste en el cálculo de los valores iniciales de  $X$  y  $W$  directamente a partir de  $O$  mediante una modificación de la solución algebraica de Schöneman.

1.- **Fase de Representación Óptima.**

- (a) Calculamos  $D$  mediante el uso de  $X$  y  $W$ .
- (b) Se obtienen las disparidades  $D$  de la representación óptima en el sentido de los mínimos cuadrados de  $D^*$ , las observaciones  $O$  y las restricciones de medidas necesarias.
- (c) Se normaliza apropiadamente.

2.- **Fase de Terminación.** Se determina si la razón de la mejora del SS-TRESS es suficientemente baja para terminar. Si es así, se obtienen los resultados y se finaliza. Si no, se continúa en el siguiente paso.

3.- **Fase de estimación del modelo.**

- (a) Se calculan los estimadores mínimo cuadráticos de  $W$  a partir de la anterior configuración  $X$  y de la nueva matriz  $D$  proveniente del paso 1.c mediante regresión.
- (b) Se impone la restricción de no negatividad sobre  $W$  si fuese necesario, mediante las técnicas que se desarrollarán a tal efecto.
- (c) Se calcula el nuevo estimador mínimo cuadrático de la configuración  $X$  a partir de los nuevos pesos calculados en los dos pasos anteriores y a partir de la matriz  $D$  calculada en el paso 1.c mediante Newton-Raphson.
- (d) Se vuelve al paso 1.a.

Antes de comenzar a describir cada una de las fases, debemos hacer un comentario en el sentido de que en los desarrollos posteriores nos centraremos en el modelo euclídeo ponderado aplicado a datos simétricos sin observaciones faltantes. Esto se hace para simplificar el desarrollo. Un modelo euclídeo no ponderado se ajusta a partir de éste, sin más que saltar la fase de estimación de  $W$ , lo cual es equivalente a tomar los pesos igual a 1. Los datos asimétricos pueden manejarse cambiando los índices y órdenes de las matrices. Los datos faltantes se tratan eliminándolos del criterio de optimización y con estimaciones de éstos a través del modelo obtenido.

0.- **FASE DE INICIACIÓN.**

Supongamos que hay  $R$  matrices de productos escalares  $P_r$  una para cada sujeto  $r$ , de órdenes  $n$  ( $n$ =número de estímulos) de forma que

$$P_r = XW_rX^t$$

donde  $W_r$  es una matriz diagonal de pesos asociada al sujeto  $r$  y  $W$  es la matriz rectangular de los pesos de todos los sujetos.

La forma de obtener la matriz  $P_r$  nos la proporciona de forma aproximada el desarrollo

$$P_{ijr} = \frac{1}{2} (O_{i.r}^2 + O_{.jr}^2 - O_{..r}^2 - O_{ijr}^2)$$

donde

- $O_{i.r}^2 = \frac{1}{n} \sum_{k=1}^n O_{ikr}^2$
- $O_{.jr}^2 = \frac{1}{n} \sum_{k=1}^n O_{kjr}^2$
- $O_{..r}^2 = \frac{1}{n} \sum_{t=1}^n \sum_{k=1}^n O_{tkr}^2$

Además debemos tener en cuenta los siguientes casos anómalos:

- (a) Si los datos son asimétricos, promediamos cada  $O_{ijr}^2$  y  $O_{jir}^2$  dentro de cada matriz  $r$  antes del doble centrado.
- (b) Si los datos contienen elementos faltantes, cada elemento es estimado como la media de los juicios del sujeto.
- (c) La condicionalidad de los datos es ignorada.

El problema consiste en obtener  $X$  y  $W_r$  a partir de las  $P_r$  bajo la hipótesis de que  $X$  es una matriz por columnas de rango completo y los elementos diagonales de  $W_r$  son estrictamente positivos para al menos un sujeto.

Tal y como se ha planteado, el problema resulta muy general en el sentido de que no se ha impuesto ninguna restricción sobre las  $W_r$ . Por ello, de no decir nada al respecto, podríamos encontrarnos con que si  $T$  es una matriz no singular, diagonal, de orden  $m$  (siendo  $m$  el número de dimensiones) entonces

$$P_r = XT(T_{-1}W_rT_{-1})TX^t$$

con lo que la definición de  $W$  sería demasiado imprecisa. Así pues, si consideramos la expresión  $P_r = XW_rX^t$  y promediamos según el número de individuos tendremos

$$P = \frac{1}{R} \sum_{r=1}^R P_r = X \left( \frac{1}{R} \sum_{r=1}^R W_r \right) X^t$$

Llamando

$$D = \frac{1}{R} \sum_{r=1}^R W_r$$

y si se toma  $D = I$  el problema se reduce a encontrar  $X$  de forma que

$$P = XX^t$$

donde  $P$ , tal y como indicábamos antes, será el promedio de las  $P_r$ .

Una vez llegados a ese punto, la solución de la ecuación anterior queda determinada bien por Cholesky o bien por la descomposición espectral de  $P$ . Así pues, se tiene que  $X$  queda determinada salvo una rotación que notaremos por su matriz ortogonal  $K$ .

$$X = YK$$

Si despejamos  $W_r$  de la primera ecuación se tiene que

$$(X^tX)^{-1}X^tP_rX(X^tX)^{-1} = W_r$$

Además, se sabe que  $W_r$  podría ser diagonal. De esa forma, podemos seleccionar la matriz de rotación  $K$  de forma que

$$W_r = K^t(Y^tY)^{-1}Y^tP_rY(Y^tY)^{-1}K$$

sea diagonal para cada  $i$ .

(Nota:  $KK^t = K^tK = I$  y  $K^{-1} = K^t$ ).

Sea

$$C_r = (Y^t Y)^{-1} Y^t P_r Y (Y^t Y)^{-1}$$

Estamos ante el caso en que cualquier combinación lineal de  $R$  matrices  $C_r$  (con raíces diferentes) se usan para encontrar una rotación  $K$ . Supongamos que dicha combinación lineal es posible y la notamos por  $e$ . Entonces, se calcula el conjunto de vectores propios de

$$C_e = \sum_{r=1}^R e_r C_r$$

para encontrar  $K$  y calcular las  $W_r$  de la expresión

$$K^t C_r K = W_r$$

Así se obtiene la configuración  $X$  y los pesos  $W$ . Resulta evidente, dadas la hipótesis anteriores, que la solución es única (salvo permutaciones de las dimensiones). Hay que resaltar que la hipótesis de que existe una combinación lineal  $e$  es esencialmente equivalente a la hipótesis de que los pesos, al menos para un individuo  $r$ , son diferentes.

Tal y como hemos podido comprobar, en el desarrollo anterior necesitamos determinar  $P$  y  $e$ . Al primer problema se le ha dado solución. Una vez obtenida  $P$  se descompone de la forma

$$P = Y Y^t$$

para obtener  $Y$  que será la configuración arbitraria que mejor reproduce el promedio de los productos escalares. En cuanto al segundo problema de definir la mejor orientación de la configuración y los pesos asociados, se resuelve obteniendo una matriz de rotación  $K$  que diagonalice simultáneamente las matrices  $C_r$  tanto como sea posible lo cual se realiza mediante el procedimiento descrito en el trabajo de Young, Takane y Lewyckyj (1978).

### 1.- FASE DE REPRESENTACIÓN ÓPTIMA.

En esta fase se tratará de obtener una representación de los datos brutos óptima en el sentido de maximizar la correlación entre las observaciones y el modelo, respetando las características de las medidas de las observaciones. En esta fase se supondrá que  $X$  y  $W$  son constantes y se buscará  $D$  de forma que resuelva

$$\text{Min}_D [\Phi^2(D/X, W)]$$

- (a) **Cálculo de las distancias.** El primer paso consiste en calcular las  $D_r^*$  a partir de las actuales  $W$  y  $X$ .
- (b) **Representación óptima.** Se trata de transformar la representación óptima actual. Para ello, definimos una transformación lineal de las distancias, de forma que dicha transformación que notaremos por  $l$  minimiza el SSTRESS en el sentido de los mínimos cuadrados para valores fijos de  $W$  y de  $X$ . Así pondremos

$$d_{ijr}^2 = l(d_{ijr}^{*2})$$

Puesto que estamos haciendo regresión mínimo cuadrática de  $d_{ijr}^{*2}$  sobre  $O_{ijr}^2$  bajo las restricciones de medida mencionadas anteriormente, podemos representarlo tal y como se describe en el trabajo de Young, de Leew y Takane (1976) como un operador proyección cuya matriz será

$$E = Z(Z^t Z)^{-1} Z^t$$

donde  $Z$  será en general la matriz de los vectores que definen el espacio en el que se encuentra el vector de  $d_{ijr}^{*2}$  sobre el que se hace regresión.

Para la transformación de razón o proporción,  $Z$  se toma como  $O$ . Para la transformación de intervalo,  $Z$  se reduce al caso anterior después de haber estimado la correspondiente constante aditiva. En ambos casos, se obtiene el estimador mínimo cuadrático a través de las reglas clásicas de regresión.

En los casos ordinal y nominal,  $Z$  se define como una matriz de variables *falsas* que únicamente indican la necesidad de satisfacer las restricciones de medida. Para la transformación continua-ordinal los elementos relacionados conllevan violaciones del orden y para la transformación discreta-ordinal los elementos relacionados conllevan observaciones que son categóricamente equivalentes. Para el caso discreto-nominal,  $Z$  indica que las distancias que corresponden a observaciones categóricamente equivalentes van a ser ligadas. Obviamente, los estimadores mínimo cuadráticos en este caso sólo se forman con medias simples. Por último, para el caso continuo-nominal,  $Z$  indica las distancias que caen fuera del intervalo deseado. Aquí, los estimadores mínimo-cuadráticos serán los extremos de los intervalos para las distancias que caen fuera y las propias distancias para las que caen dentro.

Hay que resaltar que para algunas transformaciones,  $Z$  es conocida antes de que el análisis sea efectuado y en otros casos no. Especialmente, para las transformaciones discretas, excepto la discreta-ordinal,  $Z$  se conoce a priori y para las restantes  $Z$  sólo se conoce después de efectuado el análisis. Además, en esos casos  $Z$  varía de una iteración a otra, dependiendo de la naturaleza de las distancias. Para el estudio del método en cuestión, lo realmente interesante, es que se puede mostrar el proceso de la representación óptima como un operador proyección tal y como antes afirmábamos. Esto quiere decir que si definimos un vector columna  $d^*$  que contiene los  $Rn(n-1)/2$  elementos  $d_{ijr}^{*2}$  y otro vector columna  $d$  que contiene los correspondientes  $d_{ijr}^2$ , entonces,

$$d = Ed^*$$

Otro dato importante a destacar es el hecho de que esa ecuación viene dada en términos de datos incondicionales pero puede extenderse fácilmente a los datos condicionales. Para las matrices de datos condicionales, se define  $Z_r$  para cada individuo separadamente y después se construye la matriz  $Z$  por bloques y diagonal. Para matrices de datos condicionadas por filas, se define  $Z_{ir}$  para cada fila de cada matriz individual y se construye la matriz diagonal por bloques  $Z$  con las  $Z_{ir}$ .

La importancia fundamental de poder utilizar la ecuación anterior, es que ahora puede expresarse fácilmente el SSTRESS en forma matricial y de forma completa en términos de distancias. Por tanto, si definimos  $\tilde{E} = I - E$  entonces podremos escribir el SSTRESS de la forma

$$\Phi^2(X, W, D) = d^{*t} \tilde{E} d^*$$

y el SSTRESS normalizado de la forma

$$\Phi^2(X, W, D) = \frac{d^{*t} \tilde{E} d^*}{d^{*t} d^*} = d^{*t} \tilde{E} (d^{*t} d^*)^{-1} \tilde{E} d^*$$

Finalmente, vamos a hacer hincapié en el procedimiento para la estimación de la constante aditiva cuando los datos se definen en el nivel de medida de intervalo. En ese caso, para  $a$  y  $b$  desconocidos se tiene que

$$\begin{aligned} d_{ijr} &= a(O_{ijr}) + b \\ d_{ijr}^2 &= a^2(O_{ijr})^2 + 2ab(O_{ijr}) + b^2 \end{aligned}$$

Redefiniendo el problema de la forma

$$d_{ijr} = \alpha + \beta(O_{ijr}) + \gamma(O_{ijr})^2$$

se tratará ahora de encontrar  $\alpha$ ,  $\beta$  y  $\gamma$  bajo la restricción de que

$$\beta^2 = 4\alpha\gamma$$

Consideremos ahora las siguientes definiciones.

- Definimos el vector de parámetros  $\chi^t = (\alpha, \beta, \gamma)$ .
- Definimos una matriz  $R[n(n-1)/2] \times 3$  que notaremos por  $O$  (aunque es distinta de la  $O$  anterior), formada por polinomios de segundo grado de observaciones (uno en la primera columna, las observaciones en la segunda columna y las observaciones al cuadrado en la tercera columna).
- Definimos un vector columna  $d^*$  que contiene los  $R[n(n-1)/2]$   $d_{ijr}^{*2}$  en el mismo orden que los  $O_{ijr}$  de  $O$ .

Con esos elementos puede escribirse el SSTRESS de la forma

$$\Phi^2(\chi, \lambda | O, d^*) = (d^* - O_\chi)^t (d^* - O_\chi) + \lambda(\beta^2 - 4\alpha\gamma)$$

el cual minimizaremos resolviendo para  $\chi$  y  $\lambda$  (multiplicador Lagrangiano).

El estimador mínimo cuadrático para los parámetros restringidos es

$$\hat{\chi} = (O^t O)^{-1} O^t d^* + \lambda q$$

Para resolver el multiplicador Lagrangiano, se define

$$(O^t O)^{-1} g = q$$

donde

$$q^t = [q_1, q_2, q_3]$$

de forma que

$$g' = [-2\gamma, \beta, -2\alpha]$$

son las derivadas de la ecuación de restricción

$$\beta^2 = 4\alpha\gamma$$

Habr a pues que resolver

$$(\hat{\beta} + \lambda q_2)^2 = 4(\hat{\alpha} + \lambda q_1)(\hat{\gamma} + \lambda q_3)$$

De ah ı, se elegir a la mejor de las dos soluciones (la que minimiza el SSTRESS) evaluando el conjunto de elementos de  $\hat{\chi}$  asociado a cada soluci on.

(c) **Normalizaci on.**

El tercer paso consiste en normalizar la soluci on. Hay que hacer aqu ı dos consideraciones separadas, referentes a la normalizaci on de los par ametros (configuraci on, pesos y valores de la representaci on

óptima) y a la normalización de la función de pérdida, la cual se hace en cada iteración, mientras que para los parámetros, sólo es necesaria al final.

En cuanto a la normalización de los parámetros hay que decir que se normaliza la configuración de forma que la proyección media para cada dimensión vale cero y la varianza sobre la proyección de cada dimensión vale uno. Además se normalizarán los datos optimizados y no los originales como sucede en la mayoría de los métodos. El que sea así, se debe a problemas en el caso de tener matrices condicionales por filas.

Otro aspecto en lo referente a la normalización de los parámetros es normalizar la matriz de pseudoproductos escalares de forma que la suma de los cuadrados valga una constante.

Para normalizar la función de pérdida, se procede respecto a los datos optimizados  $D$ . Así pues, tendremos

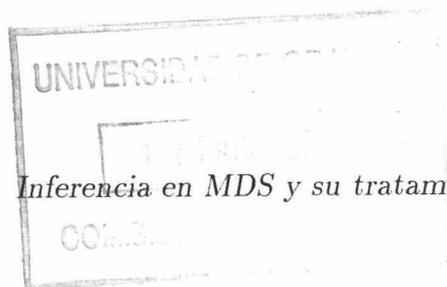
- i. Para datos incondicionales, calculamos la normalización del STRESS de la forma

$$\Phi^2 = \frac{\sum_{r=1}^R \sum_{i=1}^n \sum_{j=1}^{i-1} (d_{ijr}^2 - d_{ijr}^{*2})^2}{\sum_{r=1}^R \sum_{i=1}^n \sum_{j=1}^{i-1} d_{ijr}^4}$$

- ii. Para datos condicionales, se normaliza el SSTRESS para cada matriz y se obtiene su promedio

$$\Phi^2 = \frac{1}{R} \sum_{r=1}^R \left[ \frac{\sum_{i=1}^n \sum_{j=1}^{i-1} (d_{ijr}^2 - d_{ijr}^{*2})^2}{\sum_{i=1}^n \sum_{j=1}^{i-1} d_{ijr}^4} \right]$$

- iii. Finalmente para datos condicionales por fila, se calcula el SSTRESS normalizado para cada fila y después se obtiene el promedio.



*Inferencia en MDS y su tratamiento computacional.*

$$\Phi^2 = \frac{1}{Rn} \sum_{r=1}^R \sum_{i=1}^n \left[ \frac{\sum_{j=1}^{i-1} (d_{ijr}^2 - d_{ijr}^{*2})^2}{\sum_{j=1}^{i-1} d_{ijr}^4} \right]$$

## 2.- FASE DE TERMINACIÓN.

En esta fase únicamente se determina el valor del SSTRESS y se compara con el valor anterior. Si la diferencia es menor que una cantidad determinada se termina, si no se prosigue.

## 3.- FASE DE ESTIMACIÓN DEL MODELO.

En esta fase se resuelven sucesivamente dos problemas de mínimos cuadrados condicionales. La primera subfase resuelve el problema

$$\text{Min}_W[\Phi^2(W/X, D)]$$

y la segunda subfase resuelve

$$\text{Min}_X[\Phi^2(X/W, D)].$$

### (a) Cálculo de los pesos.

Para la estimación de  $W$  se obtienen las derivadas parciales del SSTRESS,  $\Phi^2(X, W, D)$  respecto a los elementos de  $W$  y se igualan a cero. Resolviendo el sistema en  $W$  se obtiene el resultado. Para simplificar la derivación se define una matriz  $Y$  de orden dado por  $[n(n-1)/2] \times M$ , cuyas columnas están formadas por la proyección en dimensión  $a$  de la distancia no ponderada entre los estímulos  $(i, j)$ , o sea donde

$$y_{ija} = (x_{ia} - x_{ja})^2$$

Se define igualmente la matriz  $D$  de orden  $R \times n(n-1)/2$  cuyas filas contienen las  $n(n-1)/2$  observaciones óptimas para cada individuo almacenando dichos datos en la misma situación que los de  $Y$ .

**Proposición 2.4.6** *El estimador mínimo cuadrático de la matriz de ponderaciones  $W$  viene determinado por la expresión*

$$W = DY(Y^tY)^{-1}$$

DEMOSTRACIÓN.

Tenemos que

$$d_{ijr}^2 \cong d_{ijr}^{*2} = \sum_{a=1}^M w_{ra} y_{ija}$$

o matricialmente

$$D^* \cong D = WY^t$$

Podemos poner

$$\begin{aligned} WY^t &= D^* \\ WY^tY(Y^tY)^{-1} &= D^*Y(Y^tY)^{-1} \\ W &= D^*Y(Y^tY)^{-1} \end{aligned}$$

Como  $D \cong D^*$  podemos decir que el estimador mínimo cuadrático de  $W$  es

$$W = DY(Y^tY)^{-1}$$

C.Q.D.

Existe una dificultad en la utilización de la regresión para la obtención de  $W$  ya que la condición de no negatividad de los pesos puede ser violada. Veamos a continuación como se resuelve el problema.

En primer lugar, se observa que el proceso de estimación presentado anteriormente es independiente para cada individuo, o sea, los valores estimados para los pesos de un individuo no afectan a los

de otro. Esto proviene del hecho de que el SSTRESS, el cual puede escribirse con la notación anterior como

$$\Phi^2(Y, W, D) = \text{tr} (D - WY^t)^t (D - WY^t)$$

puede descomponerse como la suma de elementos, cada uno de los cuales es una función de un único sujeto. Puesto que los pesos para un sujeto son independientes de los de otro, podemos imponer una restricción de no negatividad sobre los sujetos con pesos negativos sin tener que modificar los pesos de los demás. No obstante, hay que destacar que los pesos para un sujeto dado, no son independientes entre sí por lo que no puede asignárseles simplemente a los pesos negativos el valor cero para que se satisfaga la restricción, ya que de lo contrario, se destruiría la propiedad de mínimos cuadrados de los pesos estimados.

Para solucionar el problema, en primer lugar se obtienen los estimadores mínimo cuadráticos de  $W$  sin restricción, a través de la proposición anterior, los cuales serán los estimadores mínimo cuadráticos para los sujetos con pesos no negativos. Para aquellos sujetos que tengan pesos negativos, se procede de la siguiente forma: se elige uno de los pesos negativos del sujeto y se iguala a cero, que será el estimador mínimo cuadrático de dicho peso bajo la restricción de no negatividad, suponiendo constantes los demás pesos. El estimador mínimo cuadrático para un único peso del sujeto  $r$  será

$$w_{ra} = \frac{\left( d_r - \sum_{b \neq a} w_{rb} y_b \right)^t y_a}{(y_a^t y_a)}$$

donde  $y_a$  es la  $a$ -ésima columna de  $Y$ , la cual contiene los cuadrados de las distancias entre los puntos, como proyecciones en la  $a$ -ésima dimensión. Si este estimador mínimo cuadrático no restringido resulta negativo, se toma como valor cero. El proceso se repite hasta que todos los pesos del sujeto sean no negativos.

(b) **Cálculo de las coordenadas.**

La segunda subfase de la fase de estimación del modelo consiste en la determinación de las coordenadas de los estímulos  $X$ . Esta subfase

es más complicada ya que las derivadas parciales del SSTRESS con respecto a los elementos de  $X$  no son lineales en  $x_{ja}$  por lo que las derivadas constituyen un sistema de ecuaciones cúbicas. Veamos detalladamente el proceso de estimación de las coordenadas.

La estimación se realiza considerando los valores anteriormente calculados de  $D$  y  $W$ . Las coordenadas se determinan de una en una, minimizando el SSTRESS. Veámoslo.

**Teorema 2.4.7** *En la situación anterior, la ecuación que determina la estimación para cada coordenada  $x_{le}$  viene dada por*

$$\frac{1}{R} \sum_{r=1}^R c_r 4w_{re}^2 u_{ler} = 0$$

donde

$$c_r = \frac{1}{\sum_{i=1}^n \sum_{j=1}^{i-1} d_{ijr}^4}$$

$$u_{ler} = \sum_{j=1}^n [x_{le}^3 - 3x_{le}^2 x_{je} + 2x_{le} x_{je}^2 - b_{ljre}^2 x_{le} + b_{ljre}^2 x_{je}]$$

$$a_{ljre}^2 = \frac{d_{ijr}^2 - \sum_{a \neq e}^M w_{ra} (x_{ia} - x_{ja})^2}{w_{re}^2}$$

$$b_{ljre}^2 = a_{ljre}^2 - x_{je}^2$$

DEMOSTRACIÓN.

Consideremos

$$\Phi^2 = \frac{1}{R} \sum_{r=1}^R \frac{\sum_{i=1}^n \sum_{j=1}^{i-1} (d_{ijr}^{*2} - d_{ijr}^2)^2}{\sum_{i=1}^n \sum_{j=1}^{i-1} d_{ijr}^4}$$

Como  $\sum_{i=1}^n \sum_{j=1}^{i-1} d_{ijr}^4$  es constante respecto a  $x_{le}$ , si llamamos

$$c_r = \frac{1}{\sum_{i=1}^n \sum_{j=1}^{i-1} d_{ijr}^4}$$

podemos poner

$$\Phi^2 = \frac{1}{R} \sum_{r=1}^R c_r \sum_{i=1}^n \sum_{j=1}^{i-1} (d_{ijr}^{*2} - d_{ijr}^2)^2 = \frac{1}{R} \sum_{r=1}^R c_r \phi_r$$

donde  $\phi_r$  será la cantidad

$$\phi_r = \sum_{i=1}^n \sum_{j=1}^{i-1} (d_{ijr}^{*2} - d_{ijr}^2)^2$$

Así pues, se ha descompuesto  $\Phi^2$  en sumas separadas de las cantidades  $\phi_r$ , cada una de las cuales puede optimizarse con respecto a  $x_{le}$ . Como

$$d_{ijr}^{*2} = \sum_{a=1}^M w_{ra} (x_{ia} - x_{ja})^2$$

se tiene que

$$\begin{aligned} \phi_r &= \sum_{i=1}^n \sum_{j=1}^{i-1} \left[ \sum_{a=1}^M w_{ra} (x_{ia} - x_{ja})^2 - d_{ijr}^2 \right]^2 = \\ &= \sum_{i=1}^n \sum_{j=1}^{i-1} \left[ \sum_{a \neq e}^M w_{ra} (x_{ia} - x_{ja})^2 + w_{re} (x_{ie} - x_{je})^2 - d_{ijr}^2 \right]^2 = \\ &= \sum_{i=1}^n \sum_{j=1}^{i-1} \left[ d_{ijr}^2 - \sum_{a \neq e}^M w_{ra} (x_{ia} - x_{ja})^2 - w_{re} (x_{ie} - x_{je})^2 \right]^2 = \end{aligned}$$

$$w_{re}^2 \sum_{i=1}^n \sum_{j=1}^{i-1} \left[ \frac{d_{ijr}^2 - \sum_{a \neq e}^M w_{ra} (x_{ia} - x_{ja})^2}{w_{re}^2} - (x_{ie} - x_{je})^2 \right]^2$$

Sea

$$a_{ijre}^2 = \frac{d_{ijr}^2 - \sum_{a \neq e}^M w_{ra} (x_{ia} - x_{ja})^2}{w_{re}^2}$$

por lo que

$$\begin{aligned} \phi_r &= w_{re}^2 \sum_{i=1}^n \sum_{j=1}^{i-1} [a_{ijre}^2 - (x_{ie} - x_{je})^2]^2 = \\ &w_{re}^2 \sum_{i=1}^n \sum_{j=1}^{i-1} (a_{ijre}^2 - x_{ie}^2 + 2x_{ie}x_{je} - x_{je}^2)^2 \end{aligned}$$

Se trata de minimizar el SSTRESS respecto a  $x_{le}$ . Puesto que

$$\frac{\partial \Phi^2}{\partial x_{le}} = \frac{1}{R} \sum_{r=1}^R c_r \frac{\partial}{\partial x_{le}} \phi_r$$

bastará con calcular la derivada de  $\phi_r$  respecto a  $x_{le}$  con lo que

$$\frac{\partial}{\partial x_{le}} \phi_r = \frac{\partial}{\partial x_{le}} w_{re}^2 \sum_{i=1}^n \sum_{j=1}^{i-1} (a_{ijre}^2 - x_{je}^2 - x_{ie}^2 + 2x_{ie}x_{je})^2 = (*)$$

Llamando

$$\gamma_{ijre} = a_{ijre}^2 - x_{je}^2 - x_{ie}^2 + 2x_{ie}x_{je}$$

tendremos

$$\begin{aligned}
(*) &= \frac{\partial}{\partial x_{le}} w_{re}^2 \sum_{i=1}^n \sum_{j=1}^{i-1} \gamma_{ijre}^2 = \\
&w_{re}^2 \frac{\partial}{\partial x_{le}} \sum_{i=1}^{l-1} \sum_{j=1}^{i-1} \gamma_{ijre}^2 + w_{re}^2 \frac{\partial}{\partial x_{le}} \sum_{j=1}^{l-1} \gamma_{ijre}^2 + \\
&w_{re}^2 \frac{\partial}{\partial x_{le}} \sum_{i=l+1}^n \sum_{\substack{j=1 \\ j \neq l}}^{i-1} \gamma_{ijre}^2 + w_{re}^2 \frac{\partial}{\partial x_{le}} \sum_{i=l+1}^n \gamma_{ijre}^2 = (*)
\end{aligned}$$

Tenemos que

$$\gamma_{ijre}^2 = a_{ijre}^2 - x_{je}^2 - x_{ie}^2 + 2x_{ie}x_{je}$$

luego

- Si  $i \neq l$  y  $j \neq l$ ,

$$\frac{\partial}{\partial x_{le}} \gamma_{ijre}^2 = 0$$

- Si  $i = l$  y  $j \neq l$ ,

$$\frac{\partial}{\partial x_{le}} \gamma_{ijre}^2 = 2(x_{je} - x_{le})$$

- Si  $i \neq l$  y  $j = l$ ,

$$\frac{\partial}{\partial x_{le}} \gamma_{ijre}^2 = 2(x_{ie} - x_{le})$$

- Si  $i = l$  y  $j = l$ ,

$$\frac{\partial}{\partial x_{le}} \gamma_{ijre}^2 = -2x_{le}$$

luego

$$(*) = w_{re}^2 \left[ 2 \sum_{j=1}^{l-1} \gamma_{ljre}^2 2(x_{je} - x_{le}) + 2 \sum_{i=l+1}^n \gamma_{ilre}^2 2(x_{ie} - x_{le}) \right] =$$

$$4w_{re}^2 \left[ \sum_{j=1}^{l-1} \gamma_{ljre}(x_{je} - x_{le}) + \sum_{i=l+1}^n \gamma_{ilre}(x_{ie} - x_{le}) \right]$$

Como  $\gamma_{llre} = 0$ , podemos escribir el sumatorio para los  $i \geq l$  en lugar de los  $i > l$  con lo que tendremos

$$(*) = 4w_{re}^2 \left[ \sum_{j=1}^{l-1} \gamma_{ljre}(x_{je} - x_{le}) + \sum_{i=l}^n \gamma_{ilre}(x_{ie} - x_{le}) \right]$$

Como  $\gamma_{ijre} = \gamma_{jire}$  podemos escribir el segundo sumatorio de la forma

$$4w_{re}^2 \left[ \sum_{j=1}^{l-1} \gamma_{ljre}(x_{je} - x_{le}) + \sum_{i=l}^n \gamma_{lire}(x_{ie} - x_{le}) \right]$$

y sustituyendo  $i$  por  $j$  como índice queda

$$4w_{re}^2 \left[ \sum_{j=1}^{l-1} \gamma_{ljre}(x_{je} - x_{le}) + \sum_{j=l}^n \gamma_{ljre}(x_{je} - x_{le}) \right] =$$

$$4w_{re}^2 \sum_{j=1}^n \gamma_{ljre}(x_{je} - x_{le}) =$$

$$4w_{re}^2 \sum_{j=1}^n (a_{ljre}^2 - x_{je}^2 - x_{le}^2 + 2x_{le}x_{je})(x_{je} - x_{le}) = (*)$$

si llamamos ahora

$$b_{ljre}^2 = a_{ljre}^2 - x_{je}^2$$

será

$$(*) = 4w_{re}^2 \sum_{j=1}^n (b_{ljre}^2 - x_{le}^2 + 2x_{le}x_{je})(x_{je} - x_{le}) =$$

$$4w_{re}^2 \sum_{j=1}^n [b_{ljre}^2 x_{je} - b_{ljre}^2 x_{le} - x_{je} x_{le}^2 + x_{le}^3 + 2x_{le} x_{je}^2 - 2x_{le}^2 x_{je}]$$

y agrupando en  $x_{le}$  tendremos

$$(*) = 4w_{re}^2 \sum_{j=1}^n [x_{le}^3 - 3x_{le}^2 x_{je} + 2x_{le} x_{je}^2 - b_{ljre}^2 x_{le} + b_{ljre}^2 x_{je}]$$

Así pues, sustituyendo la expresión anterior en la ecuación

$$\frac{\partial \Phi^2}{\partial x_{le}} = \frac{1}{R} \sum_{r=1}^R c_r \frac{\partial}{\partial x_{le}} \phi_r = 0$$

se tiene lo que queríamos.

C.Q.D.

La ecuación dada por el teorema, se resuelve mediante procedimientos numéricos. Para ello, se suponen constantes las coordenadas excepto la  $x_{le}$ . Una vez resuelto para  $x_{le}$ , reemplazamos dicho valor en la iteración anterior y se procede a estimar la siguiente coordenada. Así, hasta estimar las coordenadas del estímulo  $l$ . Se continúa estímulo a estímulo hasta haber obtenido la matriz de configuración  $X$ . Una vez finalizado el proceso, se vuelve al principio del algoritmo de optimización y se comienza una nueva iteración.

#### 2.4.2 El método INDSCAL.

INDSCAL es un método para el tratamiento de datos de disimilaridad con diferencias individuales, el cual presupone que los datos están relacionados estrechamente con las distancias euclídeas ponderadas entre los puntos de un espacio común subyacente en el conjunto de los estímulos, a través de funciones lineales, monótonas o de cualquier otro tipo específico de función. Esta relación se puede expresar de la forma ya conocida

$$f_r(d_{ijr}) \cong d_{ijr}^*$$

$$d_{ijr}^* = \left[ \sum_{a=1}^M w_{ra} (x_{ia} - x_{ja})^2 \right]^{1/2}$$

donde  $d_{ijr}$  son las disimilaridades entre cada pareja de estímulos  $(i, j)$  observadas por el individuo  $r$ , y donde las funciones  $f_r$  serán consideradas generalmente lineales para el caso métrico y monótonas para el no métrico, aunque pueden ser de cualquier tipo deseado por el investigador y previamente especificado. Hay que destacar no obstante que se consideran diferentes funciones  $f_r$  para cada individuo  $r$  de forma separada, o sea, que las funciones pueden ser de diferentes tipos, según los datos del individuo en cuestión, por lo que INDSCAL es en este sentido un modelo *matricialmente condicional*.

Los datos de partida del modelo INDSCAL están constituidos por matrices tridimensionales (*estímulos*  $\times$  *estímulos*  $\times$  *individuos*) de disimilaridad, simétricas en los índices referentes a los estímulos. Por otro lado, la salida que ofrece INDSCAL está formada por dos matrices, una que contiene las coordenadas de los estímulos y otra que contiene lo que se denomina el espacio de los sujetos en la cual se encuentran los pesos asignados según los diferentes sujetos.

Para formalizar matemáticamente el desarrollo hay que hacer una distinción en primer lugar, en cuanto al espacio de representación de los estímulos analizados. En todo modelo ponderado pueden considerarse dos espacios de representación; uno el denominado *espacio de representación general* o global, en el que se representan los estímulos considerando simultáneamente toda la información proveniente de la totalidad de los individuos estudiados y otro el denominado *espacio de percepción privado* que está asociado con cada individuo de forma separada, en el que las coordenadas de los puntos se han obtenido teniendo en cuenta exclusivamente la información proveniente del individuo en cuestión.

Teniendo presente la separación entre esos dos conceptos, consideremos  $x_{ia}$  la coordenada en la dimensión  $a$  del estímulo  $i$  en el espacio general de los estímulos  $X$ . Por otra parte, podemos considerar igualmente la coordenada en la dimensión  $a$  del estímulo  $i$  pero expresada no en el espacio general sino en el espacio particular asociado al sujeto  $r$ , y que notaremos por  $x_{iar}$ . Debido a que tenemos la relación entre distancias

$$d_{ijr}^2 = \sum_{a=1}^M w_{ra}(x_{ia} - x_{ja})^2 = \sum_{a=1}^M (w_{ra}^{1/2}x_{ia} - w_{ra}^{1/2}x_{ja})^2$$

si llamamos

$$x_{iar} = w_{ra}^{1/2} x_{ia}$$

tendremos

$$d_{ijr}^{*2} = \sum_{a=1}^M (x_{iar} - x_{jar})^2$$

Luego la relación entre  $x_{ia}$  y  $x_{iar}$  viene dada por la expresión

$$x_{iar} = w_{ra}^{1/2} x_{ia}$$

La relación anterior, suministra una interpretación alternativa de la métrica euclídea ponderada, de forma que el modelo INDSCAL puede expresarse en términos de una clase particular de transformaciones del espacio conjunto general, además del cálculo de la métrica euclídea clásica. La clase de las transformaciones puede describirse algebraicamente, como transformaciones lineales con matriz de transformación diagonal.

Veamos a continuación las diferentes clases de metodologías INDSCAL según el tipo de análisis.

### INDSCAL métrico.

Esta versión métrica parte de un conjunto de matrices simétricas de disimilaridades de forma que si los datos iniciales son similaridades se convierten en disimilaridades multiplicándolos por -1 o bien restándolos del mayor valor.

Una vez obtenidos los datos este modelo utiliza el procedimiento clásico de Torgerson descrito en apartados anteriores para la resolución.

Sobre esta metodología clásica, hay que decir que más bien puede ser utilizada como una primera aproximación que como una conclusión según comentábamos anteriormente. Por ello puede decirse que INDSCAL propiamente dicho, para el caso métrico puede expresarse como el método basado en productos escalares que pasamos a definir a continuación.

### El modelo INDSCAL en forma de producto escalar.

La relación que anteriormente expresábamos entre las coordenadas consideradas en el espacio general o en el particular de un estímulo, para el sujeto  $r$ , puede expresarse matricialmente de la forma

$$X_r = W_r^{1/2} X$$

donde  $X_r$  es la configuración asociada al individuo  $r$  y  $X$  es la configuración global.  $W_r$  la matriz diagonal de los pesos asignados por el individuo  $r$  a las dimensiones de representación que serán obligatoriamente las mismas para cada individuo.

Dadas las distancias entre los estímulos para cada individuo  $r$ , las cuales pueden obtenerse por ejemplo por el procedimiento descrito en el apartado anterior, podemos obtener los productos escalares asociados mediante la relación establecida por el método clásico. Dados pues los productos escalares  $b_{ijr}$  podemos expresar éstos en relación con las coordenadas de los estímulos de la forma

$$B_r = X_r X_r^t = X W_r^{1/2} (W_r^{1/2})^t X^t = X W_r X^t$$

y expresando dicha relación en coordenadas tenemos

$$b_{ijr} = \sum_{a=1}^M w_{ra} x_{ia} x_{ja}$$

Se trata de resolver la ecuación

$$B_r = X W_r X^t$$

El método INDSCAL trabaja bajo una hipótesis simple: Sea  $y_{ijr}$  un dato de una matriz tridimensional. Cada uno de esos valores es una función de un conjunto de parámetros de la forma

$$y_{ijr} = \sum_{a=1}^M a_{ra} b_{ia} c_{ja}$$

En primer lugar, si comparamos esa expresión con la del producto escalar, se trata de resolver el sistema trilineal, donde  $a_{ra} = w_{ra}$ ,  $b_{ia} = x_{ia}$  y  $c_{ja} = x_{ja}$ .

Para resolver el problema procedemos de la siguiente forma: En primer lugar, situamos cada matriz  $B_r$  en un vector fila con  $n^2$  elementos y situamos todos esos vectores por filas en una matriz  $B$ ,  $R \times n^2$ . De igual forma, almacenamos las diagonales de las  $R$  matrices de pesos  $W_r$  en una matriz  $W$ ,  $R \times M$ . Por último, se construye una matriz  $V$  de la siguiente forma; la primera columna se construye para  $i = 1$  y para  $j = 1$ , calculando los productos

$x_{1a}x_{1a}, \forall a = 1, \dots, M$ ; la segunda columna de  $V$  se calcula de forma análoga pero  $i = 1, j = 2$ . Así se procede hasta completar los  $M \times n^2$  valores de  $V$ .

Una vez obtenidas las matrices anteriores, podemos escribir la relación conjunta, para todos los individuos simultáneamente, de los productos escalares de la forma

$$B = WV$$

Como no es posible explicar exactamente los productos escalares, dada la forma aproximada de su obtención, parece más lógico expresar la relación anterior de la forma

$$B \approx WV$$

Si se toma una configuración inicial  $X$ , por ejemplo, mediante el procedimiento clásico, se tendrán  $B$  y  $V$  fijos, por lo que la matriz  $W$  puede estimarse mediante regresión mínimo cuadrática de la siguiente forma:

La función de pérdida será

$$L = \text{tr} [(B - WV)^t(B - WV)]$$

y derivando tenemos que

$$\begin{aligned} \frac{\partial}{\partial W} L &= \frac{\partial}{\partial W} \text{tr} [(B - WV)^t(B - WV)] = \\ &= \frac{\partial}{\partial W} \text{tr} [(B^t - V^t W^t)(B - WV)] = \\ &= \frac{\partial}{\partial W} \text{tr} [B^t B - B^t W V - V^t W^t B + V^t W^t W V] = \\ &= \frac{\partial}{\partial W} \text{tr}(B B^t) - \frac{\partial}{\partial W} \text{tr}(B^t W) - \frac{\partial}{\partial W} \text{tr}(V^t W^t B) + \frac{\partial}{\partial W} \text{tr}(V^t W^t W V) = (*) \end{aligned}$$

- (1)  $\frac{\partial}{\partial W} \text{tr}(B^t B) = 0$ , por ser  $B$  cte.
- (2)  $\frac{\partial}{\partial W} \text{tr}(B^t W V) \stackrel{(P1)}{=} \frac{\partial}{\partial W} \text{tr}(V B^t W) = (V B^t)^t = B V^t$   
(P1) Invarianza frente a permutaciones cíclicas.

- (3)  $\frac{\partial}{\partial W} \text{tr}(V^t W^t B) \stackrel{(P2)}{=} \frac{\partial}{\partial W} \text{tr}[(V^t W^t B)^t] =$   
 $= \frac{\partial}{\partial W} \text{tr}(B^t W V) \stackrel{(2)}{=} B V^t$   
(P2) Invarianza frente a trasposición.
- (4)  $\frac{\partial}{\partial W} \text{tr}(V^t W^t W V) \stackrel{(P1)}{=} \frac{\partial}{\partial W} \text{tr}[(V V^t W^t W)] \stackrel{(P3)}{=}$   
 $= \frac{\partial}{\partial W} \text{tr}[(V V^t W^t)_c W] + \frac{\partial}{\partial W} \text{tr}[V V^t W^t W_c] =$   
 $= (V V^t W^t)^t + \frac{\partial}{\partial W} \text{tr}[W_c^t W (V V^t)] =$   
 $= W V V^t + \frac{\partial}{\partial W} \text{tr}[(V V^t W_c^t) W] =$   
 $= W V V^t + W V V^t = 2 W V V^t$   
(P3) Regla del producto. El subíndice  $c$  indica constante.

luego

$$(*) = -2 B V^t + 2 W V V^t = O \Rightarrow$$

$$B V^t = W V V^t \Rightarrow$$

$$B V^t = W (V V^t)$$

Así pues, el estimador mínimo cuadrático de  $W$  será

$$W = B V^t (V V^t)^{-1}$$

Sabemos que las filas de  $W$  son las diagonales de las matrices de pesos individuales  $W_r$ . No obstante no puede esperarse que esas  $R$  matrices de pesos conjuntamente bajo la configuración inicial  $X$ , den el estimador mínimo cuadrático de cada matriz  $B_r$ . Así pues, se necesita calcular una configuración  $X^{(1)}$  más apropiada, relativa a los  $W_r$  dados.

Para ello en primer lugar se calculan los productos  $W_r X^t$  para cada  $r$  y se sitúan las  $(R M \times n)$  matrices en una matriz  $H$ ,  $(M \times n R)$ . Se crea igualmente

una matriz  $\dot{B}$  con las  $R$  matrices  $B_r$  de forma idéntica a la anterior. Con esas matrices, partiendo de la ecuación

$$B_r = XW_rX^t$$

se obtiene la ecuación

$$\dot{B} \approx XH$$

Ahora debe encontrarse la matriz  $X$ , ( $n \times M$ ). Para ello, se procede como en el caso anterior. La función de pérdida será ahora

$$L = \text{tr} [(\dot{B} - XH)^t(\dot{B} - XH)]$$

luego

$$\frac{\partial}{\partial X} L = -2\dot{B}H^t + 2XHH^t = 0$$

de donde se obtiene que el estimador mínimo cuadrático de  $X$  es

$$X^{(1)} = \dot{B}H^t(HH^t)^{-1}$$

Con esta nueva matriz  $X^{(1)}$  se determina la matriz  $W^{(1)}$  y después otra matriz  $X^{(2)}$  con la que comenzaremos el proceso nuevamente hasta obtener convergencia y una solución final de  $X$  y  $W$ .

## 2.5 La interpretación final en MDS.

En este apartado trataremos de poner de manifiesto algunos aspectos fundamentales que influyen de manera decisiva en el planteamiento de un problema con MDS y en la posterior interpretación de los resultados obtenidos. Para situar estas conclusiones hay que tener presente en todo momento que MDS es una herramienta de la estadística puramente aplicada, por lo que su desarrollo no tiene sentido al margen de la interpretación de los resultados obtenidos. Así pues y dada la importancia que en MDS tiene la aplicación práctica, consideremos a continuación una serie de aspectos importantes en todo estudio con MDS.

### 2.5.1 La Configuración.

Cuando se efectúa un análisis MDS en su forma más simple, partiendo de la tabla de similitudes o disimilaridades, se obtiene como resultado las coordenadas de los puntos. No obstante, puesto que esas coordenadas se han calculado en base únicamente a que los puntos que representan, mantengan entre sí unas distancias dadas, éstos no serán susceptibles de una interpretación directa, ya que siempre es posible efectuar una rotación o en general un movimiento rígido a la configuración, manteniendo las distancias entre los puntos aunque lógicamente, variando sustancialmente las coordenadas de éstos.

Así pues, siempre debemos tener presente que las soluciones del MDS ordinario o a dos vías (no así en MDS a tres vías o MDS con diferencias individuales), siempre están sujetas a rotaciones por lo que los ejes de coordenadas no tienen una significación especial en la configuración obtenida, sino únicamente la de unas rectas, que como podrían haber sido otras cualesquiera, permiten representar los puntos, sin que tengan ninguna preponderancia frente a otra posible elección como sistema de referencia.

Un aspecto que es muy importante en MDS lo constituyen las direcciones en la configuración obtenida, las cuales tienen unas interpretaciones muy interesantes. Dicho de otro modo, las posiciones en la configuración, pueden asociarse con alguna característica del elemento que es representado. De hecho, estudiar esas características es una de las razones para el empleo del MDS. Una forma práctica de descubrir esas características es simplemente mediante observación de la configuración, recordando qué cosas se conocen sobre los objetos. Sin embargo, este método es limitado por cuanto requiere de un gran esfuerzo memorístico y una gran capacidad de representación mental, así como limita las dimensiones en las que se puede trabajar al no poder imaginar la mente humana más de tres. Por todas estas razones se puede concluir que este método no es muy recomendable.

El método que normalmente se utiliza está basado en la regresión lineal. Supongamos que tenemos una variable asociada al elemento en cuestión, de forma que se sospecha que esa variable tiene alguna relación con la posición de ese ítem en la configuración. Una forma de poder confirmarlo, es mediante regresión lineal múltiple de esa variable sobre las otras variables que constituyen cada coordenada del punto, o sea, si la variable es  $V$  y si suponemos una configuración tridimensional, fijándonos en el  $i$ -ésimo ítem, tendremos por un lado el valor de  $V$  asociado a éste,  $v_i$  y por otro, sus coordenadas en la configuración

$(x_{i1}, x_{i2}, x_{i3})$ . Se tratará pues de ajustar

$$v_i = a + b_1x_{i1} + b_2x_{i2} + b_3x_{i3}$$

El coeficiente de correlación múltiple dará una medida de como de bien se ajusta el plano a la variable. Así pues, la idea de utilizar la regresión para facilitar la interpretabilidad de la configuración consiste en que, una vez representada la nube de puntos en una dimensión adecuada y en un espacio adecuado, de forma que el valor asociado de la función de error haya resultado suficientemente pequeño como para que dicha configuración sea representativa, se consideran una serie de variables  $V$  que puedan suministrar alguna información en cuanto a la posición de los puntos en el espacio en que se representan, ya que como antes afirmábamos, el sistema de referencia en el que éstos se representan, no es significativo en cuanto a la interpretación de los resultados. Al efectuar regresión de esas variables con las coordenadas de los puntos, elegiremos de entre ellas, las que resulten con mayor coeficiente de correlación, representando sus correspondientes rectas de regresión en el espacio de la configuración, e interpretando la posición de los puntos, no respecto al sistema de referencia, sino según la posición de éstos respecto de las rectas representadas.

Este método no solo se usa en el sentido descrito (o sea, una vez obtenida la dimensionalidad del problema, según el valor de la función de error, se representan en el espacio obtenido, tanto los puntos, como las rectas de regresión antes mencionadas), sino que además ayuda en muchas ocasiones a elegir igualmente la dimensionalidad del problema. Así pues, una vez obtenidos los puntos, si al efectuar regresión y calcular los cosenos directores de la recta con respecto a los ejes de coordenadas, resulta que sólo algunos de estos cosenos son suficientemente grandes mientras que el resto no lo son, podríamos interpretar este hecho como que a pesar de que el valor de la función que mide el error (STRESS, SSTRESS, etc. . .) en la dimensión dada es pequeño, no todas las dimensiones resultan susceptibles de ser interpretadas, por lo que se elegiría una nueva dimensionalidad para el problema, de acuerdo con el número de cosenos directores suficientemente grandes, volviendo por tanto a obtener las coordenadas de los puntos en esa nueva dimensión, así como una nueva representación de la recta de regresión en el subespacio elegido.

En definitiva, una vez que se ha deducido lo más razonablemente posible, la dimensionalidad del problema, se pasaría a estudiar las relaciones de los

puntos representados, con respecto a esos nuevos ejes que se han obtenido. Lógicamente, la elección de las variables  $V$  se realizará con arreglo a los factores que interese poner de manifiesto en el problema en cuestión y con arreglo a los cuales queramos interpretar las posiciones de los puntos de la configuración obtenida.

Algunas veces, se puede observar una estructura en el espacio multidimensional, distinta de la suministrada por la interpretación dimensional. Por ejemplo, regiones locales en el espacio, pueden tener un significado asociado con otras características. Esto suele ocurrir cuando, por ejemplo, una solución bidimensional se ha obtenido para datos cuya dimensión apropiada es mayor. Una razón importante por la que una interrelación local puede revelar otras relaciones en los datos, es porque el enfoque se dirige a pequeñas distancias (grandes similitudes), mientras que una aproximación dimensional se centra más en grandes distancias. Se ha demostrado que el análisis de correlaciones en pequeños espacios revela estructuras importantes entre grupos de datos. Así pues, deben utilizarse todas las medidas que se tengan al alcance para el desarrollo del estudio, por lo que la interpretación local, debe usarse como un complemento del análisis dimensional.

Existen también métodos sistemáticos, cuyo objeto se centra en encontrar agrupamientos naturales entre los puntos, basados directamente en la matriz de proximidades. Los grupos pueden dibujarse en el espacio multidimensional como lazos alrededor del punto cuyo estímulo es relevante. Una vez hecho eso, se pueden buscar características comunes a los objetos de un grupo. Una forma más simple de incluir esta información en el espacio multidimensional desde la matriz original de datos, consiste en dibujar una línea entre cada par de objetos cuya proximidad exceda de una cantidad fijada por el investigador. La presencia de líneas largas y cruzadas (basadas en las proximidades), indican una gran resíduo que podría tenerse en cuenta en la interpretación de la configuración. Así, si se trabaja sobre un plano expresado en alguna dimensión mayor, estas discrepancias podrían deberse a otras dimensiones que no se han tenido presentes en el desarrollo del estudio. Por ejemplo, agrupamientos en una configuración bidimensional, no serían probablemente válidos hasta que los puntos de cada grupo no pudiesen conectarse bien entre sí y conectarse mal con los de otro grupo.

### 2.5.2 La dimensión del problema.

No existe en principio ningún método estadístico aceptable que asegure la dimensionalidad en el estudio de unos datos, por lo que en esta sección nos dedicaremos más a la dimensionalidad apropiada para unos datos, que a la correcta. El concepto de dimensionalidad o de número de dimensiones, se refiere al número de ejes coordenados en el que representar los datos o al número de factores en el análisis factorial. La dimensionalidad no es necesariamente el número de características relevantes. De cualquier forma, aunque una característica o atributo tenga un efecto en la configuración, es posible que ésta no se refleje en un nuevo eje, debido a que esté correlada con otra dimensión, o que sólo afecte a un subconjunto de estímulos, etc... Así, si su presencia se puede hacer efectiva por otros medios, no hay por qué mostrarla como una dimensión. El número de direcciones interpretables puede ser menor que la dimensionalidad del espacio. Por ejemplo, el investigador puede interpretar únicamente una o dos dimensiones, incluso cuando los datos parezcan requerir tres.

A la hora de decidir cuantas dimensiones son apropiadas, la bondad del ajuste es sin duda una consideración muy importante. Una medida del ajuste muy generalizada en MDS es como ya hemos indicado el *stress* que está definido como la raíz cuadrada de la suma de los residuos al cuadrado normalizada, de forma que pequeños valores de éste indican un buen ajuste. Cuando se consideran los valores del *stress*, es importante recordar que la interpretación de los valores, desafortunadamente, depende al menos de todas las elecciones posibles o parámetros posibles.

Es necesario tener presente que la interpretación del valor del *stress* depende del número de objetos  $n$  y de la dimensión  $M$ . No obstante, si  $n > 4M$ , aproximadamente, éstos no influyen en la interpretación del *stress*. Cuando  $n$  se acerca a  $M$ , suceden grandes cambios. Por ejemplo, en muchas situaciones, un valor de *stress* de 0.02 indica un buen ajuste. Por otro lado, para 7 objetos y 3 dimensiones, un *stress* de 0.02 o menor ocurre en un 50% de los casos, por lo que éste, no es un tamaño suficientemente pequeño para indicar un buen ajuste. Como base de interpretación del *stress*, veamos cual es el conjunto de condiciones más usuales.

1. Conocimiento y utilización de media matriz de disimilaridades sin diagonal.

2. Ausencia de datos faltantes en las proximidades o presencia de pocos.
3. Ausencia de datos con replicaciones.

El hecho de introducir los elementos de la diagonal de la matriz de proximidades no cambia la configuración, pero aumenta el valor del *stress* sobre todo si las  $d_{ijr}$  son muy diferentes de las  $d_{jir}$ . Además, introducir observaciones que constituyen réplicas de las proximidades incrementa el *stress*, lo cual ocurre de forma excesiva si las observaciones repetidas son muy diferentes unas de otras.

Si se tienen en cuenta éstas junto con otras precauciones, debe suponerse que en cada caso de dimensionalidad, el valor del *stress* es el mínimo posible para cada caso.

Existen a nivel general, dos aproximaciones para la utilización del *stress* como guía de la dimensionalidad: la Estadística y la Intuitiva. La primera es más objetiva aunque las tablas necesarias para su utilización abarcan un rango limitado de valores de  $n$ . Veamoslas detenidamente.

### El método Estadístico.

El método estadístico está basado en la comparación de la gráfica del *stress* (o en general de la función que mide el grado de ajuste), dibujada como puntos unidos por segmentos. Posteriormente, se dibuja una gráfica simulada con los valores reales de la dimensión y el error, mediante el método de Monte Carlo, comparándose para encontrar la dimensión de los datos. Veamos una explicación más detallada del proceso.

La idea subyacente que existe una configuración real de los datos en alguna dimensión, la cual denotaremos por  $M_t$ . Se supone que las proximidades se han generado a partir de las distancias en la configuración exacta pero con errores aleatorios de algún tipo que se incorporan al nivel controlado por un valor  $e$  y posiblemente también con distorsiones monótonas incorporadas. Cuando las proximidades son representadas en diferentes escalas correspondientes a distintas dimensiones, la gráfica de puntos del *stress* frente a  $M$  depende de  $M_t$  y de  $e$ . Así, cada combinación de  $M_t$  y de  $e$  se corresponde a una apariencia particular de la gráfica de puntos. Usando el método de Monte Carlo, se generan muchas configuraciones y se representan en cada escala, de forma que las gráficas de puntos se catalogan con arreglo al par  $(M_t, e)$ , comparando sus resultados con los de los datos e infiriendo el valor de  $M_t$  y de  $e$  para la obtención de una gráfica lo más parecida a los datos que se poseen. Si la

representación de nuestros puntos es muy diferente de las salidas del método de Monte Carlo, podría deberse a una convergencia incompleta o a un mínimo local del *stress*. Igualmente se observa que para grandes valores de  $e$ , las diferencias son prácticamente inapreciables para los puntos representados en diferentes salidas de  $M_t$ , por lo que no es conveniente usar grandes valores de  $e$  para determinar  $M_t$ .

### El método intuitivo.

Este método está basado en la experiencia y en la intuición. Si el valor de  $M_t$  es distinto a 1, el *stress* para  $M = 1$  suele ser grande. Así pues, esto sugiere que si el *stress* para  $M = 1$  está próximo a 0.15, la gráfica suele tener un codo en  $M_t = M$ , a menos que  $M_t = 1$ .

### 2.5.3 Algunas consideraciones complementarias.

Veamos a continuación algunas consideraciones importantes para la elección de la dimensión del problema. Estas consideraciones son producto de la experiencia y deben tenerse muy en cuenta ya que no sólo es importante el mecanismo matemático para la resolución de un problema con MDS sino que la experiencia constituye en la mayoría de las ocasiones el factor determinante para la obtención de soluciones.

### Interpretabilidad.

En MDS es posible que, por ejemplo, en dos dimensiones no tengamos una interpretación de los resultados y sin embargo en tres dimensiones exista una clara interpretación de la solución obtenida. Esto es debido a que la relación entre las diferentes dimensionalidades es más complicada en este método que en otros. Así, en el análisis de componentes principales, por ejemplo, la configuración en  $M - 1$  dimensiones se puede obtener de la de  $M$  sin más que omitir la última coordenada, ocurriendo algo parecido en el análisis factorial, mientras que para MDS la cuestión es más complicada. Omitir la segunda coordenada en una configuración bidimensional, es tanto como proyectar sobre el eje de abscisas los valores obtenidos. En MDS, la configuración unidimensional se puede obtener algunas veces de forma aproximada, proyectando sobre alguna recta en un espacio bidimensional, aunque en otros casos esto no es posible.

Podemos generalizar todo lo expuesto al caso de tres dimensiones y para eliminar la última coordenada de una configuración tridimensional se proyecta ésta sobre el plano formado por los dos primeros ejes, aunque en MDS, a partir de la configuración tridimensional, se puede obtener la bidimensional proyectando ésta sobre *algún* plano del espacio tridimensional y sólo en algunos casos.

Debido a todo esto, la forma en que cambia la interpretación de una dimensión a la siguiente, puede ser complicada en MDS. De cualquier forma, la interpretabilidad juega un papel importante en la elección de una dimensionalidad de entre las propuestas por la bondad del ajuste. Así, si en dos dimensiones existen dos o más direcciones interpretables y en una tercera dimensión no afecta mucho, se elige la bidimensional, aunque el que el investigador no interprete la otra dimensión no quiere decir que ésta no sea interpretable.

Cuando la configuración bidimensional no se parece a la proyección perpendicular del espacio tridimensional, podría ocurrir que ninguna dirección de la configuración bidimensional sea interpretable, mientras que una, dos o tres direcciones del tridimensional si lo sean. Para interpretar una configuración tridimensional, frecuentemente ayuda (si es posible), el descubrir qué direcciones se parecen a la configuración bidimensional, procedimiento que podemos hacer extensivo a más dimensiones. Cuando visualmente las direcciones pueden detectarse, se recurre a procedimientos estadísticos como regresión lineal, correlación canónica o análisis factorial. La regresión múltiple, como antes apuntábamos, puede ayudar a elegir la dimensión adecuada. Por ejemplo, si algunas variables (vectores propios) tienen mayor coeficiente de correlación múltiple en una configuración en dimensión  $M + 1$  que en dimensión  $M$ , será preferible la dimensión mayor.

Independientemente de la dimensionalidad, si la configuración sitúa muy mal los datos, resulta peligroso de usar. Ninguna interpretación aparente es satisfactoria y las tendencias reales no están claras. Usar más dimensiones podría mejorar el ajuste aunque si se usa un gran número de dimensiones, la configuración se adapta al error aleatorio de los datos, lo cual puede hacer difícil el ver los aspectos interesantes. Si estas dos reglas desechan todas las dimensiones, los datos deben tratarse de otra forma y no con MDS.

### La facilidad en el uso de los resultados.

Puesto que trabajar en dos dimensiones es más fácil que con más, la facilidad en el uso de los resultados, es una consideración a tener presente. Por ejemplo, cuando una configuración en MDS es deseada fundamentalmente como fundamento para mostrar resultados de agrupamiento, una configuración bidimensional es preferible a otra. Incluso la proyección bidimensional de una configuración en una dimensión mayor es preferible en muchos casos para este fin. Una configuración en una dimensión elevada, sólo es deseable cuando se van a utilizar otras técnicas complementarias para encontrar estructuras de interés. No es usual examinar una única configuración cuando existen en ella dimensiones que no se comprenden lo suficiente; normalmente, éstas son direcciones en la configuración pero pueden ser planos, curvas y otro tipo de estructuras. Así, aunque parezca que se debe ajustar una configuración en cuatro dimensiones, se suele trabajar también en dos o tres para detectar estos casos. De cualquier forma, se puede comprobar que los aspectos más importantes de elevadas dimensiones aparecen en una configuración bidimensional, aunque para clarificarlos se deba recurrir a dimensiones más altas en la mayoría de los casos.

### Estabilidad y consideraciones relativas.

El MDS así como los datos de los que se deriva, están sujetos a variabilidades aleatorias. Incluso para datos *buenos*, no es fácil encontrar mínimos locales que difieran considerablemente de la mejor solución posible. Una forma simple de evaluar la estabilidad de las dimensiones y de otros aspectos de la configuración es dividir los datos en varias formas y aplicar MDS a cada parte. Por ejemplo, si el número de objetos lo permite, podrían dividirse éstos en grupos aleatorios. Después usaríamos el procedimiento *jackknife* expuesto por Tuckey.

Otro procedimiento consiste en que a partir de los datos se formen otros mediante una razonable alteración de algún efecto aleatorio sobre los originales, o sea, datos afectados de algún proceso de ruido deliberadamente introducido y controlado.

De todo lo expuesto se deduce que una decisión sobre la dimensionalidad de los datos puede estar influida notablemente por la estabilidad de las soluciones.

## Capítulo 3

### Inferencia en MDS.

Hasta ahora, en los dos capítulos anteriores, he planteado la problemática del MDS bajo un punto de vista estrictamente del análisis de datos. Se trata en este tema de abordar e intentar resolver varios problemas que plantean los métodos antes expuestos.

Antes de comenzar a describir los problemas mencionados, hay que hacer nuevamente hincapié en la clasificación del MDS. Cuando he descrito la clasificación de las diferentes técnicas o modelos que engloban el MDS, lo he hecho atendiendo a diversas razones, como tipo de datos manejados, carácter de la función que relaciona las distancias con las proximidades de partida e incluso con arreglo a la función que mide la bondad del ajuste estimado. No obstante, en esta ocasión nos referiremos a un punto de vista del MDS que atiende a la intención de la técnica en sí, más que a la forma del modelo o al tipo de datos empleados.

Todos los métodos antes descritos, constituyen métodos que podrían englobarse dentro de las técnicas *exploratorias* del MDS o sea, como una primera aproximación a la resolución del problema, sobre todo para la determinación de la dimensión, obtención de la configuración asociada e interpretación de soluciones. No obstante, un segundo aspecto a tratar en todo problema de MDS estriba en la necesidad de técnicas denominadas *confirmatorias*. En este apartado englobamos el trabajo que presentamos a continuación y en el que se pretende ofrecer un método para el contraste de resultados obtenidos con cualquiera de las técnicas métricas del MDS.

En general, la distinción expuesta anteriormente, puede hacerse extensiva a todo el análisis de datos, o sea, entre aquellos análisis que son primordialmente

exploratorios y aquellos que intencionadamente son confirmatorios. El primero de dichos análisis está diseñado para permitir la observación de datos bajo una gran variedad de procedimientos, con la intención de descubrir diversas relaciones de interés para el investigador. Las técnicas exploratorias deben ser tan flexibles como sea posible y tolerantes con el mayor abanico posible de tipos de datos. El énfasis fundamental de estos análisis, no estriba tanto en las decisiones en torno a las relaciones específicas o aspectos intrínsecos de los datos, como en la exposición clara de los sucesos. En relación a esto, podemos decir que ALSCAL e INDSCAL fueron diseñados para actuar, en primer lugar, como técnicas exploratorias, aunque se hayan empleado frecuentemente para propósitos relativamente confirmatorios.

El análisis confirmatorio, se encuentra enfocado en relación a aspectos específicos de los datos y a cuestiones particulares en torno a éstos. Supone que el investigador posee una idea clara del modelo a emplear, del desarrollo de los datos y de la estructura que pretende investigar. El objetivo de esta técnica, consiste en tomar decisiones referentes a algunas hipótesis. Para ilustrar con un ejemplo, digamos que la hipótesis más utilizada en MDS, hace referencia a cual es la dimensionalidad más apropiada para la representación y estudio de los datos. Otras hipótesis relativamente frecuentes las constituyen, el estudio de si datos provenientes de dos o más grupos de individuos, pueden representarse bajo la misma configuración, si un modelo ponderado es mejor que un modelo no ponderado o si una configuración con restricciones representa a los datos tan bien como lo haría una configuración no restringida.

Puesto que todas esas cuestiones se plantean de forma precisa, requiriendo además el investigador una respuesta tan clara y concisa como sea posible, la precisión de la estimación en el análisis confirmatorio adquiere una importancia fundamental.

El trabajo que se presenta a continuación, está basado en el modelo desarrollado por Ramsay y ofrece una versión mejorada de éste, del cual puede decirse, en líneas generales, que está diseñado para el estudio de una amplia familia de datos, como es la constituida por los datos que se obtienen a través de las denominadas técnicas de clasificación directa mediante escalas, estudiando dichos valores bajo el punto de vista del análisis de datos confirmatorio. Además el modelo que presentamos está diseñado, por un lado, para el tratamiento de una gama más general de datos que incluye los analizados por Ramsay así como datos provenientes de técnicas que relajan las condiciones de positividad, entre otras, como por ejemplo, los datos que provienen de *disparidades*

o sea, de valores obtenidos a través de transformaciones monótonas iniciales, estimadas a partir de los datos originales. Por otro lado, está igualmente diseñado de forma que permite complementar los modelos antes expuestos, o sea, modelos como ALSCAL e INDSCAL, en lugar de ofrecer exclusivamente un modelo alternativo. Lo realmente razonable sería utilizar el modelo probabilístico, no como modelo para la determinación de la dimensión del espacio de configuración y de la matriz de coordenadas, sino como una segunda etapa de contraste.

### 3.1 Desarrollo estadístico del MDS.

El análisis de disimilaridades provenientes de un único sujeto o de varios sujetos se ha convertido en un problema lo suficientemente importante como para que sea estudiado estadísticamente. Las primeras técnicas de MDS, basadas en el análisis propio de la matriz de productos escalares tal y como en los capítulos anteriores exponíamos, resultan insuficientes. Esta carencia es debida, posiblemente, a que los productos escalares están basados en los cuadrados de las observaciones originales e inevitablemente contienen mayores errores típicos. Estos modelos emplean igualmente modelos de error que resultan irreales, ya que el uso de un simple criterio de mínimos cuadrados es equivalente a la hipótesis estadística de que los productos escalares son independientes y están normalmente distribuidos en torno a los valores reales.

Por otro lado, dentro de las técnicas de MDS tanto métrico como no métrico, hay que añadir, que el MDS no métrico sufre en general del grado de libertad usado al ajustar la transformación monótona supergeneral a la variable dependiente, tal y como se describía en el capítulo anterior. Esas transformaciones monótonas, ignoran consideraciones como la suavidad y el hecho de que al menos ciertos tipos de datos requieran de transformaciones muy especializadas. Como consecuencia de todo esto, las técnicas no métricas favorecen la inestabilidad cuando son aplicadas sobre matrices de datos que tienen un orden relativamente pequeño respecto al número de dimensiones elegido en la configuración ajustada o cuando se aplican a matrices de datos que tienen un error medible como sucede en los problemas en que interviene un único sujeto. De cualquier forma, cualquier análisis serio debe manejar la posibilidad de que una determinada transformación de los datos mejore el ajuste estimado.

Cuando se realiza un análisis de datos, sea cual sea la técnica particular

empleada en la investigación, aparecen implícitamente diversas estructuras enfocadas a describir la forma en que las observaciones varían en torno a sus valores ajustados, lo cual, generalmente, ofrece un punto de vista acerca de la bondad de la técnica empleada con arreglo a los datos estudiados. Además, las distribuciones de estos errores o residuos son consideradas generalmente de forma implícita en el modelo, por lo que resulta bastante complejo su manejo, así como la interpretación de la influencia que sus fluctuaciones ejercen en el total del modelo. Por tanto, si se desea realizar una investigación mediante un modelo de análisis de datos cualquiera, es necesario dar una expresión explícita del modelo de error que interviene y estructurar la posibilidad de comparación entre diversos modelos de error, bien dentro de un mismo estudio o bien entre estudios diferentes.

En el análisis de datos, puede decirse que aflora una carencia grave desde el punto de vista técnico y que arrastran en general la totalidad de las técnicas. Esta carencia consiste en la necesidad de poder contrastar diversos elementos estructurales de los modelos, teniendo en cuenta la imposibilidad del manejo de una distribución sobre los datos de partida. Dicha carencia se viene resolviendo a través del empleo de técnicas de simulación como Jackknife y Bootstrap, estando orientada su utilización más hacia la validación de un modelo que a la inferencia sobre diversos aspectos del mismo.

En el trabajo que presentamos, se emplea en principio, exclusivamente la estimación por máxima verosimilitud, la cual ofrece la posibilidad de efectuar la confrontación a que antes nos referíamos del modelo de error, además de las ventajas intrínsecas a ella, como el uso de las técnicas del análisis de residuos para la realización de una posible revisión sobre el modelo de error empleado. Además, la estimación máximo verosímil ofrece, aunque asintóticamente, una función de bondad de ajuste y un criterio de decisión para cuestiones como dimensionalidad, etc...

Las técnicas de máxima verosimilitud empleadas en este trabajo, tienen la suficiente eficiencia para ser útiles, tanto en el caso en que se tiene un único sujeto y un número modesto de estímulos, como cuando se estudian simultáneamente opiniones ofrecidas por más de un sujeto, supuesto que se ha elegido un determinado modelo de error. También consideraremos el problema del ajuste de una transformación a los datos. Todos los procedimientos que se desarrollan en este trabajo pretenden ofrecer una solución al problema de la ineficiencia en el análisis de productos escalares y el problema del vasto número de grados de libertad en aproximaciones no métricas.

### 3.2 MDS sin diferencias individuales.

En primer lugar, vamos a concretar la notación que se va a seguir a lo largo de todo el trabajo, comenzando por la descripción que utilizaremos en este primer apartado. Para fijar ideas, supongamos en principio, que a  $R$  individuos o sujetos les pedimos que muestren su opinión en torno a la semejanza o mejor dicho a la desemejanza entre cada una de las parejas que se hayan estudiado de un total de  $n$  estímulos u objetos, a través de algún procedimiento de escala, previamente establecido. Así pues, se dispone de  $R$  matrices  $n \times n$  de *disimilaridades* entre cada pareja de estímulos, cuyos elementos notaremos por  $d_{ijr}$ , con  $i, j = 1, \dots, n$  y con  $r = 1, \dots, R$ . El problema del MDS consiste fundamentalmente en la búsqueda de  $n$  puntos, cuyas coordenadas notaremos por  $x_{im}$  y que representarán a los  $n$  estímulos en un espacio común de dimensión  $K$ . Se trata pues de encontrar dichos puntos de forma que la distancia entre el punto  $i$  y el  $j$ , que notaremos  $d_{ij}^*$ , sea lo más parecida posible al valor esperado de las correspondientes disimilaridades observadas  $d_{ijr}$ .

Bajo un punto de vista estadístico, podemos formular este primer modelo denominado *two-way MDS* o *MDS a dos vías* y que fue diseñado para el estudio de datos de disimilaridad provenientes de un único sujeto, por lo que también puede considerarse el índice  $r$  en este caso, como indicativo de replicación, de la siguiente forma: Consideremos la variable aleatoria  $D_{ij}$ , que toma valores  $d_{ijr}$ , indicando disimilaridad entre el estímulo  $i$  y el estímulo  $j$  apreciada por el sujeto  $r$  (resp. réplica  $r$ ). Así pues, podemos considerar a  $d_{ij}^*$  como la distancia exacta, o sea, teórica y sin error y a  $d_{ijr}$  como la  $r$ -ésima observación de esa distancia o bien como la  $r$ -ésima replicación de la distancia observada a través de la variable aleatoria  $D_{ij}$ . Supondremos que las observaciones  $d_{ijr}$  son independientes e idénticamente distribuidas según una función de densidad que notaremos por  $f$ . Se trata ahora de ver cual es la elección de  $f$  más adecuada al problema que hemos planteado.

#### 3.2.1 Modelización de la variabilidad residual.

Para considerar una elección conveniente de la densidad asociada a la variable  $D_{ij}$ , Ramsay [1977] argumenta que deben tenerse en cuenta los siguientes aspectos:

- La consistencia del dominio con los datos tratados.

- El hecho de que la varianza posea una estructura adecuada a la de los datos estudiados en el modelo.
- La coherencia de los principales momentos de la densidad respecto a lo que empíricamente muestran los datos.

El procedimiento más simple para evaluar la densidad es por tanto, evaluar el dominio y evaluar sus primeros momentos. En el tipo de datos analizados por Ramsay, constituidos por valores obtenidos a través de técnicas de clasificación directa, las observaciones, en términos de distancias, serán no negativas (este punto será ampliado posteriormente, ya que no refleja aspectos como los mencionados en Heiser[1991]). Hay que hacer especial hincapié en que esto no ocurría así en los primeros trabajos de MDS, en los que se aplicaban técnicas univariantes, las cuales obtenían valores de escala con orígenes arbitrarios, por lo que automáticamente, surge un primer motivo para argumentar la necesidad de extensión del modelo, bajo un punto de vista de localización. En el trabajo que presentamos, comentaremos en primer lugar las técnicas de MDS para el análisis de disimilaridades obtenidas mediante procedimientos de clasificación directa, extendiéndolas y perfeccionándolas paso a paso, hasta llegar finalmente al modelo general propuesto. Así pues, tal y como se comenta en el trabajo de Ramsay[1977], los datos así obtenidos tienden a ser:

- No negativos (Será generalizado posteriormente a datos que si pueden serlo).
- Tienen un origen definido en el caso de las disimilaridades (también será extendido al caso de datos de disimilaridad con origen indefinido y por tanto, estimado) o bien son transformables tomando simétricos de forma que resulten no negativos y posean un origen definido, en el caso de juicios de similaridad.

En vista de todo lo que acabamos de exponer, podemos resumir diciendo que las características determinantes para la elección de la función de densidad  $f$  que represente la distribución de las disimilaridades, obtenidas según el procedimiento general son:

1. **El Dominio.** La densidad  $f$  debe estar definida sobre valores reales y no negativos, dadas las características de las disimilaridades en el caso de los

datos analizados por Ramsay, siendo simplemente reales, sin someterlos a ninguna restricción, para la familia general de datos de tipo intervalo y escala analizados en este trabajo.

2. **La Media.** La media de cada variable  $D_{ij}$  debería estar próxima al valor deseado  $d_{ij}^*$ , aunque como argumenta Ramsay [1969] no es necesario que sean exactamente iguales. Por ejemplo, cuando se utilizan técnicas de clasificación directa con escalas positivas, en muchas ocasiones, se obtienen medias positivas incluso cuando el valor de la distancia es cero, por estar evaluando objetos idénticos. Esta conclusión es debida, fundamentalmente, a que los individuos se muestran normalmente reacios a clasificar a los estímulos con disimilaridad cero, incluso cuando el estímulo es el mismo. Este hecho corresponde a la ley psicológica de Weber. Debido a esta razón, verificada empíricamente y en consonancia con dicha ley, en este tipo de datos de disimilaridad positiva, parece lógico no considerar un dominio de disimilaridad con origen en cero, al no ocurrir en ningún caso valores por debajo de un valor real, mayor estrictamente que cero, pero indeterminado. Por tanto, considerar dominios con orígenes cero, contribuye en muchos casos, a aumentar la dispersión en el análisis y por tanto, desvirtúa las conclusiones finales del análisis.
3. **La Varianza.** El desarrollo de la varianza de  $D_{ij}$  resulta especialmente importante, como puede verse en los trabajos de Ramsay [1977], [1978] y [1982]. La cuestión más crítica concierne a la relación entre la desviación típica de  $D_{ij}$  y la distancia teórica  $d_{ij}^*$ . De hecho, pueden seguirse varios caminos al respecto como considerarla constante o bien proporcional a la distancia teórica  $d_{ij}^*$ . Suponer la varianza constante puede resultar equívoco en varios sentidos. En primer lugar, si la media de  $D_{ij}$  vale cero cuando  $d_{ij}^*$  vale cero, entonces todas las desviaciones típicas valdrán cero al ser la varianza positiva. Incluso cuando éste no es el caso, solo observamos pequeñas variaciones en las disimilaridades cuando los objetos son bastante diferentes. Por otro lado, cuando existe poca variación en  $d_{ij}^*$ , se comete muy poco error al suponer homoscedasticidad. Tampoco se comete mucho error suponiendo heteroscedasticidad ya que todo lo referente a la relación entre la desviación típica y la media se favorece de forma irrelevante cuando la media tiene poca variación. Por otro lado, el caso de la desviación típica proporcional a la media es un caso práctico

muy fuerte en el que la razón de proporcionalidad es invariante frente a cambios de unidad de medida, pudiendo afirmarse que en este caso las disimilaridades obedecen la ley de Weber del continuo psicológico.

Una vez observados estos aspectos, Ramsay [1977] afirma que la ley lognormal es la adecuada para describir el comportamiento de las disimilaridades obtenidas a través de técnicas de clasificación directa de escalas positivas. Así pues, bajo la hipótesis de varianzas constantes, éste propone que

$$D_{ij} \sim \Lambda(\log(d_{ij}^*), \sigma^2)$$

luego

$$f(d_{ijr} | d_{ij}^*, \sigma^2) = \frac{1}{\sigma d_{ijr} \sqrt{2\pi}} \exp \left\{ \frac{-\log^2 \left( \frac{d_{ijr}}{d_{ij}^*} \right)}{2\sigma^2} \right\} \quad d_{ijr} > 0$$

El motivo que induce esa relación entre distancias y disimilaridades puede entenderse de la siguiente forma:

La distribución lognormal asocia probabilidades positivas a todos los posibles valores positivos de  $d_{ij}^*$ , tiene media casi igual a  $d_{ij}^*$  y su coeficiente de asimetría es positivo. Varias propuestas alternativas a la de la distribución lognormal son consideradas por Ramsay [1977] y [1982], como es el caso de la distribución normal  $N(d_{ij}^*, \sigma d_{ij}^*)$ , aunque esta propuesta implica el hecho de que se le asigne cierta probabilidad a valores negativos de las disimilaridades, lo cual puede considerarse como erróneo en el tipo de datos analizados por su modelo y de ninguna manera semejante al tratamiento que se ofrece en nuestro trabajo. No obstante, la principal distinción entre esta distribución y la lognormal estriba en torno a la consideración de simetría en los datos de disimilaridad, ya que está comprobado experimentalmente que a pesar de asociar distancias a disimilaridades, estas últimas no verifican por regla general la propiedad reflexiva. Se ofrecen otros refinamientos en cuanto a la distribución inicial como podría ser considerar la varianza descompuesta en componentes referentes a los sujetos o sea determinar la variación en  $d_{ijr}$  mediante  $\sigma_i^2 + \sigma_j^2$ , etc... Posteriormente desarrollaremos aun más las alternativas en cuanto a la distribución inicial de las disimilaridades o distribución de los residuos, cuando se describa el MDS con diferencias individuales.

Se trata ahora de ampliar la posibilidad de análisis, en el contexto del MDS probabilístico métrico, a datos de disimilaridad obtenidos a partir de procedimientos de escala e intervalo generales, que pueden violar la restricción de no negatividad impuesta por el modelo de Ramsay, o incluso, en el caso en que esto no suceda, formular un modelo que permita reducir la elevada variabilidad, que inevitablemente se produce cuando se utilizan métodos cuyo dominio excede innecesariamente el rango real de los datos estudiados. Así pues, para completar la elección de la distribución inicial de los datos, hay que introducir en el proceso un parámetro de localización que permita la inclusión de aquellos datos de disimilaridad que pueden suministrar valores no positivos, así como, tanto orígenes definidos pero no conocidos, como orígenes desconocidos y arbitrarios.

Tal y como en apartados anteriores se comenta, los diseños generales de obtención de datos de disimilaridad, no tienen por qué estar limitados a la obtención de datos no negativos. Puede ocurrir que por necesidades del estudio en particular, la obtención de las disimilaridades deba realizarse sin restricciones, para que sea posible reflejar algunas características determinadas del experimento que resultan de especial interés para el investigador. Por tanto, parece lógico que, en ocasiones, surja la necesidad de poder reflejar explícitamente valores de escala negativos que permitan expresar al sujeto su opinión, por ejemplo, ante un juicio de disimilaridad muy desfavorable o viceversa, en cuyo caso resulta necesario, para poder abordar este tipo de datos con las técnicas probabilísticas métricas existentes hasta ahora, realizar una transformación de los valores, previa y por tanto no controlada e independiente del análisis, para conseguir datos de disimilaridad no negativos. No obstante, si éste fuese el caso, el problema no plantearía graves inconvenientes hasta el punto de la necesidad de un nuevo modelo. Bastaría con incorporar una traslación en la variable en base al valor mínimo de la escala y analizar los datos resultantes aunque se perdiese la significación de los mismos. Veamos un caso interesante que plantea el problema y que era apuntado al principio de esta sección.

En muchos casos de MDS métrico, los datos son obtenidos o recogidos mediante una escala de disimilaridad con origen arbitrario (una escala de tipo intervalo). El hecho de que los cambios en la escala de disimilaridad puedan influir fuertemente en la solución obtenida e incluso en la dimensionalidad estimada de los puntos de la configuración, resulta un inconveniente en la teoría de MDS denominado *el problema de la constante aditiva*. Una descripción

detallada puede verse en Heiser[1991]. En estos casos aparece de forma más evidente la necesidad de introducir un parámetro de localización que permita analizar datos negativos y que, de producirse una transformación, ésta vaya incorporada al análisis en sí de los datos y no como un aspecto aislado y que no influye en el proceso. Para modelizar esta problemática, hay que seguir pensando en que lógicamente asociamos cantidades no negativas a las distancias teóricas  $d_{ij}^*$  (o  $d_{ijr}^*$ , como después veremos) por lo que no puede plantearse un modelo que asocie directamente un valor de disimilaridad negativo a un valor de distancia con el cual se va a representar. Es necesario ajustar una transformación que permita, por un lado, asociar valores de disimilaridad positivos a distancias y por otro, estimar la transformación necesaria a lo largo del análisis y no como elemento adicional.

Por último, aun en el caso de que no se planteen ninguno de los problemas anteriores sino que tengamos valores de disimilaridad no negativos, parece lógico, en virtud de la ya citada ley de Weber del continuo psicológico, considerada para la elección de la distribución inicial de las disimilaridades por la mayoría de los procedimientos de MDS, el hecho de estimar un origen para los datos de disimilaridad es un elemento fundamental para la reducción de la variabilidad en cuanto a la matriz de disimilaridad individual, teniendo en cuenta que puede considerarse como *imposible* en virtud de dicha ley obtener el valor cero para disimilaridades no negativas, incluso cuando se presentan al individuo estímulos que son idénticos.

Así pues, debido a que las demás características necesarias para la consideración de una distribución inicial de las disimilaridades permanecen inalteradas bajo este nuevo punto de vista, proponemos a continuación un nuevo modelo de distribución de la variabilidad residual, que puede considerarse una extensión del modelo expuesto anteriormente ya que bajo las mismas hipótesis, el modelo anterior resulta un caso particular del que se presenta y que viene dado, en una primera aproximación, bajo el punto de vista de la distribución *lognormal triparamétrica* (aunque después serán añadidos diversos aspectos adicionales), la cual conserva las propiedades esenciales descritas anteriormente y además asegura la posibilidad de relajación en cuanto a las restricciones *extremadamente severas* del dominio de la variable aleatoria a estudiar. Así, bajo la hipótesis de varianzas constantes, el modelo de distribución es

$$D_{ij} \sim \Lambda(\theta_{ij}, \log(d_{ij}^*), \sigma^2)$$

Hay que añadir finalmente que se ha elegido el modelo con parámetro de localización  $\theta_{ij}$ , es decir, dependiente de cada pareja de estímulos observados  $(i, j)$  (para MDS a tres vías dependerá de cada tripleta  $(i, j, r)$ ). Los motivos son múltiples: Por un lado, ofrece una mayor generalidad para que el investigador decida en cada caso, si desea ofrecer un origen de escala común a todos los datos de disimilaridad o bien, pretende restringir la variabilidad al máximo para las observaciones asociadas con cada pareja  $(i, j)$ . Por tanto, si lo que desea es estimar una cota inferior común para una escala de obtención de disimilaridades con posibles valores negativos, puede elegir entre las diferentes opciones de tomar una cota inferior común a todos los pares, sin más que tomar  $\theta_{ij}$  constante o bien una cota particular asociada a cada par. Posteriormente en los modelos con diferencias individuales pondremos de manifiesto el interés que puede despertar el que una vez hayan sido estimados los valores de cada  $\theta_{ijr}$ , elegir como origen de escala un valor  $\theta_r = \min_{ij} \{\theta_{ijr}\}$ . Finalmente, asociado con el último caso, hay que decir que en principio parece razonable considerar valores del parámetro de localización dependientes de cada par, ya que el modelo estadístico se formula distinguiendo una variable para cada par, puesto que lo que se modeliza realmente es la distribución de datos en torno a un supuesto valor central teórico-óptimo y que depende de cada par  $(i, j)$ .

Así pues, bajo la hipótesis anterior, se tiene una densidad

$$f(d_{ijr} | \theta_{ij}, d_{ij}^*, \sigma^2) = \frac{1}{\sigma(d_{ijr} - \theta_{ij})\sqrt{2\pi}} \exp \left\{ \frac{-\log^2 \left( \frac{d_{ijr} - \theta_{ij}}{d_{ij}^*} \right)}{2\sigma^2} \right\} \quad \forall d_{ijr} > \theta_{ij}$$

Podemos concluir al igual que lo hacíamos con la biparamétrica, diciendo que la distribución lognormal triparamétrica:

- Asocia probabilidades positivas a todos los posibles valores de  $D_{ij}$  mayores que  $\theta_{ij}$  o lo que es lo mismo, asocia probabilidades positivas a todos los posibles valores positivos  $(d_{ijr} - \theta_{ij})$  de  $d_{ij}^*$ , tomando como variable aleatoria en esta ocasión a  $(D_{ij} - \theta_{ij})$ .
- Tiene una esperanza casi igual a  $d_{ij}^*$ , salvo una transformación lineal que asegura la positividad, o lo que es equivalente, la variable  $(D_{ij} - \theta_{ij})$

posee una esperanza casi igual a  $d_{ij}^*$ , dada por

$$E[D_{ij}] = \theta_{ij} + d_{ij}^* e^{\frac{\sigma^2}{2}}$$

- Su coeficiente de asimetría es positivo, por lo cual es bastante lógica de considerar ya que la distribución de las disimilaridades, debe reflejar una asimetría patente en la mayoría de los modelos de obtención de disimilaridades y que es bien conocida en ambientes psicológicos. Esa asimetría es puesta de manifiesto igualmente en la matriz de disimilaridades y en el desarrollo matemático del modelo.
- Sigue manteniendo una desviación típica proporcional a la distancia teórica deseada sin depender del parámetro de traslación  $\theta_{ij}$ , ya que

$$\text{Var}[D_{ij}] = (d_{ij}^*)^2 [e^{\sigma^2} (e^{\sigma^2} - 1)]$$

Este modelo que acabamos de proponer será igualmente extendido al caso en que esten presentes diferencias individuales, en el cual habrá que efectuar alguna consideración adicional sobre el parámetro de acotación de la localización. Igualmente serán propuestos bajo la hipótesis de lognormalidad triparamétrica, diferentes variantes en cuanto a la modelización de la varianza, para reducir la dispersión en los datos de disimilaridad, tanto como sea posible, sin el empleo de una transformación adicional de los mismos, problema que resulta de gran interés sobre los experimentos de perfil psicológico y que es debido a la naturaleza intrínseca de los individuos encuestados, por lo que no resulta un aspecto controlable a priori y que conlleva que el diseño de procedimientos de disminución de dicha dispersión, constituya un factor fundamental en MDS y en general en la mayoría de las técnicas de análisis de datos, siendo éste el motivo por el que posteriormente se introducen diversas transformaciones, que contribuyen a la resolución del problema.

### 3.2.2 Estimación en el modelo Euclídeo homocedástico biparamétrico.

En esta sección vamos a desarrollar matemáticamente el primer modelo estadístico que se plantea a raíz de las conclusiones anteriores, bajo la ley lognormal biparamétrica. El procedimiento de estimación considerado será el de Máxima Verosimilitud. Sea pues el modelo

$$D_{ij} \sim \Lambda(\log(d_{ij}^*), \sigma^2), \quad \forall i, j = 1, \dots, n.$$

donde

$$d_{ij}^* = \sqrt{\sum_{m=1}^K (x_{im} - x_{jm})^2}$$

siendo

$K$  = dimensión del espacio de la configuración.

Bajo esta situación la verosimilitud adopta la forma

$$L_{\{d_{ijr}\}}(\sigma^2, X) = \prod_i \prod_j \prod_r \frac{1}{\sigma d_{ijr} \sqrt{2\pi}} \exp \left\{ \frac{-\log^2 \left( \frac{d_{ijr}}{d_{ij}^*} \right)}{2\sigma^2} \right\}$$

donde los productos se supondrán lógicamente entre los datos realmente observados. Así pues la log-verosimilitud será

$$\log L_{\{d_{ijr}\}}(\sigma^2, \{x_{pm}\}_{pm}) = -\frac{1}{2} \left( \frac{S}{\sigma^2} + M \log(\sigma^2) \right) - \sum_i \sum_j \sum_r \log(d_{ijr}) - M \log(\sqrt{2\pi})$$

donde

$$S = \sum_i \sum_j \sum_r \log^2 \left( \frac{d_{ijr}}{d_{ij}^*} \right)$$

y  $M$  es el número de datos observados.

Veamos a continuación los estimadores de los parámetros del modelo. En este modelo aparece un estadístico suficiente para el parámetro de dispersión. Este hecho no solo simplifica el problema de la estimación de  $\sigma^2$  sino que también implica que la estimación de la matriz de configuración  $X$  puede efectuarse independientemente de la de  $\sigma^2$ .

**Proposición 3.2.1** *El estimador por máxima verosimilitud de la varianza asociada a los residuos, en el modelo Euclídeo, homoscedástico, lognormal bi-paramétrico, es:*

$$\hat{\sigma}^2 = \frac{S}{M}$$

donde

$$S = \sum_i \sum_j \sum_r \log^2 \left( \frac{d_{ijr}}{d_{ij}^*} \right)$$

y  $M$  es el número de datos observados.

Otros estimadores de  $\sigma^2$  tales como un estimador insesgado o un estimador Bayesiano pueden verse en Ramsay[1977]. Puesto que en este trabajo nos decidimos por la estimación máximo verosímil, nos quedaremos con el correspondiente estimador de los parámetros sin considerar en general otros procedimientos de estimación.

Se trata pues de calcular el estimador por máxima verosimilitud de la matriz de configuración  $X$ , lo cual será efectuado, derivando en la función de log-verosimilitud. La estimación de dicha matriz de configuración, que vendrá dada en una dimensión previamente predeterminada, se efectúa a través de las componentes de la función de distancia  $d_{ij}^*$ . Hay que señalar finalmente, que puesto que  $S$  es un estadístico suficiente para  $\sigma^2$ , dada la matriz de configuración  $X$ , basta con sustituir  $\sigma^2$  por  $\hat{\sigma}^2$  en la función de log-verosimilitud y maximizar con respecto a  $X$ , para obtener un estimador global de la matriz de configuración. Veamos en primer lugar un resultado necesario para lo que sigue.

**Lema 3.2.2** *En la situación anterior, consideremos  $x_i$  y  $x_j$  dos puntos representantes de los estímulos  $i$  y  $j$  y supongamos que la configuración se encuentra determinada en dimensión  $K$ . Así pues será*

$$x_i = (x_{i1}, \dots, x_{im}, \dots, x_{iK})$$

$$x_j = (x_{j1}, \dots, x_{jm}, \dots, x_{jK})$$

Consideremos la distancia euclídea entre esos dos puntos que notamos por

$$d_{ij}^* = \sqrt{\sum_{m=1}^K (x_{im} - x_{jm})^2}$$

Entonces:

$$\frac{\partial d_{ij}^*}{\partial x_{pq}} = 0, \quad \text{si } p \neq i, j$$

$$\text{Si } p = i, \quad \frac{\partial d_{pj}^*}{\partial x_{pq}} = \frac{(x_{pq} - x_{jq})}{d_{pj}^*}$$

$$\text{Si } p = j, \quad \frac{\partial d_{ip}^*}{\partial x_{pq}} = \frac{(x_{pq} - x_{iq})}{d_{ip}^*}$$

Veamos ahora el cálculo del E.M.V. para la matriz de configuración  $X$ .

**Teorema 3.2.3** *El estimador por máxima verosimilitud de la matriz de configuración  $X$ , en el modelo euclídeo, homocedástico, lognormal triparamétrico, vendrá dado por el sistema de ecuaciones:*

$$x_{pq} \sum_j t_{pj} = \sum_j x_{jq} t_{pj}$$

$$\forall p = 1, \dots, n. \quad \text{y} \quad \forall q = 1, \dots, K.$$

donde, para cada  $p$  y cada  $j$ ,

$$t_{pj} = \frac{1}{(d_{pj}^*)^2} \sum_r \left[ \log \left( \frac{d_{pjr}}{d_{pj}^*} \right) + \log \left( \frac{d_{jpr}}{d_{pj}^*} \right) \right]$$

Este resultado puede verse comentado en Ramsay[1977], aunque no ha sido probado explícitamente en dicho trabajo. La demostración del mismo no se expondrá detalladamente en este apartado, ya que puede obtenerse como caso particular de un resultado paralelo que será demostrado posteriormente para el caso general del modelo euclídeo, homocedástico, lognormal triparamétrico. En nuestro trabajo se llega a una conclusión equivalente a la obtenida por Ramsay[1977], pero mejorada, por un lado bajo el punto de vista computacional y por otro en el sentido de que las  $t_{pj}$ , no dependerán del estimador de la varianza, como ocurre con la solución que se ofrece en el trabajo de Ramsay[1977]. En dicho trabajo se afirma que la expresión que adoptan las  $t_{pj}$  viene dada por

$$t_{pj} = \sigma^2 \frac{1}{d_{pj}^*} \frac{\partial}{\partial d_{pj}^*} \sum_r \left[ \log \left( f(d_{pjr} \mid d_{pj}^*, \sigma^2) \right) + \log \left( f(d_{jpr} \mid d_{pj}^*, \sigma^2) \right) \right]$$

Veamos que ambas expresiones de las  $t_{pj}$  coinciden,  $\forall p, j$ .  
Tenemos que

$$D_{ij} \sim \Lambda(\log(d_{ij}^*), \sigma^2)$$

luego

$$f(d_{ijr} | d_{ij}^*, \sigma^2) = \frac{1}{\sigma d_{ijr} \sqrt{2\pi}} \exp \left\{ \frac{-\log^2 \left( \frac{d_{ijr}}{d_{ij}^*} \right)}{2\sigma^2} \right\} \quad \forall d_{ijr} > 0$$

de donde, tomando logaritmos, se obtiene que

$$\log f(d_{ijr}) = -\log(\sigma) - \log(d_{ijr}) - \log(\sqrt{2\pi}) - \frac{1}{2\sigma^2} \log^2 \left( \frac{d_{ijr}}{d_{ij}^*} \right)$$

luego

$$\log f(d_{pjr} | d_{pj}^*, \sigma^2) = -\frac{1}{2\sigma^2} \log^2 \left( \frac{d_{pjr}}{d_{pj}^*} \right) - \log(\sigma) - \log(d_{pjr}) - \log(\sqrt{2\pi})$$

derivando respecto a  $d_{pj}^*$  se tiene

$$\begin{aligned} \frac{\partial}{\partial d_{pj}^*} \log f(d_{pjr} | d_{pj}^*, \sigma^2) &= -\frac{1}{2\sigma^2} \frac{\partial}{\partial d_{pj}^*} \log^2 \left( \frac{d_{pjr}}{d_{pj}^*} \right) = \\ &= \frac{1}{\sigma^2} \frac{1}{d_{pj}^*} \log \left( \frac{d_{pjr}}{d_{pj}^*} \right) \end{aligned}$$

Por otro lado,

$$\frac{\partial}{\partial d_{pj}^*} \log f(d_{jpr} | d_{pj}^*, \sigma^2) = \frac{1}{\sigma^2} \frac{1}{d_{pj}^*} \log \left( \frac{d_{jpr}}{d_{pj}^*} \right)$$

Así pues, se concluye que

$$\sigma^2 \frac{1}{d_{pj}^*} \frac{\partial}{\partial d_{pj}^*} \sum_r \left[ \log \left( f(d_{pjr} | d_{pj}^*, \sigma^2) \right) + \log \left( f(d_{jpr} | d_{pj}^*, \sigma^2) \right) \right] =$$

$$\begin{aligned}
&= \sigma^2 \frac{1}{d_{pj}^*} \sum_r \left[ \frac{1}{\sigma^2 d_{pj}^*} \log \left( \frac{d_{pjr}}{d_{pj}^*} \right) + \frac{1}{\sigma^2 d_{pj}^*} \log \left( \frac{d_{jpr}}{d_{pj}^*} \right) \right] = \\
&= \frac{1}{(d_{pj}^*)^2} \sum_r \left[ \log \left( \frac{d_{pjr}}{d_{pj}^*} \right) + \log \left( \frac{d_{jpr}}{d_{pj}^*} \right) \right] = t_{pj}
\end{aligned}$$

Por tanto, queda probado que ambas expresiones son equivalentes, siendo la expresión que se ofrece en este trabajo una alternativa analítica de mayor consideración. Además está favorecida, tal y como antes mencionábamos, por el hecho de la independencia del parámetro que determina la varianza del modelo, lo que ofrece la posibilidad de estimación disjunta de ambos tipos de parámetros.

### 3.2.3 Estimación en el modelo Euclídeo homocedástico triparamétrico.

Se trata ahora de ofrecer una solución al problema que antes planteábamos para lo cual y en la línea antes descrita, ofrecemos una primera modificación al modelo propuesto por Ramsay[1977], bajo el punto de vista de la distribución lognormal triparamétrica, como justifica al hablar sobre la distribución de los residuos y la necesidad de introducir un parámetro de localización. Así pues, en este primer caso en que nuevamente consideramos varianzas constantes, espondremos que las variables aleatorias  $D_{ij}$  se distribuyen de la forma

$$D_{ij} \sim \Lambda(\theta_{ij}, \log(d_{ij}^*), \sigma^2); \quad i, j = 1, \dots, n.$$

donde

$$d_{ij}^* = \sqrt{\sum_{m=1}^K (x_{im} - x_{jm})^2}$$

con  $K$  = dimensión del espacio de la configuración.

En el modelo que presentamos aparecen tres bloques de parámetros a estimar: los parámetros de localización  $\theta_{ij}$ , el valor común de la varianza  $\sigma^2$  y la matriz de configuración  $X = (x_{pq})_{pq}$ , que será, según las hipótesis de partida, una matriz  $n \times K$ . Por tanto, en este modelo, la verosimilitud depende de tres clases de parámetros diferentes a estimar que constituyen un total de  $n \times K + 2$ ,

si se consideran constantes los parámetros de localización, o sea, si se considera  $\theta_{ij} = \theta, \forall i, j$ . En caso contrario, aparecerán un total de  $n \times K + n^2 + 1$  parámetros, siendo

$$L_{\{d_{ijr}\}}(\theta_{ij}, \sigma^2, X) = \prod_i \prod_j \prod_r \frac{1}{\sigma(d_{ijr} - \theta_{ij})\sqrt{2\pi}} \exp \left\{ \frac{-\log^2 \left( \frac{d_{ijr} - \theta_{ij}}{d_{ij}^*} \right)}{2\sigma^2} \right\}$$

Nuevamente supondremos que los índices varían sobre los datos que realmente han sido observados sin considerar aquellos posibles datos que por alguna razón no lo hayan sido. Por tanto, la log-verosimilitud quedará, de manera parecida al caso biparamétrico y considerando nuevamente la hipótesis de independencia, de la forma

$$\log L_{\{d_{ijr}\}}(\theta_{ij}, \sigma^2, \{x_{pm}\}_{pm}) = -\frac{1}{2} \left[ \frac{S}{\sigma^2} + M \log(\sigma^2) \right] - \sum_i \sum_j \sum_r \log(d_{ijr} - \theta_{ij}) - M \log(\sqrt{2\pi})$$

donde

$$S = \sum_i \sum_j \sum_r \log^2 \left( \frac{d_{ijr} - \theta_{ij}}{d_{ij}^*} \right)$$

y  $M$  es el número de datos observados.

Como puede comprobarse, la forma de la verosimilitud puede considerarse idéntica al caso biparamétrico, salvo el parámetro de localización  $\theta_{ij}$ . Así pues, si  $\theta_{ij}$  es conocido para el modelo, la estimación no ofrece ninguna dificultad y puede ser considerada idéntica a la del caso biparamétrico, salvo que donde aparecía  $d_{ijr}$  ahora se tiene  $(d_{ijr} - \theta_{ij})$ . En el caso general en que  $\theta_{ij}$  no es conocida, se estimará a partir de la muestra estudiada, de forma que si su valor es cero, el modelo se reduce al caso biparamétrico y el empleo de este modelo no mejoraría la resolución del problema del MDS. Si por el contrario y como es de esperar en unos datos afectados fuertemente por el efecto de la ley de Weber, el valor de  $\theta_{ij}$  queda desplazado a la derecha del cero, o bien, si tomara un valor

negativo para datos de disimilaridad con valores claramente negativos (siempre pensando como parece lógico, en el modelo de lognormal triparamétrica dado por la distribución de  $(D_{ij} - \theta_{ij})$  y no en el de la distribución de  $(D_{ij} + \theta_{ij})$  que podría ser enunciado equivalentemente), la presentación de este modelo mejora sin duda alguna al modelo simple presentado por Ramsay. Por tanto, vamos a considerar la estimación de los parámetros del modelo, incluida la estimación de  $\theta_{ij}$ , lo cual será efectuado nuevamente bajo el punto de vista de la estimación por máxima verosimilitud, comentando posteriormente los inconvenientes algebraicos que se plantean en este caso, para la resolución de las ecuaciones de verosimilitud.

**Proposición 3.2.4** *El estimador por máxima verosimilitud del parámetro de la varianza asociada a los residuos, en el modelo euclídeo, homocedástico, lognormal triparamétrico con parámetro de localización  $\theta_{ij}$ , es:*

$$\hat{\sigma}^2 = \frac{S}{M}$$

donde

$$S = \sum_i \sum_j \sum_r \log^2 \left( \frac{d_{ijr} - \theta_{ij}}{d_{ij}^*} \right)$$

y  $M$  es el número de datos observados.

DEMOSTRACIÓN.

Se tiene que

$$\log L = \log L_{\{d_{ijr}\}}(\theta_{ij}, \sigma^2, \{x_{pm}\}_{pm}) =$$

$$-\frac{1}{2} \left[ \frac{S}{\sigma^2} + M \log(\sigma^2) \right] - \sum_i \sum_j \sum_r \log(d_{ijr} - \theta_{ij}) - M \log(\sqrt{2\pi})$$

donde  $S$  y  $M$  son como antes. Derivando respecto de  $\sigma^2$  se obtiene que

$$\frac{\partial}{\partial \sigma^2} \log L = -\frac{1}{2} \left[ \frac{-S}{\sigma^4} + \frac{M}{\sigma^2} \right] = 0$$

$$\Rightarrow -\frac{S}{\sigma^2} + M = 0$$

luego

$$\hat{\sigma}^2 = \frac{S}{M}$$

donde

$$S = \sum_i \sum_j \sum_r \log^2 \left( \frac{d_{ijr} - \theta_{ij}}{d_{ij}^*} \right)$$

y  $M$  es el número de datos observados.

C.Q.D.

Tal y como puede apreciarse, el estimador de máxima verosimilitud de  $\sigma^2$ , depende tanto de la matriz de configuración  $X$  como del parámetro de localización  $\theta_{ij}$ . Así pues, en este caso, no resulta posible su estimación al margen de la de los otros parámetros, ya que cuando  $\theta_{ij}$  no es conocido no cabe más que realizar la estimación simultánea de todos los parámetros a través de la resolución iterativa de las ecuaciones de verosimilitud. Después comentaremos ampliamente el cálculo analítico de los estimadores a través de la solución del sistema de ecuaciones de verosimilitud. Veamos ya la estimación de  $\theta_{ij}$ . Para ello, consideraremos las dos posibilidades, tanto  $\theta_{ij}$  variable, como  $\theta_{ij}$  constante.

**Proposición 3.2.5** *El estimador de máxima verosimilitud del parámetro de localización en el modelo euclídeo, homocedástico, lognormal triparamétrico, viene determinado por la ecuación de verosimilitud:*

1. Para el caso de  $\theta_{pq}$ , variable.

$$\sum_r \left[ \log \left( \frac{d_{pqr} - \theta_{pq}}{d_{pq}^*} \right) + \sigma^2 \right] \frac{1}{d_{pqr} - \theta_{pq}} = 0$$

2. Para el caso de  $\theta_{pq} = \theta$ , constante.

$$\sum_i \sum_j \sum_r \left[ \log \left( \frac{d_{ijr} - \theta}{d_{ij}^*} \right) + \sigma^2 \right] \frac{1}{d_{ijr} - \theta} = 0$$

donde las sumas están dadas sobre los valores realmente observados.

DEMOSTRACIÓN.

• Cálculo del estimador  $\hat{\theta}_{pq}$ .

Se tiene que

$$\log L = -\frac{1}{2} \left[ \frac{S}{\sigma^2} + M \log(\sigma^2) \right] - \sum_i \sum_j \sum_r \log(d_{ijr} - \theta_{ij}) - M \log(\sqrt{2\pi})$$

donde  $S$  y  $M$  son como antes. Derivando respecto de  $\theta_{pq}$  se obtiene que

$$\frac{\partial}{\partial \theta_{pq}} \log L = -\frac{1}{2\sigma^2} \frac{\partial}{\partial \theta_{pq}} S - \sum_r \frac{\partial}{\partial \theta_{pq}} \log(d_{pqr} - \theta_{pq}) = (*)$$

puesto que

$$\begin{aligned} \frac{\partial}{\partial \theta_{pq}} S &= \frac{\partial}{\partial \theta_{pq}} \sum_i \sum_j \sum_r \log^2 \left( \frac{d_{ijr} - \theta_{ij}}{d_{ij}^*} \right) \\ &= -2 \sum_r \log \left( \frac{d_{pqr} - \theta_{pq}}{d_{pq}^*} \right) \frac{1}{d_{pqr} - \theta_{pq}} \end{aligned}$$

volviendo a la ecuación anterior se tiene que

$$(*) = \frac{1}{\sigma^2} \sum_r \log \left( \frac{d_{pqr} - \theta_{pq}}{d_{pq}^*} \right) \frac{1}{d_{pqr} - \theta_{pq}} + \sum_r \frac{1}{d_{pqr} - \theta_{pq}} = 0$$

de donde multiplicando por  $\sigma^2$  y simplificando, queda:

$$\sum_r \left[ \log \left( \frac{d_{pqr} - \theta_{pq}}{d_{pq}^*} \right) + \sigma^2 \right] \frac{1}{d_{pqr} - \theta_{pq}} = 0$$

• Cálculo del estimador  $\hat{\theta}$ .

Nuevamente

$$\log L = -\frac{1}{2} \left[ \frac{S}{\sigma^2} + M \log(\sigma^2) \right] - \sum_i \sum_j \sum_r \log(d_{ijr} - \theta) - M \log(\sqrt{2\pi})$$

donde ahora  $M$  es como antes y  $S$  es:

$$S = \sum_i \sum_j \sum_r \log^2 \left( \frac{d_{ijr} - \theta}{d_{ij}^*} \right)$$

Derivando respecto de  $\theta$  se obtiene que

$$\frac{\partial}{\partial \theta} \log L = -\frac{1}{2\sigma^2} \frac{\partial}{\partial \theta} S - \sum_i \sum_j \sum_r \frac{\partial}{\partial \theta} \log(d_{ijr} - \theta) = (*)$$

puesto que

$$\begin{aligned} \frac{\partial}{\partial \theta} S &= \frac{\partial}{\partial \theta} \sum_i \sum_j \sum_r \log^2 \left( \frac{d_{ijr} - \theta}{d_{ij}^*} \right) \\ &= -2 \sum_i \sum_j \sum_r \log \left( \frac{d_{ijr} - \theta}{d_{ij}^*} \right) \frac{1}{d_{ijr} - \theta} \end{aligned}$$

volviendo a la ecuación anterior

$$(*) = \frac{1}{\sigma^2} \sum_i \sum_j \sum_r \log \left( \frac{d_{ijr} - \theta}{d_{ij}^*} \right) \frac{1}{d_{ijr} - \theta} + \sum_i \sum_j \sum_r \frac{1}{d_{ijr} - \theta} = 0$$

multiplicando nuevamente por  $\sigma^2$  y simplificando, queda:

$$\sum_i \sum_j \sum_r \left[ \log \left( \frac{d_{ijr} - \theta}{d_{ij}^*} \right) + \sigma^2 \right] \frac{1}{d_{ijr} - \theta} = 0$$

C.Q.D.

La ecuación de verosimilitud que acabamos de obtener, depende tanto de  $\sigma^2$  como de las  $d_{ij}^*$ , por lo que para su estimación no puede desligarse de la estimación de esos parámetros. Posteriormente, en el modelo euclídeo ponderado, se comentará detenidamente el proceso de estimación de Cohen del parámetro de localización y se ofrecerá un procedimiento alternativo para su estimación.

Veamos finalmente el aspecto que toman las ecuaciones de verosimilitud para la estimación de la matriz de configuración  $X$ . Como podrá apreciarse,

la estimación de la configuración se realiza a través de la definición de distancia  $d_{ij}^*$  que sea considerada, por lo que depende totalmente de dicha definición. En este caso la consideramos euclídea, aunque después se verán algunas extensiones en cuanto a este aspecto. Veamos ya la expresión formal de todo lo expuesto.

**Teorema 3.2.6** *Consideremos en el modelo euclídeo, homocedástico, lognormal triparamétrico, una representación en un espacio de configuración conjunta de dimensión  $K$ . Notemos por  $x_i$  y  $x_j$  a los puntos de dicho espacio, representantes de los estímulos  $i$  y  $j$  respectivamente,*

$$x_i = (x_{i1}, \dots, x_{im}, \dots, x_{iK})$$

$$x_j = (x_{j1}, \dots, x_{jm}, \dots, x_{jK})$$

Entonces, el estimador por máxima verosimilitud de la matriz de configuración  $X$ , vendrá dado por el sistema de ecuaciones:

$$x_{pq} \sum_j t_{pj} = \sum_j x_{jq} t_{pj}$$

$$\forall p = 1, \dots, n. \quad y \quad \forall q = 1, \dots, K.$$

donde, para cada  $p$  y cada  $j$ ,

$$t_{pj} = \frac{1}{(d_{pj}^*)^2} \sum_r \left[ \log \left( \frac{d_{pjr} - \theta_{pj}}{d_{pj}^*} \right) + \log \left( \frac{d_{jpr} - \theta_{jp}}{d_{pj}^*} \right) \right]$$

DEMOSTRACIÓN.

Se tiene que

$$\log L = \log L_{\{d_{ijr}\}}(\theta_{ij}, \sigma^2, \{x_{pm}\}_{pm}) =$$

$$-\frac{1}{2} \left[ \frac{S}{\sigma^2} + M \log(\sigma^2) \right] - \sum_i \sum_j \sum_r \log(d_{ijr} - \theta_{ij}) - M \log(\sqrt{2\pi})$$

donde

$$S = \sum_i \sum_j \sum_r \log^2 \left( \frac{d_{ijr} - \theta_{ij}}{d_{ij}^*} \right)$$

y  $M$  = el número de datos observados.

Así pues, tendremos que

$$\frac{\partial \log L}{\partial x_{pq}} = -\frac{1}{2} \left[ \frac{1}{\sigma^2} \frac{\partial S}{\partial x_{pq}} \right] = \frac{\partial S}{\partial x_{pq}} = 0$$

Así pues, basta con maximizar  $S$  con respecto a  $x_{pq}$  para obtener el máximo de la función de log-verosimilitud en ese punto.

$$\begin{aligned} \frac{\partial S}{\partial x_{pq}} &= \frac{\partial}{\partial x_{pq}} \sum_i \sum_j \sum_r \log^2 \left( \frac{d_{ijr} - \theta_{ij}}{d_{ij}^*} \right) = \\ &= -2 \sum_i \sum_j \sum_r \log \left( \frac{d_{ijr} - \theta_{ij}}{d_{ij}^*} \right) \frac{\partial}{\partial x_{pq}} \log(d_{ij}^*) = \\ &= -2 \sum_i \sum_j \sum_r \log \left( \frac{d_{ijr} - \theta_{ij}}{d_{ij}^*} \right) \frac{1}{d_{ij}^*} \frac{\partial}{\partial x_{pq}} d_{ij}^* = \\ &= -2 \sum_i \sum_j \frac{1}{d_{ij}^*} \frac{\partial}{\partial x_{pq}} d_{ij}^* \sum_r \log \left( \frac{d_{ijr} - \theta_{ij}}{d_{ij}^*} \right) = \\ &= -2 \sum_j \frac{1}{d_{pj}^*} \frac{\partial}{\partial x_{pq}} d_{pj}^* \sum_r \log \left( \frac{d_{pjr} - \theta_{pj}}{d_{pj}^*} \right) - 2 \sum_i \frac{1}{d_{ip}^*} \frac{\partial}{\partial x_{pq}} d_{ip}^* \sum_r \log \left( \frac{d_{ipr} - \theta_{ip}}{d_{ip}^*} \right) = \\ &= -2 \sum_j \frac{1}{d_{pj}^*} \frac{(x_{pq} - x_{jq})}{d_{pj}^*} \sum_r \log \left( \frac{d_{pjr} - \theta_{pj}}{d_{pj}^*} \right) - \\ &\quad - 2 \sum_i \frac{1}{d_{ip}^*} \frac{(x_{pq} - x_{iq})}{d_{ip}^*} \sum_r \log \left( \frac{d_{ipr} - \theta_{ip}}{d_{ip}^*} \right) = (*) \end{aligned}$$

Considerando ambos sumatorios en  $j$  y utilizando la simetría en las distancias se tiene

$$(*) = -2 \sum_j \frac{1}{d_{pj}^*} \frac{(x_{pq} - x_{jq})}{d_{pj}^*} \sum_r \log \left( \frac{d_{pjr} - \theta_{pj}}{d_{pj}^*} \right) -$$

$$\begin{aligned}
 & -2 \sum_j \frac{1}{d_{pj}^*} \frac{(x_{pq} - x_{jq})}{d_{pj}^*} \sum_r \log \left( \frac{d_{jpr} - \theta_{jp}}{d_{pj}^*} \right) = \\
 & = -2 \sum_j \frac{1}{d_{pj}^*} \frac{(x_{pq} - x_{jq})}{d_{pj}^*} \sum_r \left[ \log \left( \frac{d_{pjr} - \theta_{pj}}{d_{pj}^*} \right) + \log \left( \frac{d_{jpr} - \theta_{jp}}{d_{pj}^*} \right) \right]
 \end{aligned}$$

Igualando a cero y simplificando queda la expresión

$$\sum_j \frac{1}{(d_{pj}^*)^2} (x_{pq} - x_{jq}) \sum_r \left[ \log \left( \frac{d_{pjr} - \theta_{pj}}{d_{pj}^*} \right) + \log \left( \frac{d_{jpr} - \theta_{jp}}{d_{pj}^*} \right) \right] = 0$$

Descomponiendo la diferencia entre los valores de  $X$  se tiene

$$\begin{aligned}
 & x_{pq} \sum_j \frac{1}{(d_{pj}^*)^2} \sum_r \left[ \log \left( \frac{d_{pjr} - \theta_{pj}}{d_{pj}^*} \right) + \log \left( \frac{d_{jpr} - \theta_{jp}}{d_{pj}^*} \right) \right] = \\
 & = \sum_j \frac{1}{(d_{pj}^*)^2} x_{jq} \sum_r \left[ \log \left( \frac{d_{pjr} - \theta_{pj}}{d_{pj}^*} \right) + \log \left( \frac{d_{jpr} - \theta_{jp}}{d_{pj}^*} \right) \right]
 \end{aligned}$$

Si llamamos

$$t_{pj} = \frac{1}{(d_{pj}^*)^2} \sum_r \left[ \log \left( \frac{d_{pjr} - \theta_{pj}}{d_{pj}^*} \right) + \log \left( \frac{d_{jpr} - \theta_{jp}}{d_{pj}^*} \right) \right]$$

se obtiene finalmente que

$$x_{pq} \sum_j t_{pj} = \sum_j x_{jq} t_{pj} \quad \forall p = 1, \dots, n \text{ y } \forall q = 1, \dots, K$$

C.Q.D.

El modelo de MDS métrico sin diferencias individuales, cuya matriz de configuración acabamos de estimar y en general, el MDS en su totalidad, sufre de un grave inconveniente que necesariamente debe tenerse en cuenta a la hora de la estimación y sobre todo, en el momento de la interpretación de los resultados. En muchas ocasiones surgen una serie de indeterminaciones, que generalmente van asociadas al modelo de distancia que es utilizado en cada

ocasión y que afectan a la estimación de los parámetros del modelo. En el modelo euclídeo sin diferencias individuales que acaba de ser estimado, parece evidente que tanto una traslación como una rotación, no afectan en ninguna medida a la forma de la matriz de configuración estimada  $X$ . Realmente, lo que define una distancia como euclídea, no es el sistema de coordenadas tal y como en algunas ocasiones suele ocurrir al confundir distancia euclídea con coordenadas cartesianas. La distancia euclídea se define por su invarianza, exclusivamente, con respecto a dos tipos de transformaciones de los puntos y del sistema de coordenadas: traslaciones y rotaciones. Ambas transformaciones mantienen fijos los ángulos entre los puntos y las dimensiones. Una medida de distancia será por tanto no euclídea, si no es invariante respecto a dichas transformaciones o bien si es invariante frente a otras transformaciones distintas a las mencionadas.

Para el modelo de distancia considerado en este apartado, cualquier movimiento rígido deja invariante la matriz de configuración  $X$ , ya que dicho movimiento no altera la matriz de la métrica, que en esta ocasión es la matriz identidad. Por esa razón, debido a la *invarianza traslacional* y a la *invarianza rotacional* de este modelo, surgen una serie de indeterminaciones que, en este caso, afectan a la estimación de la matriz de configuración conjunta  $X$  y por tanto, a la estimación de otros parámetros que dependen de ella.

En general, en aquellos modelos de la estadística y en particular en MDS, en los que aparecen parámetros que no están especificados de forma única con respecto a los datos, suele ser práctico seguir uno de los siguientes procedimientos:

1. Reducir el número de parámetros independientes, intentando relacionar alguno o algunos de ellos en términos de los otros.
2. Estimar solamente ciertas funciones de los parámetros, que a priori sí se encuentren definidas de forma única por los datos.
3. Imponer algunas restricciones sobre los parámetros del modelo, que reduzcan el espacio paramétrico hasta obtener una región donde éstos queden determinados de forma única a través de los datos.

De entre estas tres metodologías, solo puede afirmarse que tenga sentido en el entorno del MDS una de ellas; concretamente la tercera. La simetría de la parametrización hace que parezca poco natural fijar unos pocos parámetros en

términos de los restantes. Además, no existen funciones de los parámetros que sean de un interés tan especial como para su identificación. Así pues, parece más natural eliminar la invarianza traslacional y rotacional de las coordenadas mediante el establecimiento de una serie de restricciones que definan de forma única sus orientaciones y representaciones en un espacio euclídeo. En este sentido, se considerarán en este trabajo las siguientes restricciones:

$$\sum_i x_{im} = 0, \quad \forall m = 1, \dots, K$$

$$\sum_i x_{im} x_{il} = 0, \quad \forall m, l = 1, \dots, K, \quad m \neq l$$

La primera ecuación implica que el origen del sistema de coordenadas se fija en el *centroide* o centro de gravedad de la configuración, con lo que se evita la invarianza traslacional. La eliminación de la invarianza rotacional constituye un problema más complicado de eliminar. Para evitarla, no resulta posible utilizar las técnicas del análisis factorial, dada la indeterminación de orígenes del MDS. Así pues, a través de la segunda ecuación se pretende eliminar dicha invarianza, al obligar a la configuración a poseer una orientación en el sentido de los *ejes principales*, o sea, la primera dimensión tendrá la mayor variabilidad posible en sus coordenadas, la segunda la mayor variabilidad posible dada la primera, etc... Esta segunda restricción, sin embargo, no puede ser aplicada cuando la matriz de la métrica es diagonal, como es nuestro caso, ya que entonces, una rotación haría que dicha matriz tuviese valores distintos de cero fuera de los elementos diagonales. Además, tiene el inconveniente de que para configuraciones en dimensiones superiores a dos, puede resultar confusa la interpretación de la representación que se obtenga, dada la indeterminación asociada. Por tanto, en este modelo no será considerada, con lo que no hay más remedio que mantener una invarianza rotacional.

Por tanto, teniendo en cuenta estas consideraciones, parece necesario volver a estimar la matriz de configuración  $X$  para obtener el *estimador restringido* asociado al criterio de Máxima Verosimilitud.

**Corolario 3.2.7** *Consideremos en el modelo euclídeo, homocedástico, lognormal triparamétrico, una representación en un espacio de configuración conjunta de dimensión  $K$ . Sean  $x_i$  y  $x_j$ , los puntos de dicho espacio representantes de los estímulos  $i$  y  $j$  respectivamente,*

$$x_i = (x_{i1}, \dots, x_{im}, \dots, x_{iK})$$

$$x_j = (x_{j1}, \dots, x_{jm}, \dots, x_{jK})$$

Entonces, el estimador por máxima verosimilitud de la matriz de configuración  $X$ , sujeto a las restricciones impuestas por

$$\sum_i x_{im} = 0, \quad \forall m = 1, \dots, K$$

vendrá dado por el sistema de ecuaciones implícitas:

$$x_{pq} = \frac{\sum_j t_{pj} x_{jq} - \mu_q \sigma^2}{\sum_j t_{pj}} \quad p = 1, \dots, n. \quad y \quad q = 1, \dots, K.$$

donde, para cada  $p, j$  y  $q$ ,

$$t_{pj} = \frac{1}{(d_{pj}^*)^2} \sum_r \left[ \log \left( \frac{d_{pjr} - \theta_{pj}}{d_{pj}^*} \right) + \log \left( \frac{d_{jpr} - \theta_{jp}}{d_{pj}^*} \right) \right]$$

y donde el multiplicador de Lagrange  $\mu_q$ , viene determinado a través de

$$\mu_q = \frac{1}{\sigma^2} \frac{\sum_p \sum_j u_p t_{pj} x_{jq}}{\sum_p u_p} \quad \text{con} \quad u_p = \frac{1}{\sum_j t_{pj}}$$

DEMOSTRACIÓN.

Para la estimación por máxima verosimilitud bajo restricciones de la matriz de configuración  $X$ , vamos a emplear el método de los multiplicadores de Lagrange, cuyos coeficientes notaremos por  $\mu_m$ . Así pues, tratamos de maximizar la función

$$Q(X) = \log L_{\{d_{ijr}\}}(\theta_{ij}, \sigma^2, \{x_{pm}\}_{pm}) + \sum_m \mu_m \left( \sum_i x_{im} \right)$$

Por tanto, derivando  $\log L_{\{d_{ijr}\}}(\theta_{ij}, \sigma^2, \{x_{pm}\}_{pm})$  respecto de un valor  $x_{pq}$  se tiene, en virtud del teorema anterior, que:

$$\frac{\partial}{\partial x_{pq}} \log L = \frac{1}{\sigma^2} \sum_j \frac{1}{(d_{pj}^*)^2} (x_{pq} - x_{jq}) \sum_r \left[ \log \left( \frac{d_{pjr} - \theta_{pj}}{d_{pj}^*} \right) + \log \left( \frac{d_{jpr} - \theta_{jp}}{d_{pj}^*} \right) \right]$$

Por otro lado

$$\frac{\partial}{\partial x_{pq}} \sum_m \mu_m \left( \sum_i x_{im} \right) = \mu_q$$

Si llamamos

$$t_{pj} = \frac{1}{(d_{pj}^*)^2} \sum_r \left[ \log \left( \frac{d_{pjr} - \theta_{pj}}{d_{pj}^*} \right) + \log \left( \frac{d_{jpr} - \theta_{jp}}{d_{pj}^*} \right) \right]$$

las ecuaciones de verosimilitud pueden escribirse de la forma

$$\frac{\partial}{\partial x_{pq}} Q(X) = x_{pq} \sum_j t_{pj} - \sum_j t_{pj} x_{jq} + \mu_q \sigma^2 = 0$$

$$\frac{\partial}{\partial \mu_q} Q(X) = \sum_i x_{iq} = 0$$

Si despejamos, en la primera ecuación,  $x_{pq}$  se tiene que

$$x_{pq} = \frac{\sum_j t_{pj} x_{jq} - \mu_q \sigma^2}{\sum_j t_{pj}}$$

Para obtener  $\mu_q$ , basta con sumar la expresión anterior en  $p$  y se obtiene

$$0 = \sum_p x_{pq} = \sum_p \frac{\sum_j t_{pj} x_{jq}}{\sum_j t_{pj}} - \mu_q \sigma^2 \sum_p \frac{1}{\sum_j t_{pj}}$$

de donde, despejando en  $\mu_q$ , se tiene finalmente

$$\mu_q = \frac{1}{\sigma^2} \frac{\sum_p \sum_j u_p t_{pj} x_{jq}}{\sum_p u_p} \quad \text{con} \quad u_p = \frac{1}{\sum_j t_{pj}}$$

C.Q.D.

Para finalizar el estudio en este primer modelo, habría que hacer hincapié en dos aspectos fundamentales: Por un lado, estudiar la resolución a través de procedimientos numéricos de dichas ecuaciones y por otro, estudiar la distribución asintótica de los estimadores que de esas ecuaciones se deduzcan.

La primera cuestión, aunque es de carácter exclusivamente analítico, será tratada en un apartado especial, en donde se ofrecerá un resumen de las líneas generales que se siguen para la resolución de sistemas de ecuaciones no lineales, en el entorno del análisis numérico, aplicadas a la estimación por máxima verosimilitud. Concretamente, serán descritos sin entrar en profundidad en su desarrollo, los métodos principales que usualmente son empleados para la resolución de sistemas de ecuaciones de verosimilitud. Además de éstos métodos derivados del de Newton-Raphson, en este caso será descrito un método particular que ha sido desarrollado para dar solución a los sistemas de ecuaciones implícitas que acabamos de obtener. No obstante, dicho método no resulta óptimo cuando el número de parámetros aumenta, por lo que en general, la resolución de los sistemas de ecuaciones que aparecen en este trabajo, serán resueltos a través del método *Scoring* de Fisher y que será descrito a continuación.

Por último, la segunda cuestión que se plantea referente a la obtención de la distribución asintótica de los estimadores, así como a la obtención de regiones de confianza asintóticas y contrastes de hipótesis, será igualmente tratada, de forma que se dará solución a este problema, bajo la hipótesis de verificación de una serie de condiciones de regularidad que a tal efecto serán definidas.

#### **3.2.4 Estimación en el modelo Euclídeo heteroscedástico triparamétrico.**

Nos planteamos ahora la posibilidad de que la varianza dependa de cada par de individuos realmente observados, en lugar de suponerla constante en todo el proceso. Así pues, frente al modelo anterior en el que se consideraba la varianza constante para todos los individuos o réplicas y todos los pares de estímulos, se trata de ofrecer ahora una extensión, en base a consideraciones de variabilidad, por ahora, intrínsecas a cada pareja de individuos observados y nuevamente en el marco de una métrica euclídea y una distribución inicial

de las disimilaridades según una ley lognormal triparamétrica. Además, la obtención de los resultados que se exponen en esta sección puede reducirse al caso biparamétrico sin más que considerar el parámetro de localización  $\theta_{ij} = 0$ , para todo par de estímulos  $i$  y  $j$ .

Para justificar la necesidad de descomponer la variabilidad, es necesario fijar una perspectiva desde el punto de vista experimental. Las distancias no se ajustan perfectamente a los valores obtenidos de disimilaridad, por lo que como después se verá, se requiere de una transformación de los datos para la mejora de dicho ajuste. No obstante, incluso bajo la mejor de las transformaciones, las distancias tampoco se ajustarán perfectamente a las disimilaridades transformadas. Los factores que hacen que los residuos tengan un valor distinto de cero son muchos y muy diversos. Existen dos formas de controlar esas fuentes de variación en los residuos o errores. Por un lado, teniendo en cuenta varias componentes de variación de forma que sea posible a posteriori un mayor refinamiento del análisis a través de estimaciones más precisas. Por otro lado, un resumen de las componentes de la variación puede igualmente añadir aspectos importantes a la interpretación de los datos.

Por tanto, el problema fundamental consiste en analizar a posteriori la varianza de los residuos. En este apartado, se tratará de separar esa variabilidad en componentes simples, que en esta ocasión serán específicas únicamente de los estímulos y que notaremos por  $\gamma_{ij}$ . Estas cantidades pueden definirse como el *error típico entre pares*. Dicha posibilidad de descomposición fue introducida por Ramsay [1978] aunque referida al modelo de diferencias individuales, biparamétrico y sometida a restricciones para su estimación bajo ese modelo. En este caso se estimará sin someterla a restricciones ya que no tienen sentido en el contexto en que nos situamos.

Además de lo anterior, otro aspecto a considerar consiste en establecer explícitamente la posibilidad de que cada estímulo aporte un valor independiente al montante total de la varianza para cada par. Esto se realiza a través de la descomposición de la variabilidad  $\gamma_{ij}$  en componentes específicas de cada estímulo y que notaremos por  $\alpha_i$ . Estas componentes serán interpretadas por tanto como las contribuciones relativas a la varianza, por parte de cada estímulo del par. Por ejemplo, las percepciones de los sujetos sobre un estímulo particular  $i$ , podrían ser especialmente variables o estar indeterminadas, causando clasificaciones de pares que ocasionen que los estímulos posean un alto grado de varianza. Esto podría quedar reflejado en un valor de la componente correspondiente a dicho estímulo  $\alpha_i$ , próximo al montante total de la variabi-

lidad de cada par  $\gamma_{ij}$ . Por el contrario, un cierto estímulo podría actuar como estímulo típico o de referencia respecto a aquellos con los que es comparado, causando por tanto menor variabilidad en aquellos juicios en los que éste se encuentra implicado y quedando por tanto identificado, a través de un valor de su componente de variabilidad  $\alpha_i$ , más bajo que los del resto de los estímulos con los que es comparado.

El modelo que se plantea ahora, siguiendo la misma notación que en el caso anterior, con parámetro de localización variante  $\theta_{ij}$ , puede enunciarse como

$$D_{ij} \sim \Lambda(\theta_{ij}, \log(d_{ij}^*), \sigma_{ij}^2), \quad \forall i, j = 1, \dots, n.$$

donde

$$d_{ij}^* = \sqrt{\sum_{m=1}^K (x_{im} - x_{jm})^2}$$

con  $K$  = dimensión del espacio de la configuración.

Además, son consideradas a su vez dos estructuras de descomposición de la variabilidad del modelo, en base a componentes específicas asociadas a los estímulos, que permiten poner de manifiesto la relación entre la variabilidad total en el par y el montante asociado a cada estímulo de dicho par. Esto se hace en base a dos modelos de descomposición: un modelo aditivo y uno multiplicativo.

En resumen, serán consideradas las tres estructuras siguientes:

- Variabilidad conjunta del par.

$$\sigma_{ij}^2 = \gamma_{ij}^2$$

- Modelo aditivo.

Este modelo se propone por Ramsay [1982] para el caso de MDS con diferencias individuales y viene dado por la media aritmética de las dos componentes asociadas a los estímulos. Estas componentes serán definidas por un valor  $\alpha_i^2$ , asociado a cada estímulo  $i$  y un valor  $\alpha_j^2$  para el estímulo  $j$ . Por tanto, en éste se tiene la relación:

$$\sigma_{ij}^2 = \frac{\alpha_i^2 + \alpha_j^2}{2}$$

- Modelo multiplicativo.

Este modelo es propuesto en este trabajo y su justificación principal estriba en el hecho de que permite una estimación menos costosa que la del modelo aditivo. Viene dado a través de la media geométrica de las dos componentes individuales que intervienen en cada  $\sigma_{ij}^2$ , en consonancia con la estructura multiplicativa del parámetro de la varianza, bajo una descomposición estructural que será expuesta en el modelo con diferencias individuales. Por tanto, se tiene:

$$\sigma_{ij}^2 = \sqrt{\alpha_i^2 \alpha_j^2}$$

Bajo la situación que se acaba de describir, la verosimilitud asociada a la muestra adopta la expresión,

$$L_{\{d_{ijr}\}}(\theta_{ij}, \gamma_{ij}^2, X) = \prod_i \prod_j \prod_r \frac{1}{\gamma_{ij}(d_{ijr} - \theta_{ij})\sqrt{2\pi}} \exp \left\{ \frac{-\log^2 \left( \frac{d_{ijr} - \theta_{ij}}{d_{ij}^*} \right)}{2\gamma_{ij}^2} \right\}$$

donde los productos se supondrán lógicamente entre los datos realmente observados. La log-verosimilitud quedará,

$$\begin{aligned} \log L_{\{d_{ijr}\}}(\theta_{ij}, \gamma_{ij}^2, \{x_{pm}\}_{pm}) &= -\frac{1}{2} \sum_i \sum_j \left[ \frac{1}{\gamma_{ij}^2} S_{ij} + R \log(\gamma_{ij}^2) \right] - \\ &\quad - \sum_i \sum_j \sum_r \log(d_{ijr} - \theta_{ij}) - M \log(\sqrt{2\pi}) \end{aligned}$$

donde ahora,

$$S_{ij} = \sum_r \log^2 \left( \frac{d_{ijr} - \theta_{ij}}{d_{ij}^*} \right)$$

y  $M$  es el número de datos observados.

Nuevamente se procederá a la estimación de los diferentes parámetros asociados al modelo, bajo el punto de vista de la estimación por máxima verosimilitud.

**Proposición 3.2.8** *En el modelo euclídeo, heteroscedástico lognormal triparamétrico, el estimador de máxima verosimilitud del parámetro de variabilidad conjunta entre el par  $(p, q)$ , viene dado por la expresión,*

$$\hat{\gamma}_{pq}^2 = \frac{1}{2R} \sum_r \left[ \log^2 \left( \frac{d_{pqr} - \theta_{pq}}{d_{pq}^*} \right) + \log^2 \left( \frac{d_{qpr} - \theta_{qp}}{d_{pq}^*} \right) \right] \quad \forall p, q = 1, \dots, n.$$

DEMOSTRACIÓN.

Considerando la log verosimilitud asociada a la muestra,

$$-\frac{1}{2} \sum_i \sum_j \left[ \frac{1}{\gamma_{ij}^2} S_{ij} + R \log(\gamma_{ij}^2) \right] - \sum_i \sum_j \sum_r \log(d_{ijr} - \theta_{ij}) - M \log(\sqrt{2\pi})$$

con,

$$S_{ij} = \sum_r \log^2 \left( \frac{d_{ijr} - \theta_{ij}}{d_{ij}^*} \right)$$

y considerando la simetría dada por  $\gamma_{pq}^2 = \gamma_{qp}^2$ , se obtiene sin más que derivar, la relación,

$$\frac{\partial}{\partial \gamma_{pq}^2} = \left[ \frac{-1}{\gamma_{pq}^4} S_{pq} + R \frac{1}{\gamma_{pq}^2} \right] - \left[ \frac{-1}{\gamma_{pq}^4} S_{pq} + R \frac{1}{\gamma_{pq}^2} \right] = 0$$

de donde, simplificando y reagrupando convenientemente se deduce que

$$\frac{-1}{\gamma_{pq}^2} (S_{pq} + S_{qp}) + 2R = 0$$

y despejando, finalmente

$$\gamma_{pq}^2 = \frac{1}{2R} \sum_r \left[ \log^2 \left( \frac{d_{pqr} - \theta_{pq}}{d_{pq}^*} \right) + \log^2 \left( \frac{d_{qpr} - \theta_{qp}}{d_{pq}^*} \right) \right]$$

C.Q.D.

Veamos ahora como queda la estimación de las componentes que expresan la contribución relativa que cada estímulo aporta a la variabilidad entre cada pareja de objetos estudiados.

José Fernando Vera Vera.

**Teorema 3.2.9** Consideremos la hipótesis de un modelo euclídeo, heteroscedástico, lognormal triparamétrico, sin diferencias individuales. Supongamos que la variabilidad del modelo asociada a cada par de estímulos  $\gamma_{ij}^2$ , queda expresada según las estructuras:

1. Modelo aditivo.

$$\gamma_{ij}^2 = \frac{\alpha_i^2 + \alpha_j^2}{2}$$

2. Modelo multiplicativo.

$$\gamma_{ij}^2 = \sqrt{\alpha_i^2 \alpha_j^2}$$

Entonces, el estimador de máxima verosimilitud de la componente de variabilidad asociada a cada individuo, viene expresada a través de las siguientes ecuaciones:

1. En el modelo aditivo.

$$\hat{\alpha}_p^2 = \frac{\sum_j [(S_{pj} + S_{jp}) - R\alpha_j^2] t_{pj}^2}{R \sum_j t_{pj}^2}; \quad p = 1, \dots, n.$$

2. En el modelo multiplicativo.

$$\hat{\alpha}_p^2 = \left[ \frac{1}{2Rn} \sum_j \frac{S_{pj} + S_{jp}}{|\alpha_j|} \right]^2$$

donde

$$S_{pj} = \sum_r \log^2 \left( \frac{d_{pjr} - \theta_{pj}}{d_{pj}^*} \right); \quad p, j = 1, \dots, n.$$

y

$$t_{pj}^2 = \frac{1}{(\alpha_p^2 + \alpha_j^2)^2}$$

DEMOSTRACIÓN.

Vamos a separar la demostración para cada uno de los modelos enunciados.

• Cálculo del estimador en el modelo aditivo.

En este caso, la log-verosimilitud puede expresarse de la forma:

$$\begin{aligned} \log L = & -\frac{1}{2} \sum_i \sum_j \left[ R \log(\alpha_i^2 + \alpha_j^2) + \frac{2}{\alpha_i^2 + \alpha_j^2} S_{ij} \right] + \\ & + \frac{M}{2} \log(2) - \sum_i \sum_j \sum_r \log(d_{ijr} - \theta_{ij}) - M \log(\sqrt{2\pi}) \end{aligned}$$

siendo  $M$  el número total de muestras observadas y

$$S_{ij} = \sum_r \log^2 \left( \frac{d_{ijr} - \theta_{ij}}{d_{ij}^*} \right)$$

Por tanto, derivando respecto  $\alpha_p^2$ , e igualando a cero se tiene

$$\frac{\partial \log L}{\partial \alpha_p^2} = \frac{\partial}{\partial \alpha_p^2} \sum_i \sum_j \left[ R \log(\alpha_i^2 + \alpha_j^2) + \frac{2}{\alpha_i^2 + \alpha_j^2} S_{ij} \right] = 0$$

de donde se obtiene,

$$\begin{aligned} & \sum_i \left( \frac{\partial}{\partial \alpha_p^2} \alpha_i^2 \right) \sum_j \left[ \frac{R}{\alpha_i^2 + \alpha_j^2} - \frac{2S_{ij}}{(\alpha_i^2 + \alpha_j^2)^2} \right] + \\ & + \sum_j \left( \frac{\partial}{\partial \alpha_p^2} \alpha_j^2 \right) \sum_i \left[ \frac{R}{\alpha_j^2 + \alpha_i^2} - \frac{2S_{ji}}{(\alpha_j^2 + \alpha_i^2)^2} \right] = 0 \end{aligned}$$

Calculando la derivada y simplificando queda,

$$\sum_j \frac{R}{\alpha_p^2 + \alpha_j^2} - \sum_j \frac{S_{pj} + S_{jp}}{(\alpha_p^2 + \alpha_j^2)^2} = 0$$

Luego, si multiplicamos y dividimos por  $\alpha_p^2 + \alpha_j^2$  dentro del primer sumatorio y si llamamos,

$$t_{pj}^2 = \frac{1}{(\alpha_p^2 + \alpha_j^2)^2}$$

simplificando queda:

$$R\alpha_p^2 \sum_j t_{pj}^2 + R \sum_j \alpha_j^2 t_{pj}^2 - \sum_j (S_{pj} + S_{jp}) t_{pj}^2 = 0$$

Por tanto, agrupando convenientemente y despejando se obtiene,

$$\alpha_p^2 = \frac{\sum_j [(S_{pj} + S_{jp}) - R\alpha_j^2] t_{pj}^2}{R \sum_j t_{pj}^2}$$

- Cálculo en el modelo multiplicativo.  
En este caso, la log-verosimilitud queda

$$\begin{aligned} \log L = & -\frac{1}{2} \sum_i \sum_j \left[ R \log(\text{sqrta}_i^2 \alpha_j^2) + \frac{1}{\text{sqrta}_i^2 \alpha_j^2} S_{ij} \right] - \\ & - \sum_i \sum_j \sum_r \log(d_{ijr} - \theta_{ij}) - M \log(\sqrt{2\pi}) \end{aligned}$$

Si derivamos en la expresión anterior respecto a  $\alpha_p^2$  e igualamos a cero, resulta

$$\frac{\partial \log L}{\partial \alpha_p^2} = \sum_i \sum_j \left[ \left( \frac{R}{(\alpha_i^2 \alpha_j^2)^{1/2}} - \frac{S_{ij}}{(\alpha_i^2 \alpha_j^2)} \right) \left( \frac{\partial}{\partial \alpha_p^2} (\alpha_i^2 \alpha_j^2)^{1/2} \right) \right] = 0$$

de donde, derivando y simplificando se obtiene que

$$\begin{aligned} & \sum_j \frac{\alpha_j^2}{(\alpha_i^2 \alpha_j^2)^{1/2}} \left( \frac{R}{(\alpha_p^2 \alpha_j^2)^{1/2}} - \frac{S_{pj}}{(\alpha_p^2 \alpha_j^2)} \right) + \\ & + \sum_i \frac{\alpha_i^2}{(\alpha_i^2 \alpha_p^2)^{1/2}} \left( \frac{R}{(\alpha_p^2 \alpha_i^2)^{1/2}} - \frac{S_{ip}}{(\alpha_p^2 \alpha_i^2)} \right) = 0 \end{aligned}$$

Renombrando el sumatorio en  $i$  por  $j$  y simplificando queda

$$\sum_j (\alpha_j^2)^{1/2} \left( \frac{R}{(\alpha_p^2 \alpha_j^2)^{1/2}} - \frac{S_{pj}}{(\alpha_p^2 \alpha_j^2)} \right) +$$

$$+ \sum_j (\alpha_j^2)^{1/2} \left( \frac{R}{(\alpha_p^2 \alpha_j^2)^{1/2}} - \frac{S_{jp}}{(\alpha_p^2 \alpha_j^2)} \right) = 0$$

Simplificando nuevamente y reagrupando se obtiene por último la expresión:

$$2Rn(\alpha_p^2)^{1/2} = \sum_j \frac{S_{pj} + S_{jp}}{(\alpha_j^2)^{1/2}}$$

de donde finalmente,

$$\alpha_p^2 = \left[ \frac{1}{2Rn} \sum_j \frac{S_{pj} + S_{jp}}{|\alpha_j|} \right]^2$$

C.Q.D.

Respecto de los resultados que acabamos de obtener, hay que decir que la estimación de las componentes  $\alpha_p^2$  suele efectuarse a posteriori para el modelo aditivo, debido al gran costo computacional que supone la resolución de dichas ecuaciones. Para ello, basta tener en cuenta que

$$t_{pj}^2 = \frac{1}{2\gamma_{ij}^2}$$

Para resolver este problema computacional, se ha definido el modelo multiplicativo, más consonante con la estructura de la varianza. La estimación bajo esta hipótesis estructural supone poco tiempo de cálculo en el cómputo total de los estimadores del modelo. Además, la elección de una media geométrica para la descomposición del parámetro de la varianza, está más relacionada con la estructura de la varianza de los modelos lognormales. De cualquier forma, a lo largo de este trabajo se mantienen ambos modelos, ya que cada problemática particular puede requerir de uno u otro indistintamente.

Veamos a continuación la estimación directa para este modelo heteroscedástico, de los parámetros de localización, nuevamente bajo la doble consideración de suponer dependencia o no de los estímulos.

**Proposición 3.2.10** *El estimador de máxima verosimilitud del parámetro de localización en el modelo euclídeo, heteroscedástico, lognormal triparamétrico,*

con parámetro de dispersión  $\gamma_{ij}^2$  asociado a cada par de estímulos  $(i, j)$ , viene determinado por la ecuación implícita:

1. Para el caso de  $\theta_{pq}$ , variante.

$$\hat{\theta}_{pq} = \frac{\sum_r \frac{d_{pqr}}{(d_{pqr} - \theta_{pq})^2} \left[ \log \left( \frac{d_{pqr} - \theta_{pq}}{d_{pq}^*} \right) + \gamma_{pq}^2 \right]}{\sum_r \frac{1}{(d_{pqr} - \theta_{pq})^2} \left[ \log \left( \frac{d_{pqr} - \theta_{pq}}{d_{pq}^*} \right) + \gamma_{pq}^2 \right]}$$

2. Para el caso de  $\theta$ , constante.

$$\hat{\theta} = \frac{\sum_i \sum_j \sum_r \frac{1}{d_{ijr} - \theta} \left[ \frac{1}{\gamma_{ij}^2} \log \left( \frac{d_{ijr} - \theta}{d_{ij}^*} \right) + \frac{d_{ijr}}{d_{ijr} - \theta} \right]}{\sum_i \sum_j \sum_r \frac{1}{(d_{ijr} - \theta)^2}}$$

DEMOSTRACIÓN.

• Cálculo del estimador  $\hat{\theta}_{pq}$ .

Tenemos que

$$\begin{aligned} \log L = & -\frac{1}{2} \sum_i \sum_j \left[ \frac{1}{\gamma_{ij}^2} S_{ij} + R \log(\gamma_{ij}^2) \right] - \\ & - \sum_i \sum_j \sum_r \log(d_{ijr} - \theta_{ij}) - M \log(\sqrt{2\pi}) \end{aligned}$$

con,

$$S_{ij} = \sum_r \log^2 \left( \frac{d_{ijr} - \theta_{ij}}{d_{ij}^*} \right)$$

luego derivando respecto a  $\theta_{pq}$  se obtiene

$$\frac{\partial \log L}{\partial \theta_{pq}} = \frac{1}{\gamma_{pq}^2} \sum_r \log \left( \frac{d_{pqr} - \theta_{pq}}{d_{pq}^*} \right) \frac{1}{d_{pqr} - \theta_{pq}} + \sum_r \frac{1}{d_{pqr} - \theta_{pq}} = 0$$

de donde, multiplicando y dividiendo por  $d_{pqr} - \theta$  queda:

$$\sum_r \frac{d_{pqr} - \theta_{pq}}{(d_{pqr} - \theta_{pq})^2} \left[ \log \left( \frac{d_{pqr} - \theta_{pq}}{d_{pq}^*} \right) + \gamma_{pq}^2 \right] = 0$$

Descomponiendo y despejando se obtiene finalmente,

$$\theta_{pq} = \frac{\sum_r \frac{d_{pqr}}{(d_{pqr} - \theta_{pq})^2} \left[ \log \left( \frac{d_{pqr} - \theta_{pq}}{d_{pq}^*} \right) + \gamma_{pq}^2 \right]}{\sum_r \frac{1}{(d_{pqr} - \theta_{pq})^2} \left[ \log \left( \frac{d_{pqr} - \theta_{pq}}{d_{pq}^*} \right) + \gamma_{pq}^2 \right]}$$

- Cálculo del estimador  $\hat{\theta}$ .

Ahora, la log-verosimilitud tiene la forma

$$\begin{aligned} \log L = & -\frac{1}{2} \sum_i \sum_j \left[ \frac{1}{\gamma_{ij}^2} S_{ij} + R \log(\gamma_{ij}^2) \right] - \\ & - \sum_i \sum_j \sum_r \log(d_{ijr} - \theta) - M \log(\sqrt{2\pi}) \end{aligned}$$

con,

$$S_{ij} = \sum_r \log^2 \left( \frac{d_{ijr} - \theta}{d_{ij}^*} \right)$$

Derivando nuevamente en la log-verosimilitud, pero ahora respecto de  $\theta$ , se tiene:

$$\frac{\partial \log L}{\partial \theta} = \sum_i \sum_j \frac{1}{\gamma_{ij}^2} \sum_r \frac{1}{d_{ijr} - \theta} \log \left( \frac{d_{ijr} - \theta}{d_{ij}^*} \right) + \sum_i \sum_j \sum_r \frac{1}{d_{ijr} - \theta} = 0$$

Multiplicando y dividiendo en el último sumatorio por  $d_{ijr} - \theta$  y simplificando se obtiene

$$\sum_i \sum_j \sum_r \frac{1}{\gamma_{ij}^2 (d_{ijr} - \theta)} \log \left( \frac{d_{ijr} - \theta}{d_{ij}^*} \right) + \sum_i \sum_j \sum_r \frac{d_{ijr}}{(d_{ijr} - \theta)^2} =$$

$$= \theta \sum_i \sum_j \sum_r \frac{1}{(d_{ijr} - \theta)^2}$$

Finalmente, despejando  $\theta$  en esta expresión y agrupando queda,

$$\theta = \frac{\sum_i \sum_j \sum_r \frac{1}{d_{ijr} - \theta} \left[ \frac{1}{\gamma_{ij}^2} \log \left( \frac{d_{ijr} - \theta}{d_{ij}^*} \right) + \frac{d_{ijr}}{d_{ijr} - \theta} \right]}{\sum_i \sum_j \sum_r \frac{1}{(d_{ijr} - \theta)^2}}$$

C.Q.D.

Respecto a las ecuaciones implícitas que aparecen en el resultado anterior para la estimación del parámetro de localización, solo nos queda por apuntar que su resolución no se puede efectuar independientemente de la estimación de los demás parámetros. Por ello, un procedimiento de resolución puede ser a través de un procedimiento general de estimación, el cual, en cada iteración principal para la obtención de  $X$ , va iterando para resolver cada uno de los parámetros del modelo a través de una aproximación numérica por el método de Scoring de Fisher, asociado a cada ecuación implícita. No obstante, en este trabajo hemos considerado más conveniente la estimación de  $\theta_{pq}$  en virtud de los resultados de Cohen al respecto. Así pues, aunque el resultado general se verá en la sección del MDS con diferencias individuales, se anticipa aquí el siguiente resultado:

**Proposición 3.2.11** *El estimador de máxima verosimilitud del parámetro de localización en el modelo euclídeo, heteroscedástico, lognormal triparamétrico, con parámetro de dispersión  $\gamma_{ij}^2$  asociado a cada par de estímulos  $(i, j)$ , viene determinado por la ecuación implícita:*

$$\left[ \sum_r \frac{1}{d_{pqr} - \theta_{pq}} \right] \left[ \sum_r (d_{pqr} - \theta_{pq}) - \sum_r \log^2(d_{pqr} - \theta_{pq}) + \frac{1}{R} \left( \sum_r \log(d_{pqr} - \theta_{pq}) \right)^2 \right] - R \sum_r \frac{1}{d_{pqr} - \theta_{pq}} \log^2(d_{pqr} - \theta_{pq}) = 0$$

## DEMOSTRACIÓN.

Para obtener este resultado, basta con remitirse al de Cohen que puede verse, por ejemplo, en Crown & Shimizu, [1988]. Así pues, en virtud de dicho resultado, se tiene

$$\frac{\partial \log L}{\partial \log(d_{pq}^*)} = \frac{1}{\gamma_{pq}^2} \sum_r \log \left( \frac{d_{pqr} - \theta_{pq}}{d_{pq}^*} \right) = 0$$

$$\frac{\partial \log L}{\partial \gamma_{pq}} = \frac{R}{\gamma_{pq}} + \frac{1}{\gamma_{pq}^3} \sum_r \log^2 \left( \frac{d_{pqr} - \theta_{pq}}{d_{pq}^*} \right) = 0$$

$$\frac{\partial \log L}{\partial \theta_{pq}} = \frac{1}{\gamma_{pq}^2} \sum_r \frac{1}{d_{pqr} - \theta_{pq}} \log \left( \frac{d_{pqr} - \theta_{pq}^*}{d_{pqr} - \theta_{pq}} \right) + \sum_r \frac{1}{d_{pqr} - \theta_{pq}} = 0$$

Eliminando en esas ecuaciones tanto  $\log(d_{pq}^*)$ , como  $\gamma_{pq}^2$ , se tiene finalmente

$$\left[ \sum_r \frac{1}{d_{pqr} - \theta_{pq}} \right] \left[ \sum_r (d_{pqr} - \theta_{pq}) - \sum_r \log^2(d_{pqr} - \theta_{pq}) + \right. \\ \left. + \frac{1}{R} \left( \sum_r \log(d_{pqr} - \theta_{pq}) \right)^2 \right] - R \sum_r \frac{1}{d_{pqr} - \theta_{pq}} \log^2(d_{pqr} - \theta_{pq}) = 0$$

C.Q.D.

Resolviendo esa ecuación implícita a través de procedimientos numéricos tipo Newton-Raphson, se obtiene un estimador local máximo verosímil de  $\theta_{pq}$ . En el caso de  $\theta$  constante, se procede de igual forma.

Veamos ahora la estimación de la matriz de configuración  $X$  en este modelo. Para ello se realizará la estimación, por un lado sin restricciones y por tanto sujeta a indeterminaciones traslacionales y rotacionales y por otro lado, sujeta a una restricción de fijación de origen, tal y como hacíamos en el caso anterior, para intentar resolver la invarianza traslacional. El resultado análogo al modelo homocedástico puede enunciarse de la siguiente forma.

**Teorema 3.2.12** Consideremos en el modelo euclídeo, heteroscedástico, log-normal triparamétrico, una representación en un espacio de configuración conjunta de dimensión  $K$ . Notemos por  $\gamma_{ij}^2$  al parámetro de dispersión y por  $x_i$  y  $x_j$  a los puntos del espacio de configuración asociados a los estímulos  $i$  y  $j$  respectivamente.

$$x_i = (x_{i1}, \dots, x_{im}, \dots, x_{iK})$$

$$x_j = (x_{j1}, \dots, x_{jm}, \dots, x_{jK})$$

Entonces, el estimador por máxima verosimilitud de la matriz de configuración  $X$ , vendrá dado por el sistema de ecuaciones:

1. Sin restricciones.

$$x_{pq} \sum_j t_{pj} = \sum_j x_{jq} t_{pj}; \quad p = 1, \dots, n. \quad \text{y} \quad q = 1, \dots, K.$$

2. Bajo la restricción de origen dada por,

$$\sum_i x_{im} = 0, \quad m = 1, \dots, K$$

el estimador viene dado a través de:

$$x_{pq} = \frac{\sum_j t_{pj} x_{jq} - \mu_q}{\sum_j t_{pj}}$$

$$p = 1, \dots, n; \quad \text{y} \quad q = 1, \dots, K.$$

donde, para cada  $p$  y cada  $j$ ,

$$t_{pj} = \frac{1}{\gamma_{pj}^2 (d_{pj}^*)^2} \sum_r \left[ \log \left( \frac{d_{pjr} - \theta_{pj}}{d_{pj}^*} \right) + \log \left( \frac{d_{jpr} - \theta_{jp}}{d_{pj}^*} \right) \right]$$

y donde el multiplicador de Lagrange,  $\mu_q$  viene dado para cada  $q$  por

$$\mu_q = \frac{\sum_p \sum_j u_p t_{pj} x_{jq}}{\sum_p u_p}$$

siendo,

$$u_p = \frac{1}{\sum_j t_{pj}}$$

La demostración del teorema es análoga a la del modelo homocedástico y no se hará explícitamente. No obstante, surge una diferencia importante a la hora de la estimación, al considerar como modelo de varianza el modelo multiplicativo que se propone en este trabajo. Debido a la forma de dicho modelo, la estimación de  $X$  dependerá de la varianza, no a través de la variabilidad conjunta, sino a través de la variabilidad particular debida a cada estímulo. Esta ventaja que aporta la estructura multiplicativa del modelo sólo es apreciada de forma explícita, en el caso de la estimación sin restricciones. No obstante, puesto que en muchas ocasiones resulta más preferible obtener una configuración sujeta a invarianza traslacional, que otra que no lo sea, pero que implique un elevado costo computacional en el cálculo del multiplicador lagrangiano, es por lo que las ventajas de la reducción de parámetros en la estimación pueden considerarse de gran interés. Esto se pone de manifiesto en el siguiente resultado.

**Corolario 3.2.13** *Consideremos en el modelo euclídeo, heteroscedástico, log-normal triparamétrico, una representación en un espacio de configuración conjunta de dimensión  $K$ . Notemos por  $x_i$  y  $x_j$  a los puntos de dicho espacio, representantes de los estímulos  $i$  y  $j$  respectivamente.*

$$x_i = (x_{i1}, \dots, x_{im}, \dots, x_{iK})$$

$$x_j = (x_{j1}, \dots, x_{jm}, \dots, x_{jK})$$

*Supongamos igualmente que el parámetro de dispersión del modelo asociado a cada pareja de estímulos  $i$  y  $j$ , que notaremos por  $\gamma_{ij}^2$ , viene a su vez expresado en componentes independientes asociadas a los estímulos según la siguiente estructura:*

$$\gamma_{ij}^2 = (\alpha_i^2 \alpha_j^2)^{1/2}; \quad i, j = 1, \dots, n$$

Entonces, el estimador por máxima verosimilitud de la matriz de configuración  $X$ , vendrá dado por el sistema de ecuaciones:

$$x_{pq} \sum_j t'_{pj} = \sum_j x_{jq} t'_{pj}; \quad p = 1, \dots, n. \quad y \quad q = 1, \dots, K.$$

donde ahora, para cada  $p$  y cada  $j$ ,

$$t'_{pj} = \frac{1}{|\alpha_j|} \frac{1}{(d_{pj}^*)^2} \sum_r \left[ \log \left( \frac{d_{pjr} - \theta_{pj}}{d_{pj}^*} \right) + \log \left( \frac{d_{jpr} - \theta_{jp}}{d_{pj}^*} \right) \right]$$

DEMOSTRACIÓN.

La demostración de este resultado es trivial. Para ello, basta considerar la solución que se obtenía para la variabilidad conjunta, la cual venía dada por

$$x_{pq} \sum_j t_{pj} = \sum_j t_{pj} x_{jq} \quad \forall p, q$$

donde

$$t_{pj} = \frac{1}{\gamma_{pj}^2 (d_{pj}^*)^2} \sum_r \left[ \log \left( \frac{d_{pjr} - \theta_{pj}}{d_{pj}^*} \right) + \log \left( \frac{d_{jpr} - \theta_{jp}}{d_{pj}^*} \right) \right]$$

Si consideramos la descomposición en  $\gamma_{pj}^2$  y definimos

$$t'_{pj} = \frac{1}{|\alpha_j|} \frac{1}{(d_{pj}^*)^2} \sum_r \left[ \log \left( \frac{d_{pjr} - \theta_{pj}}{d_{pj}^*} \right) + \log \left( \frac{d_{jpr} - \theta_{jp}}{d_{pj}^*} \right) \right]$$

se obtiene que,

$$\frac{1}{|\alpha_p^2|} x_{pq} \sum_j t'_{pj} = \frac{1}{|\alpha_p^2|} \sum_j t'_{pj} x_{jq}$$

y simplificando se tiene lo que queríamos.

C.Q.D.

### 3.3 MDS con diferencias individuales.

Para comenzar a la descripción de este tipo de análisis con MDS, en el que se hace hincapié explícitamente en las diferencias entre los individuos a los que se ha pedido su impresión sobre los estímulos, es necesario en primer lugar fijar la notación que se va a seguir a lo largo de la sección. Supongamos en esta ocasión que a  $R$  individuos o sujetos les pedimos que muestren su opinión en  $T$  ocasiones distintas o réplicas, en torno a la desemejanza entre cada una de las parejas que se hayan estudiado de un total de  $n$  estímulos u objetos, a través de algún procedimiento de escala general previamente establecido. Por tanto, al haber introducido en esta ocasión  $T$  réplicas en el experimento, y suponiendo que no existen datos faltantes y por consiguiente, que hayan sido observadas todas las parejas de estímulos, se tendrá un total de  $T \times R$  matrices cuadradas  $n \times n$  de *disimilaridades* entre cada par de estímulos, cuyos elementos notaremos por  $d_{ijrt}$ , con  $i, j = 1, \dots, n$ ,  $r = 1, \dots, R$  y  $t = 1, \dots, T$ . El problema del MDS vuelve a ser el encontrar  $n$  puntos que representen a los  $n$  estímulos, en un espacio de configuración conjunta de *dimensión*  $K$  y cuyas coordenadas se notarán nuevamente por  $x_{im}$ , con  $m = 1, \dots, K$ . La idea por tanto es igualmente métrica, sólo que con la variante de que en esta ocasión las dimensiones del espacio de configuración conjunta, están ponderadas por cada uno de los sujetos que intervienen en el estudio, es decir, se trata de encontrar esos  $n$  puntos, de forma que la distancia entre el punto  $i$  y el punto  $j$ , ponderada por el sujeto  $r$  a través del peso positivo  $w_{rm}$  y que se notará por  $d_{ijr}^*$ , sea lo más parecida posible al valor esperado de las correspondientes disimilaridades observadas  $d_{ijrt}$ .

Este modelo se denomina *Three-Way MDS* o *MDS a tres vías* y puede ser enunciado de la siguiente forma. Consideremos la variable aleatoria  $D_{ijrt}$ , que tomará valores  $d_{ijrt}$  indicando *disimilaridad* entre el estímulo  $i$  y el estímulo  $j$ , apreciada por el sujeto  $r$  en la réplica  $t$ . Por tanto, se considera  $d_{ijr}^*$  como la *distancia* exacta o teórica y sin error y a  $d_{ijrt}$  como la  $t$ -ésima observación de esa distancia o bien como la  $t$ -ésima réplica de la distancia observada a través de la variable aleatoria  $D_{ijrt}$ . De igual forma que en el caso anterior, se supondrá que las observaciones, incluidas las réplicas,  $d_{ijrt}$ , son independientes y están distribuidas según una ley cuya función de densidad notaremos también por  $f$ .

Como consecuencia de lo que se acaba de exponer, la principal diferencia que aporta este modelo con respecto al modelo a dos vías de MDS, consiste

en la posibilidad de ponderar las opiniones individuales ante cada dimensión en el espacio de configuración conjunta. Esta propiedad viene determinada implícitamente por el modelo métrico euclídeo ponderado que se utiliza en este procedimiento, donde la distancia viene determinada según la siguiente expresión:

$$d_{ijr}^* = \left[ \sum_{m=1}^K w_{rm} (x_{im} - x_{jm})^2 \right]^{1/2}; \quad i, j = 1, \dots, n; \quad r = 1, \dots, R.$$

Notaremos por  $W$  a la matriz  $(n \times K)$  de la métrica y  $w_{rm}$  será el peso que indicará la importancia relativa de la dimensión  $m$  para el sujeto  $r$ . Por tanto, bajo esa interpretación, es evidente que el modelo de MDS a dos vías que antes se ha desarrollado puede verse como un caso particular de éste, en el que todas las dimensiones se consideran igualmente ponderadas por los individuos o réplicas implicados,  $w_{rm} = 1, \forall r, m$ .

### 3.3.1 Métricas individualizadas.

Desde un punto de vista geométrico, en el modelo de MDS sin diferencias individuales que veíamos en la sección anterior, el modelo de distancia ajustado se correspondía con un sistema de coordenadas cartesiano ortogonal. Este sistema de coordenadas está formado por  $K$  direcciones en el espacio, perpendiculares entre si. A pesar de ello, en principio, no es necesario que las direcciones utilizadas para localizar los puntos deban ser ortogonales sino que pueden ser oblicuas. El resultado en este caso para  $K$  dimensiones sigue siendo el de  $K$  coordenadas, pero cada coordenada representará ahora la localización de la proyección de un punto en un eje determinado, mediante el desplazamiento desde el punto hasta el eje en una dirección paralela a las demás dimensiones. Así pues, la distancia en un sistema de coordenadas cartesianas oblicuo, asociada al individuo  $r$ , adopta la forma:

$$d_{ijr}^* = \sqrt{\sum_m^K \sum_l^K (x_{im} - x_{jm}) w_{rml} (x_{il} - x_{jl})}$$

Como caso particular, se deduce de ésta la métrica euclídea ortogonal del modelo a dos vías, sin más que considerar  $w_{r11} = \dots = w_{rmm} = \dots = 1$  y

los demás coeficientes nulos. El tamaño de los coeficientes no diagonales indicará el grado de oblicuidad del par de dimensiones correspondiente, pudiendo suponer estos coeficientes simétricos sin pérdida de generalidad (si no, se definen para que ocurra así). Representar esa ecuación en términos matriciales se tiene:

$$d_{ijr}^* = (x_i - x_j)^t W (x_i - x_j)$$

El que la distancia sea positiva asegura el hecho de que  $W$  sea definida positiva y puesto que la suponemos simétrica, se tiene la descomposición de  $W$  en el producto de una matriz triangular y no singular  $T$  por su traspuesta, de la forma  $W = TT^t$ . Puesto que un sistema de coordenadas cartesiano es aquel para el que la matriz de su métrica es  $I$ , veamos que definiendo unas nuevas coordenadas a partir de las viejas, pueden medirse las mismas distancias en un sistema ortogonal. Sea pues  $y_i = T^t x_i$ . Entonces

$$\begin{aligned} (d_{ijr}^*)^2 &= (x_i - x_j)^t TT^t (x_i - x_j) = \\ &= (T^t x_i - T^t x_j)^t (T^t x_i - T^t x_j) = \\ &= (y_i - y_j)^t (y_i - y_j). \end{aligned}$$

Por tanto, puede interpretarse este resultado diciendo que la matriz  $T^t$  es una transformación que lleva las coordenadas de  $x_i$  expresadas en un sistema oblicuo, en las coordenadas de  $y_i$  con respecto a un sistema ortogonal. En esta sección será considerado el estudio para simplificar los cálculos de métricas oblicuas pero cuyos elementos no diagonales son nulos, quedando indeterminados los elementos diagonales. Esto se traduce en que serán consideradas matrices métricas  $W$  diagonales.

Volviendo nuevamente al punto de vista de la interpretación más que al puramente geométrico de la representación, en el modelo que será descrito a continuación, la introducción de factores de ponderación hace que por un lado puedan distinguirse dos espacios de configuración, uno para los estímulos que es conocido como el *espacio de configuración conjunta* y viene dado por la matriz  $X$  y otro para los pesos que viene determinado por la matriz de la métrica  $W$ . Ambos espacios son independientes entre si ya que no pueden representarse conjuntamente en el mismo sistema de referencia puesto que

las escalas son diferentes, a pesar de estar representados ambos en espacios de la misma dimensión. La representación de la matriz de ponderaciones, pone de manifiesto la importancia que se asigna a cada individuo con arreglo a las opiniones efectuadas o bien, si no se estima dicha matriz sino que es suministrada a priori, mostrará la importancia que a dichas opiniones asigna el investigador.

Otra interpretación de la introducción de pesos, radica en el hecho de que si los valores de los factores de ponderación son grandes, pueden interpretarse gráficamente como factores de dilatación de las dimensiones en el espacio de configuración conjunta, mientras que si los valores son pequeños aunque positivos, serán interpretados como factores de acortamiento de las escalas de dichas dimensiones. Por tanto, estos pesos producen para cada fuente de datos de disimilaridad (individuo), un espacio de estímulos que se conoce en la literatura como *privado* o *individual*, mediante otra representación, en una nueva escala de medida para las dimensiones del espacio de configuración, la cual se obtiene tal y como antes se comentaba, aumentando o disminuyendo cada dimensión del espacio de configuración conjunta, a través de la multiplicación de ésta por la raíz cuadrada del factor de ponderación correspondiente. De esa forma, el procedimiento permite el estudio de los llamados espacios de configuración individuales, para enriquecer la interpretación final de la representación.

Por último, consideremos otra característica fundamental que encierra el modelo euclídeo ponderado, referente a la eliminación de indeterminaciones en la representación. Dichas indeterminaciones están relacionadas con el modelo de métrica elegido para la representación de los estímulos en el espacio de configuración conjunta. Para dar una explicación razonada de ésta, veamos por qué se producen.

Una reflexión del  $m$ -ésimo eje del espacio de objetos es simplemente el efecto de utilizar la transformación  $x'_{im} = -x_{im}$ , donde el objeto resultante de la reflexión viene designado por el signo de apóstrofe. Por tanto, una reflexión del  $m$ -ésimo eje del espacio, reemplaza  $(x_{im} - x_{jm})^2$  por  $(x'_{im} - x'_{jm})^2$ , en el  $m$ -ésimo término de la fórmula de distancia, dejando el valor del cómputo total de dicha distancia, invariante. Además, si se añade un valor constante  $c_m$  arbitrario a cada una de las coordenadas de los objetos  $x_{im}$  y  $x_{jm}$ , efectuando por tanto un cambio de origen (o sea una *traslación*) en el  $m$ -ésimo eje, dicha constante es indiferente en el cómputo de  $([x_{im} + c_m] - [x_{jm} + c_m])^2$ , por lo que tampoco afecta una traslación al modelo métrico. De igual forma puede verse que una permutación de ejes tampoco afecta a la métrica.

Veamos por último, como afecta una rotación a éste modelo métrico. Volviendo a la notación matricial, una rotación queda representada mediante el cambio de coordenadas de un vector  $X$  a otro dado por  $T^t X$ , donde  $T$  es una matriz cuadrada que satisface la propiedad  $T^t = T^{-1}$  o sea es ortogonal. Por tanto, para que una métrica permanezca invariante frente a una rotación, debe ocurrir que la matriz de la métrica  $W$  no varíe en la transformación  $T^t W T$ , lo cual ocurre sólo cuando  $W = I$ . Nótese que siempre es posible encontrar una rotación que haga que la métrica sea  $I$ . Esto se especifica, descomponiendo  $W$  en  $VDV^t$ , donde  $V$  es la matriz de vectores propios de  $W$  y  $D$  es la matriz diagonal formada por los valores propios. Por tanto, tomando  $T = V$ , se obtiene como métrica resultante de aplicar esa rotación, una que tiene por matriz, la matriz diagonal  $D$ . Así pues, éste es el motivo de que este modelo resuelva el problema de la invarianza rotacional, por lo que la configuración resultante estará rotacionalmente determinada con respecto a las dimensiones de representación, cosa que no ocurrirá con respecto al origen, reflexiones ni permutaciones de ejes.

Como resumen de todo lo expuesto puede decirse que la razón más importante para la consideración de sistemas de coordenadas cartesianas oblicuos (con matrices métricas  $W$ ), estriba en la posibilidad de que los sujetos puedan generar disimilaridades que puedan ser mejor ajustadas, al permitir que la métrica varíe de sujeto a sujeto. Por supuesto, esto implica que la distancia varíe de sujeto a sujeto, para cualquier par de puntos. Así pues, cualquier dirección en el espacio puede ser interpretada como un punto de vista con respecto a los estímulos. El valor de una coordenada de un punto sobre el correspondiente eje definirá la variación de ese estímulo con respecto a dicho punto de vista. En particular, una dirección que sea usada también como una dimensión para la localización de los puntos, se corresponde de hecho con un punto de vista. Por tanto, de acuerdo con esa interpretación, la raíz cuadrada de los elementos de la diagonal de la matriz de la métrica, definirán la sensibilidad para la variación desde ese punto de vista. Eso es así ya que una diferencia de coordenadas particular, definida con respecto al sistema de referencia estándar, queda multiplicada por ese coeficiente (raíz cuadrada de un valor diagonal) para poder calcular la distancia con la métrica asociada al correspondiente individuo. Si el coeficiente es mayor que uno, la variación en esa dirección contribuirá de forma más decisiva o tenderá a incrementar la distancia relativa respecto al sistema de referencia estándar. Por el contrario, si se obtiene un valor diagonal menor, indicará igualmente una menor contribución en la

variación con respecto a esa dirección. En cuanto a la representación de los ejes, habrá que tener en cuenta la influencia de esos valores diagonales ya que el tamaño de un eje es inversamente proporcional al valor del correspondiente elemento diagonal de la matriz de la métrica.

Por último hay que decir que cuando se obtiene la matriz final de la métrica  $W$ , la presencia de elementos no diagonales y distintos de cero, se traduce en la representación de ejes que no son perpendiculares entre si, de forma que cuanto mayores sean los valores de dichos elementos, mayor será la oblicuidad de los ejes correspondientes.

### 3.3.2 Modelización de la variabilidad residual.

Para el desarrollo de este modelo consideraremos el caso general en que las disimilaridades puedan ser negativas. Además aquí ya tiene pleno sentido el hablar de réplicas bajo una interpretación correcta. Por tanto, todo lo que se comente a continuación se hará en el marco de las condiciones más generales en las que desarrollamos el modelo de MDS probabilístico métrico con diferencias individuales. Así pues, se supondrá que los datos de disimilaridad son obtenidos en su forma más general, bien sea porque provengan de procedimientos de escala generales de tipo intervalo o bien son valores originados por alguna transformación monótona de los datos originales (los cuales en el contexto de la regresión no lineal son conocidos como *disparidades*) o son simplemente datos de similitudes, a los cuales, para convertirlos en disimilaridades, se ha cambiado de signo (Arabie et al. [1988]). De cualquier forma, y sin volver a repetir lo que se expuso en el apartado correspondiente de la sección anterior, parece evidente la necesidad de introducir un parámetro de localización para poder modelizar la distribución de los datos de disimilaridad en un marco probabilístico.

El modelo que se propone en este apartado, engloba los modelos expuestos anteriormente, además de recoger aspectos e intrínsecos a un modelo de MDS con diferencias individuales. Por último, antes de pasar a su desarrollo, hay que apuntar que al igual que en el MDS a dos vías el parámetro de localización dependía de la variable asociada a cada par de estímulos, en este caso dependerá igualmente de la correspondiente variable que irá relacionada además de con cada par de estímulos, con el sujeto o individuo que expresa su evaluación sobre dicho par. Por tanto, el parámetro de localización será en esta ocasión un valor  $\theta_{ijr}$ , con  $i, j = 1, \dots, n$  y con  $r = 1, \dots, R$ . La introducción de este

parámetro de localización así como la flexibilidad que se desprende de las posibilidades de su utilización hacen que el modelo triparamétrico sea aun más atractivo en el caso de MDS a tres vías que en el modelo simple. Veamos algunos aspectos de interés que nos llevan a esa conclusión.

El análisis de datos en general y MDS no constituye una excepción, sufre en general de un alto grado de dispersión debido a la variabilidad introducida por la componente individual o sea, se produce mucha dispersión en los datos analizados debida a factores que dependen exclusivamente de los individuos y no del contexto general del experimento. En muchas ocasiones, los individuos utilizan únicamente las categorías pares o las impares y otras veces sólo las categorías centrales de la escala o bien las más alejadas. Esta forma de respuesta resulta en general difícil de controlar e induce en principio a que sea necesario controlar la variabilidad del experimento con factores que dependan entre otros, de los individuos que intervienen en el estudio. Por esa razón, independientemente de que aparezca un parámetro exclusivamente asociada a ella en la estructura de la varianza, surge la idea de si lo realmente interesante no es tanto que el parámetro de localización dependa formalmente de los mismos índices que la media de los datos a estimar o si debería estimarse un parámetro de localización asociado a cada individuo exclusivamente. Con ese fin, puede que en un análisis con este modelo, resulte más interesante estimar un parámetro  $\theta_r$ , o bien tomar,

$$\theta_r = \min_{ij} \{\theta_{ijr} \mid i, j = 1, \dots, n, \quad r = 1, \dots, R\}$$

Puesto que la naturaleza de los valores de disimilaridad no ha cambiado, puede suponerse como distribución inicial la distribución lognormal triparamétrica, con arreglo al modelo:

$$D_{ijrt} \sim \Lambda(\theta_{ijr}, \log(d_{ijr}^*), \sigma_{ijrt}^2)$$

Además de lo dicho anteriormente sobre el parámetro de localización y de la necesidad de que éste dependa, al igual que la media, de los tres índices de variación, hay que hacer un comentario en torno a la figura de la varianza, la cual también es considerada dependiente de cada pareja de estímulos, de cada individuo en cuestión y de cada réplica estudiada. Después se descompondrá la varianza del modelo, bajo una estructura multiplicativa entre variabilidad debida al individuo y variabilidad debida al par, de modo análogo al caso lognormal biparamétrico en Ramsay [1982]. Además, la variabilidad debida a

cada par será igualmente descompuesta bajo dos modelos: el aditivo propuesto por Ramsay [1978] y el multiplicativo propuesto en este trabajo y que es más natural, al estar en consonancia con la estructura general considerada sobre la varianza. Así pues, bajo las hipótesis que acaban de enunciarse y considerando los puntos de vista expuestos en la sección anterior, se tiene una densidad,

$$f(d_{ijrt} | \theta_{ijr}, d_{ijr}^*, \sigma_{ijrt}^2) = \frac{1}{\sigma_{ijrt}(d_{ijrt} - \theta_{ijr})\sqrt{2\pi}} \exp \left\{ \frac{-\log^2 \left( \frac{d_{ijrt} - \theta_{ijr}}{d_{ijr}^*} \right)}{2\sigma_{ijrt}^2} \right\}$$

$$d_{ijrt} > \theta_{ijr}; \quad i, j = 1, \dots, n; \quad r = 1, \dots, R; \quad t = 1, \dots, T.$$

Las propiedades que pueden asignarse a esta distribución son análogas al caso a dos vías, pudiendo resumirse:

- Asocia probabilidades positivas a todos los valores  $d_{ijrt}$  de  $D_{ijrt}$  mayores que  $\theta_{ijr}$ . Esto hay que entenderlo como que el valor mínimo fijado en la escala depende de cada individuo y de cada par de estímulos y no de la réplica del experimento, por tanto, la distribución estará definida para aquellos valores en consonancia con el tipo de dato analizado.
- El valor esperado de la variable se aproxima a  $d_{ijr}^*$  para cada réplica de la forma:

$$E[D_{ijrt}] = \theta_{ijr} + d_{ijr}^* e^{\frac{\sigma_{ijrt}^2}{2}}$$

- Su desviación típica es proporcional a la distancia teórica deseada.

$$\text{Var}[D_{ijrt}] = (d_{ijr}^*)^2 [e^{\sigma_{ijrt}^2} (e^{\sigma_{ijrt}^2} - 1)]$$

El empleo del modelo que acabamos de exponer junto con la estimación bajo máxima verosimilitud de los parámetros y en algunos casos, bajo restricciones, constituirá el fundamento del modelo que vamos a desarrollar.

En vista de lo anterior, la verosimilitud para este modelo Euclídeo ponderado, heteroscedástico, lognormal triparamétrico, viene dada por la expresión:

$$L_{\{d_{ijrt}\}}(\theta_{ijr}, \sigma_{ijrt}^2, X) = \\ = \prod_i \prod_j \prod_r \prod_t \frac{1}{\sigma_{ijrt}(d_{ijrt} - \theta_{ijr})\sqrt{2\pi}} \exp \left\{ \frac{-\log^2 \left( \frac{d_{ijrt} - \theta_{ijr}}{d_{ijr}^*} \right)}{2\sigma_{ijrt}^2} \right\}$$

donde los productos se supondrán lógicamente entre los datos realmente observados. La log-verosimilitud quedará,

$$\log L_{\{d_{ijrt}\}}(\theta_{ijr}, \sigma_{ijrt}^2, \{x_{pm}\}_{pm}) = \\ = -\frac{1}{2} \sum_i \sum_j \sum_r \sum_t \left[ \frac{1}{\sigma_{ijrt}^2} \log^2 \left( \frac{d_{ijrt} - \theta_{ijr}}{d_{ijr}^*} \right) + \log(\sigma_{ijrt}^2) \right] - \\ - \sum_i \sum_j \sum_r \sum_t \log(d_{ijrt} - \theta_{ijr}) - M \log(\sqrt{2\pi})$$

donde,  $M$  es el número de datos observados.

A partir de ahora, supondremos el modelo más general en el que la variabilidad queda descompuesta en términos independientes, cada uno de los cuales refleja un aspecto importante de la variabilidad, que el investigador puede elegir exteriorizar para una optimización del análisis. En este trabajo se pondrá de manifiesto la posibilidad de descomponer la variabilidad del modelo en tres componentes:

- Una componente,  $\gamma_{ij}$ , referente a a cada par de estímulos  $(i, j)$  estudiados y que ha sido ampliamente comentada junto con sus posibles variantes, en el apartado que en la sección anterior se dedicaba a la estimación en el modelo euclídeo heteroscedástico triparamétrico.
- Otra componente referente a la parte de la variación debida exclusivamente a los individuos implicados. Esta cantidad que será notada por  $\sigma_r$ , representa el error típico dentro de cada sujeto  $r$ . Permitir que esa componente varíe de sujeto a sujeto implica que la variación en cada

residuo o el tamaño de la desviación típica para la mejor distancia ajustada, dependa de cada sujeto. Generalmente, puede comprobarse tras el estudio de los residuos después de un análisis con MDS, que algunos sujetos tienden a producir juicios de disimilaridad que se desvían relativamente poco de las distancias ajustadas, mientras que otros tienden a producir desviaciones relativamente grandes. Por tanto, la componente de la varianza asociada a los sujetos,  $\sigma_r^2$ , puede interpretarse como un índice de la relativa falta de precisión (desde el punto de vista del modelo) de los datos de un cierto sujeto. Así pues, valores relativamente grandes pueden indicar que el sujeto debería ser eliminado de la muestra, en un escrutinio general para el análisis. También, un valor grande podría indicar un fallo a la hora de entender las instrucciones para el juicio por parte del individuo, un cambio del punto de vista durante la clasificación o simplemente un fallo al juzgar las disimilaridades apropiadamente.

- Por último, en este trabajo se propone un nuevo factor de variabilidad que será notado por  $\tau_t$  y que recoge la filosofía del parámetro  $\sigma_r$  pero referido a la réplica  $t$ . Este parámetro auna aquellos factores de variación relativos a cada réplica y puede entenderse como el error típico dentro de dicha réplica. Puede por tanto interpretarse de forma que si su valor es relativamente alto, induciría a reconsiderar en el estudio general a dicha réplica. Así pues, a través del estudio de ese parámetro (que puede estimarse a la vez que los demás o a posteriori una vez estimados los otros) puede eliminarse de la muestra, aquellas réplicas que por cualquier razón han presentado una mayor variabilidad en el modelo y por tanto, han inducido el incremento de los otros dos tipos de parámetros de variabilidad antes expuestos, sin que la conclusión deba ser la de que el modelo no se ajusta a los datos. Por tanto, este parámetro puede ser también interpretado como un parámetro de validación del modelo en cuanto a la selección de réplicas, pudiendo ser considerado tanto como una variable (según acabamos de exponer), como un valor constante igual a uno, según las peculiaridades del análisis en cuestión.

En virtud de esa descomposición, la log-verosimilitud asociada a la muestra  $\{d_{ijrt}\}_{ijrt}$ , adopta la expresión,

$$\log L = \log L_{\{d_{ijrt}\}}(\theta_{ijr}, \sigma_r^2, \gamma_{ij}^2, \tau_t^2, \{x_{pm}\}_{pm}) =$$

$$\begin{aligned}
&= -\frac{1}{2} \sum_i \sum_j \frac{1}{\gamma_{ij}^2} \sum_r \frac{1}{\sigma_r^2} \sum_t \frac{1}{\tau_t^2} \log^2 \left( \frac{d_{ijrt} - \theta_{ijr}}{d_{ijr}^*} \right) - \\
&-\frac{RT}{2} \sum_i \sum_j \log(\gamma_{ij}^2) - \frac{n^2 T}{2} \sum_r \log(\sigma_r^2) - \frac{n^2 R}{2} \sum_t \log(\tau_t^2) - \\
&-\sum_i \sum_j \sum_r \sum_t \log(d_{ijrt} - \theta_{ijr}) - M \log(\sqrt{2\pi})
\end{aligned}$$

siendo  $M$  el número total de disimilaridades observadas.

Veamos la estimación de los parámetros que intervienen en el modelo y que nos dará la solución al problema de encontrar la mejor matriz de configuración  $X$  de forma que las distancias entre sus filas sea lo más parecida posible a los valores de disimilaridad observados.

### 3.3.3 Estimación de las componentes de la varianza.

En este apartado se van a estimar, tanto las componentes principales de la descomposición de la variabilidad de los residuos que antes se han descrito, como las componentes asociadas a los estímulos en la variabilidad dentro de cada par, bajo las dos estructuras manejadas en este trabajo; la aditiva y la multiplicativa. Veamos en primer lugar la estimación de la componente asociada a los individuos,  $\sigma_s^2$ .

**Proposición 3.3.14** *El estimador por máxima verosimilitud de la componente individual de la varianza de los residuos  $\sigma_s^2$ , en el modelo euclídeo, heteroscedástico, lognormal triparamétrico, con diferencias individuales, es:*

$$\hat{\sigma}_s^2 = \frac{1}{n^2 T} \sum_i \sum_j \sum_t \frac{1}{\gamma_{ij}^2 \tau_t^2} \log^2 \left( \frac{d_{ijst} - \theta_{ijs}}{d_{ijs}^*} \right)$$

DEMOSTRACIÓN.

Reagrupando en la expresión general de la log-verosimilitud asociada a la muestra  $\{d_{ijrt}\}$ , se consigue la expresión:

$$\begin{aligned} \log L = & -\frac{1}{2} \sum_r \left[ \frac{1}{\sigma_r^2} \sum_i \sum_j \sum_t \frac{1}{\gamma_{ij}^2 \tau_t^2} \log^2 \left( \frac{d_{ijrt} - \theta_{ijr}}{d_{ijr}^*} \right) + n^2 T \log(\sigma_r^2) \right] - \\ & - \frac{RT}{2} \sum_i \sum_j \log(\gamma_{ij}^2) - \frac{n^2 R}{2} \sum_t \log(\tau_t^2) - \\ & - \sum_i \sum_j \sum_r \sum_t \log(d_{ijrt} - \theta_{ijr}) - M \log(\sqrt{2\pi}) \end{aligned}$$

Derivando respecto del valor  $\sigma_s^2$  e igualando a cero, queda la expresión,

$$-\frac{1}{\sigma_s^4} \sum_i \sum_j \sum_t \frac{1}{\gamma_{ij}^2 \tau_t^2} \log^2 \left( \frac{d_{ijst} - \theta_{ijs}}{d_{ijs}^*} \right) + \frac{n^2 T}{\sigma_s^2} = 0$$

de donde simplificando y despejando, se tiene finalmente que:

$$\sigma_s^2 = \frac{1}{n^2 T} \sum_i \sum_j \sum_t \frac{1}{\gamma_{ij}^2 \tau_t^2} \log^2 \left( \frac{d_{ijst} - \theta_{ijs}}{d_{ijs}^*} \right)$$

C.Q.D.

El estimador por máxima verosimilitud que acaba de obtenerse, se simplifica considerablemente cuando se suponen los demás factores de la varianza constantemente iguales a uno. Esto sucede en muchos análisis en los que al investigador le interesa poner de manifiesto exclusivamente la variabilidad entre sujetos y no considera las restantes componentes. Por tanto, en esos casos, el estimador de  $\sigma_s^2$  no resulta excesivamente complicado de obtener según la expresión:

$$\sigma_s^2 = \frac{1}{n^2 T} \sum_i \sum_j \sum_t \log^2 \left( \frac{d_{ijst} - \theta_{ijs}}{d_{ijs}^*} \right)$$

Esta razón que se acaba de exponer sobre mantener con valor uno las componentes entre estímulos, así como las componentes asociadas a las réplicas y a posteriori ofrecer sus correspondientes valores, es el motivo principal que induce a que en la estimación de los factores restantes, en la estructura principal de la varianza de los residuos, haya que considerar restricciones de escala

que permitan estimar un valor interpretable. Así pues, en todas las estimaciones que se consideren a continuación habrá que destacar en la estimación bajo restricciones, estando formadas dichas restricciones por ecuaciones que impongan el hecho de que el valor promedio de cada estimador sea igual a uno (que es el valor elegido, en caso de considerarlas constantes y por tratarse de una estructura multiplicativa de la varianza). No obstante y debido a que en muchos casos el costo computacional de dichas estimaciones restringidas es elevado y que en muchas situaciones prácticas, la estimación por separado de las componentes introduce demasiados parámetros en el modelo, resulta necesario ofrecer igualmente el estimador por máxima verosimilitud no restringido.

**Proposición 3.3.15** *El estimador por máxima verosimilitud del factor  $\gamma_{pq}^2$  en el modelo euclídeo, heteroscedástico, lognormal triparamétrico con diferencias individuales, viene dado por las ecuaciones implícitas siguientes:*

1. Sin restricciones.

$$\hat{\gamma}_{pq}^2 = \frac{1}{2RT} \sum_r \sum_t \frac{1}{\sigma_r^2 \tau_t^2} \left[ \log^2 \left( \frac{d_{pqrt} - \theta_{pqr}}{d_{pqr}^*} \right) + \log^2 \left( \frac{d_{qprt} - \theta_{qpr}}{d_{pqr}^*} \right) \right]$$

2. Consideremos el sistema de ecuaciones de restricción:

$$\sum_i \sum_j \gamma_{ij}^2 = n^2$$

Bajo esas restricciones, el estimador  $\hat{\gamma}_{pq}^2$  viene dado por:

$$\frac{-2RT + \left[ 4R^2 T^2 + 4\lambda \sum_r \sum_t \left( \log^2 \left( \frac{d_{pqrt} - \theta_{pqr}}{d_{pqr}^*} \right) + \log^2 \left( \frac{d_{qprt} - \theta_{qpr}}{d_{pqr}^*} \right) \right) \right]^{1/2}}{2\lambda}$$

donde

$$\lambda = \frac{1}{n^2} \sum_{i,j,r,t} \frac{1}{\sigma_r^2 \gamma_{ij}^2 \tau_t^2} \left[ \log^2 \left( \frac{d_{ijrt} - \theta_{ijr}}{d_{ijr}^*} \right) + \log^2 \left( \frac{d_{jirt} - \theta_{jir}}{d_{ijr}^*} \right) \right] - 2RT$$

DEMOSTRACIÓN.

- Estimación sin restricciones.

La log-verosimilitud adopta en este caso la expresión

$$\begin{aligned} \log L = & -\frac{1}{2} \sum_i \sum_j \left[ \frac{1}{\gamma_{ij}^2} \sum_r \sum_t \frac{1}{\sigma_r^2 \tau_t^2} \log^2 \left( \frac{d_{ijrt} - \theta_{ijr}}{d_{ijr}^*} \right) + RT \log(\gamma_{ij}^2) \right] - \\ & - \frac{n^2 T}{2} \sum_r \log(\sigma_r^2) - \frac{n^2 R}{2} \sum_t \log(\tau_t^2) - \\ & - \sum_i \sum_j \sum_r \sum_t \log(d_{ijrt} - \theta_{ijr}) - M \log(\sqrt{2\pi}) \end{aligned}$$

Derivando, respecto de un valor  $\gamma_{pq}^2$  e igualando a cero la expresión resultante se tiene:

$$\frac{-1}{\gamma_{pq}^4} \sum_r \frac{1}{\sigma_r^2} \sum_t \frac{1}{\tau_t^2} \left[ \log^2 \left( \frac{d_{pqrt} - \theta_{pqr}}{d_{pqr}^*} \right) + \log^2 \left( \frac{d_{qprr} - \theta_{qpr}}{d_{pqr}^*} \right) \right] + \frac{2RT}{\gamma_{pq}^2} = 0$$

y despejando, se obtiene el resultado

$$\gamma_{pq}^2 = \frac{1}{2RT} \sum_r \sum_t \frac{1}{\sigma_r^2 \tau_t^2} \left[ \log^2 \left( \frac{d_{pqrt} - \theta_{pqr}}{d_{pqr}^*} \right) + \log^2 \left( \frac{d_{qprr} - \theta_{qpr}}{d_{pqr}^*} \right) \right]$$

- Con restricciones.

Vamos a calcular el estimador a través del método de los multiplicadores de Lagrange. Para ello, en esta ocasión, se maximiza la función:

$$Q(\gamma_{ij}^2) = \log L - \frac{1}{4} \lambda \left( \sum_i \sum_j \gamma_{ij}^2 - n^2 \right)$$

Para ello, calculando las parciales respecto al parámetro y respecto a  $\lambda$  e igualando a cero, se tiene:

$$\frac{\partial Q}{\partial \gamma_{pq}^2} = 0$$

de donde resulta la ecuación,

$$\frac{-1}{\gamma_{pq}^4} \sum_r \sum_t \frac{1}{\sigma_r^2 \tau_t^2} \left[ \log^2 \left( \frac{d_{pqrt} - \theta_{pqr}}{d_{pqr}^*} \right) + \log^2 \left( \frac{d_{qprt} - \theta_{qpr}}{d_{pqr}^*} \right) \right] + \frac{2RT}{\gamma_{pq}^2} + \lambda = 0$$

y por otro lado,

$$\frac{\partial Q}{\partial \lambda} = \sum_i \sum_j \gamma_{pq}^2 = n^2$$

Si multiplicamos en la primera ecuación por  $\gamma_{pq}^2$  queda:

$$\lambda \gamma_{pq}^4 + 2RT \gamma_{pq}^2 - \sum_r \sum_t \frac{1}{\sigma_r^2 \tau_t^2} \left[ \log^2 \left( \frac{d_{pqrt} - \theta_{pqr}}{d_{pqr}^*} \right) + \log^2 \left( \frac{d_{qprt} - \theta_{qpr}}{d_{pqr}^*} \right) \right] = 0$$

y resolviendo el sistema obtenemos el valor de  $\gamma_{pq}^2$  siguiente

$$\frac{-2RT + \left[ 4R^2 T^2 + 4\lambda \sum_r \sum_t \left( \log^2 \left( \frac{d_{pqrt} - \theta_{pqr}}{d_{pqr}^*} \right) + \log^2 \left( \frac{d_{qprt} - \theta_{qpr}}{d_{pqr}^*} \right) \right) \right]^{1/2}}{2\lambda}$$

Veamos como se resuelve para  $\lambda$ . Para ello, considerando la primera ecuación y multiplicandola por  $\gamma_{pq}^2$ , queda:

$$2RT + \lambda \gamma_{pq}^2 = \sum_r \sum_t \frac{1}{\sigma_r^2 \gamma_{pq}^2 \tau_t^2} \left[ \log^2 \left( \frac{d_{pqrt} - \theta_{pqr}}{d_{pqr}^*} \right) + \log^2 \left( \frac{d_{qprt} - \theta_{qpr}}{d_{pqr}^*} \right) \right]$$

Si sumamos esa expresión en  $p$  y  $q$  se tiene

$$2RT n^2 + \lambda n^2 = \sum_{p,q,r,t} \frac{1}{\sigma_r^2 \gamma_{pq}^2 \tau_t^2} \left[ \log^2 \left( \frac{d_{pqrt} - \theta_{pqr}}{d_{pqr}^*} \right) + \log^2 \left( \frac{d_{qprt} - \theta_{qpr}}{d_{pqr}^*} \right) \right]$$

y despejando  $\lambda$  queda

$$\lambda = \frac{1}{n^2} \sum_{i,j,r,t} \frac{1}{\sigma_r^2 \gamma_{ij}^2 \tau_t^2} \left[ \log^2 \left( \frac{d_{ijrt} - \theta_{ijr}}{d_{ijr}^*} \right) + \log^2 \left( \frac{d_{jirt} - \theta_{jir}}{d_{ijr}^*} \right) \right] - 2RT$$

C.Q.D.

Tratamos ahora de ofrecer la estimación de las componentes de la variabilidad de los estímulos, bajo las dos estructuras presentadas en el modelo sin diferencias individuales. Por tanto, consideraremos dos descomposiciones de  $\gamma_{ij}^2$  atendiendo a una estructura aditiva y a una estructura multiplicativa. Independientemente del modelo a considerar, es necesaria la estimación de estas componentes bajo restricciones derivadas de la restricción que se aplicaba a  $\gamma_{pq}^2$ . Así pues, se tendrá que:

1. Modelo aditivo.

Este modelo descompone la variabilidad entre estímulos de la forma,

$$\gamma_{ij}^2 = \frac{\alpha_i^2 + \alpha_j^2}{2}$$

por tanto, la relación de equivalencia entre restricciones es:

$$\frac{1}{n^2} \sum_i \sum_j \gamma_{ij}^2 = 1 \Leftrightarrow \sum_i \alpha_i^2 = n$$

2. Modelo multiplicativo.

En este modelo, la variabilidad se descompone de la forma:

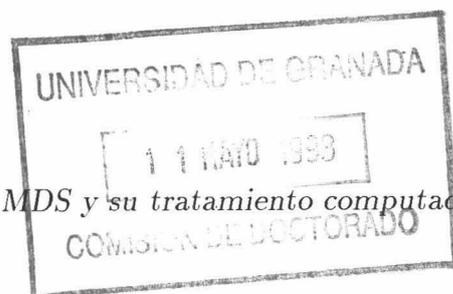
$$\gamma_{ij}^2 = \sqrt{\alpha_i^2 \alpha_j^2}$$

Por tanto, la restricción en este caso queda:

$$\frac{1}{n^2} \sum_i \sum_j \gamma_{ij}^2 = 1 \Leftrightarrow \sum_i (\alpha_i^2)^{1/2} = n$$

Utilizando dichas restricciones, puede enumerarse el siguiente resultado.

**Proposición 3.3.16** *El estimador por máxima verosimilitud de los errores típicos de los estímulos en la varianza de los residuos, bajo la restricción de escala asociada, viene expresado a través de la siguiente ecuación:*



## 1. Modelo aditivo.

$$\hat{\alpha}_p^2 = \frac{\sum_j \frac{1}{4\gamma_{pj}^4} \sum_r \sum_t \frac{1}{\sigma_r^2 \tau_t^2} \left[ \log^2 \left( \frac{d_{pjrt} - \theta_{pjr}}{d_{pjr}^*} \right) + \log^2 \left( \frac{d_{jp rt} - \theta_{jpr}}{d_{pjr}^*} \right) \right]}{\sum_j \frac{RT}{4\gamma_{pj}^4}} +$$

$$+ \frac{\sum_j \frac{RT \alpha_j^2}{4\gamma_{pj}^4} - \lambda}{\sum_j \frac{RT}{4\gamma_{pj}^4}}$$

donde

$$\lambda = \frac{1}{n} \sum_p \sum_j \frac{1}{4\gamma_{pj}^4} \sum_r \sum_t \frac{1}{\sigma_r^2 \tau_t^2} \left[ \log^2 \left( \frac{d_{pjrt} - \theta_{pjr}}{d_{pjr}^*} \right) \log^2 \left( \frac{d_{jp rt} - \theta_{jpr}}{d_{pjr}^*} \right) \right] -$$

$$- \sum_p \sum_j \frac{RT}{2\gamma_{pj}^2}$$

## 2. Modelo multiplicativo.

$$\hat{\alpha}_p^2 = \frac{\sum_j \frac{1}{|\alpha_j|} \sum_r \sum_t \frac{1}{\sigma_r^2 \tau_t^2} \left[ \log^2 \left( \frac{d_{pjrt} - \theta_{pjr}}{d_{pjr}^*} \right) \log^2 \left( \frac{d_{jp rt} - \theta_{jpr}}{d_{pjr}^*} \right) \right]}{\frac{1}{n} \sum_p \sum_j \frac{1}{\gamma_{pj}^2} \sum_r \sum_t \frac{1}{\sigma_r^2 \tau_t^2} \left[ \log^2 \left( \frac{d_{pjrt} - \theta_{pjr}}{d_{pjr}^*} \right) \log^2 \left( \frac{d_{jp rt} - \theta_{jpr}}{d_{pjr}^*} \right) \right]}$$

DEMOSTRACIÓN.

## • Modelo aditivo.

Considerando nuevamente la log-verosimilitud de la forma

$$\begin{aligned} & \log L(\{\theta_{ij}\}, \{\alpha_i^2\}, \{\sigma_r^2\}, \{\gamma_{ij}^2\}, \{x_{im}\}) = \\ & = -\frac{1}{2} \sum_i \sum_j \left[ \frac{2}{\alpha_i^2 + \alpha_j^2} \sum_r \sum_t \frac{1}{\sigma_r^2 \tau_t^2} \log^2 \left( \frac{d_{ijrt} - \theta_{ijr}}{d_{ijr}^*} \right) + RT \log(\alpha_i^2 + \alpha_j^2) \right] - \\ & \quad - M \log(2) - \frac{n^2 T}{2} \sum_r \log(\sigma_r^2) - \frac{n^2 R}{2} \sum_t \log(\tau_t^2) - \\ & \quad - \sum_i \sum_j \sum_r \sum_t \log(d_{ijrt} - \theta_{ijr}) - M \log(\sqrt{2\pi}) \end{aligned}$$

La función a maximizar será

$$Q(\alpha_p^2) = \log L - \lambda \left( \sum_i \alpha_i^2 - n \right)$$

Derivando  $Q$  respecto de  $\alpha_p^2$  y respecto de  $\lambda$  se obtienen las ecuaciones de verosimilitud dadas por,

$$\begin{aligned} \frac{\partial Q}{\partial \alpha_p^2} \sum_j \frac{(S_{pj} + S_{jp})}{(\alpha_p^2 + \alpha_j^2)^2} - \sum_j \frac{RT}{\alpha_p^2 + \alpha_j^2} - \lambda &= 0 \\ \frac{\partial Q}{\partial \lambda} = \sum_i \alpha_i^2 &= n \end{aligned}$$

donde

$$S_{pj} = \sum_r \sum_t \frac{1}{\sigma_r^2 \tau_t^2} \log^2 \left( \frac{d_{pjrt} - \theta_{pjr}}{d_{pjr}^*} \right)$$

Sumando en la primera ecuación en  $p$  y despejando  $\lambda$  se tiene que

$$\begin{aligned} \lambda &= \frac{1}{n} \sum_p \sum_j \frac{1}{4\gamma_{pj}^4} \sum_r \sum_t \frac{1}{\sigma_r^2 \tau_t^2} \left[ \log^2 \left( \frac{d_{pjrt} - \theta_{pjr}}{d_{pjr}^*} \right) \log^2 \left( \frac{d_{jp rt} - \theta_{jpr}}{d_{pjr}^*} \right) \right] - \\ & \quad - \sum_p \sum_j \frac{RT}{2\gamma_{pj}^2} \end{aligned}$$

Si en la primera ecuación de verosimilitud, despejamos  $\alpha_p$ , se obtendrá igualmente el valor deseado.

• Modelo multiplicativo.

En este modelo, la log-verosimilitud tiene la forma:

$$\begin{aligned} \log L = & -\frac{1}{2} \sum_i \sum_j \left[ \frac{1}{(\alpha_i^2 \alpha_j^2)^{1/2}} S_{ij} + RT \log\{(\alpha_i^2 \alpha_j^2)^{1/2}\} \right] - \\ & -\frac{n^2 T}{2} \sum_r \log(\sigma_r^2) - \frac{n^2 R}{2} \sum_t \log(\tau_t^2) - \\ & - \sum_i \sum_j \sum_r \sum_t \log(d_{ijrt} - \theta_{ijr}) - M \log(\sqrt{2\pi}) \end{aligned}$$

Tendremos que maximizar la función

$$Q(\alpha_p^2) = \log L - \frac{1}{2} \lambda \left( \sum_i (\alpha_i^2)^{1/2} - n \right)$$

Las ecuaciones de verosimilitud quedarán en este caso como:

$$\begin{aligned} \frac{\partial Q}{\partial \alpha_p^2} = \frac{1}{\alpha_p^2} \sum_j \frac{(S_{pj} + S_{jp})}{(\alpha_j^2)^{1/2}} - 2RTn - \lambda = 0 \\ \frac{\partial Q}{\partial \lambda} = \sum_i (\alpha_i^2)^{1/2} = n \end{aligned}$$

Si se multiplica la primera ecuación por  $\alpha_p^2$ , despejando se obtiene

$$\alpha_p^2 = \frac{\sum_j \frac{(S_{pj} + S_{jp})}{|\alpha_j|}}{2RTn + \lambda}$$

Multiplicando en la primera ecuación por  $(\alpha_p^2)^{1/2}$  y sumando en  $p$  queda

$$\lambda = \frac{1}{n} \sum_p \sum_j \frac{(S_{pj} + S_{jp})}{\gamma_{pj}^2} - 2RTn$$

Sustituyendo cada  $S_{pj}$  y  $S_{jp}$  por su valor y  $\lambda$  en la expresión de  $\alpha_p^2$ , se tiene lo deseado.

C.Q.D.

Veamos por último la estimación de la componente asociada a la variabilidad de las réplicas.

**Proposición 3.3.17** *El estimador de máxima verosimilitud del factor de variabilidad asociado a cada réplica, en el modelo euclídeo, heteroscedástico, log-normal triparamétrico con diferencias individuales, vendrá dado por el sistema de ecuaciones:*

1. Sin restricciones.

$$\hat{\tau}_u^2 = \frac{1}{n^2 R} \sum_i \sum_j \sum_r \frac{1}{\gamma_{ij}^2 \sigma_r^2} \log^2 \left( \frac{d_{ijru} - \theta_{ijr}}{d_{ijr}^*} \right)$$

2. Con restricciones.

$$\hat{\tau}_u^2 = \frac{-n^2 R + \left[ n^4 R^2 + 4\lambda \sum_i \sum_j \sum_r \frac{1}{\sigma_r^2 \gamma_{ij}^2} \log^2 \left( \frac{d_{ijru} - \theta_{ijr}}{d_{ijr}^*} \right) \right]^{1/2}}{2\lambda}$$

donde

$$\lambda = \frac{1}{T} \left[ \sum_i \sum_j \sum_r \sum_t \frac{1}{\sigma_{ijrt}^2} \log^2 \left( \frac{d_{ijrt} - \theta_{ijr}}{d_{ijr}^*} \right) \right] - n^2 R$$

DEMOSTRACIÓN.

En este caso, la log-verosimilitud quedará de la forma:

$$\log L = -\frac{1}{2} \sum_t \left[ \frac{1}{\tau_t^2} \sum_i \sum_j \sum_r \frac{1}{\gamma_{ij}^2 \sigma_r^2} \log^2 \left( \frac{d_{ijrt} - \theta_{ijr}}{d_{ijr}^*} \right) + n^2 R \log(\tau_t^2) \right] -$$

$$-\frac{n^2 T}{2} \sum_r \log(\sigma_r^2) - \frac{RT}{2} \sum_i \sum_j \log(\gamma_{ij}^2) - \\ - \sum_i \sum_j \sum_r \sum_t \log(d_{ijrt} - \theta_{ijr}) - M \log(\sqrt{2\pi})$$

- Sin restricciones.

En este caso, sin más que derivar respecto a  $\tau_u^2$  se tiene,

$$\frac{-1}{\tau_u^4} \sum_i \sum_j \sum_r \frac{1}{\gamma_{ij}^2 \sigma_r^2} \log^2 \left( \frac{d_{ijru} - \theta_{ijr}}{d_{ijr}^*} \right) + n^2 R \frac{1}{\tau_u^2} = 0$$

simplificando y despejando queda finalmente:

$$\tau_u^2 = \frac{1}{n^2 R} \sum_i \sum_j \sum_r \frac{1}{\gamma_{ij}^2 \sigma_r^2} \log^2 \left( \frac{d_{ijru} - \theta_{ijr}}{d_{ijr}^*} \right)$$

- Con restricciones.

Procediendo de forma análoga a los casos anteriores, se tiene que las ecuaciones de verosimilitud quedan:

$$\frac{-1}{\tau_u^4} \sum_i \sum_j \sum_r \frac{1}{\sigma_r^2 \gamma_{ij}^2} \log^2 \left( \frac{d_{ijru} - \theta_{ijr}}{d_{ijr}^*} \right) + n^2 R \frac{1}{\tau_u^2} + \lambda = 0 \\ \sum_t \tau_t^2 = T$$

Tomando la primera ecuación, multiplicándola por  $\tau_u^4$  y despejando queda

$$\hat{\tau}_u^2 = \frac{-n^2 R + \left[ n^4 R^2 + 4\lambda \sum_i \sum_j \sum_r \frac{1}{\sigma_r^2 \gamma_{ij}^2} \log^2 \left( \frac{d_{ijru} - \theta_{ijr}}{d_{ijr}^*} \right) \right]^{1/2}}{2\lambda}$$

Si multiplicamos en la primera ecuación por  $\tau_u^2$  y sumamos en  $u$ , se obtendrá de ahí el valor deseado de  $\lambda$ .

$$\lambda = \frac{1}{T} \left[ \sum_i \sum_j \sum_r \sum_t \frac{1}{\sigma_{ijrt}^2} \log^2 \left( \frac{d_{ijrt} - \theta_{ijr}}{d_{ijr}^*} \right) \right] - n^2 R$$

C.Q.D.

### 3.3.4 Estimación de la Configuración.

En este apartado se ofrece la estimación de la matriz de configuración  $X$  así como la estimación de la matriz de pesos  $W$  y el parámetro de localización  $\theta_{ijr}$ . Tal y como se ponía de manifiesto en la estimación de la matriz de configuración para el modelo sin diferencias individuales, resulta en ocasiones imprescindible fijar restricciones entre las coordenadas de los estímulos que eviten las indeterminaciones que inevitablemente se producen en este modelo como consecuencia de las métricas utilizadas.

Puesto que el modelo euclídeo ponderado sólo se ve afectado de una invarianza traslacional, ya que una rotación alteraría la matriz de la métrica (que suponemos diagonal), es necesario introducir restricciones sobre las coordenadas de los estímulos que fijen el origen del sistema de referencia en el centro de gravedad de la nube de puntos o centroide de la configuración. Por tanto, en el modelo que se presenta a continuación se impondrá sobre  $X$  la restricción,

$$\sum_i x_{im} = 0; \quad m = 1, \dots, K$$

Además de esa restricción sobre las coordenadas de los estímulos, en este modelo con diferencias individuales surge la necesidad de restringir también los elementos diagonales de la matriz de la métrica  $W$ . Si se observa la ecuación de la métrica,

$$d_{ijr}^* = \sqrt{\sum_m^K w_{rm} (x_{im} - x_{jm})^2}$$

resulta evidente que si se multiplican las coordenadas de los objetos,  $x_{im}$  y  $x_{jm}$  por cualquier constante arbitraria, distinta de cero, el cambio que esa multiplicación introduce en el sistema puede quedar compensado sin más que multiplicar los pesos  $w_{rm}$  por el inverso de la raíz cuadrada de dicha constante, de forma que se mantendrán invariantes las distancias.

Esta observación pone de manifiesto que la aplicación de una homotecia sobre las coordenadas de los estímulos en el espacio de configuración conjunta

puede compensarse mediante la aplicación de otra homotecia en el espacio de las ponderaciones. Esto se traduce en que existe un *problema de identificación* de los parámetros de  $X$  y de  $W$ , cuando se utiliza un modelo con diferencias individuales.

Así pues, es necesario imponer restricciones de escala que eviten la indeterminación en los modelos ponderados. El primer modelo ponderado fue el denominado INSCAL, el cual fija la escala para las coordenadas de los estímulos en el espacio de configuración conjunta, haciendo que para cada dimensión independientemente, la varianza de las coordenadas de los estímulos valga uno. Esto, debido a suponer la configuración con origen de coordenadas en el centroide, queda explícitamente indicado a través de la ecuación:

$$\sum_i x_{im}^2 = 1; \quad m = 1, \dots, K$$

Otra forma de fijar escalas consiste en imponer restricciones sobre la matriz de ponderaciones  $W$ , en lugar de imponerlas sobre la matriz de configuración (Ramsay [1982]). En este sentido y siguiendo este punto de vista, en este trabajo se impondrán las restricciones de escala dadas por las ecuaciones:

$$\sum_{r=1}^R w_{rm}^2 = R; \quad m = 1, \dots, K$$

$$\sum_{m=1}^K w_{rm}^2 = K; \quad r = 1, \dots, R$$

Estas ecuaciones representan el hecho de que si los valores diagonales de la matriz de pesos se sitúan en una matriz  $R \times K$ , tanto las medias por filas como por columnas de dicha matriz valdrán uno.

Por tanto, teniendo en cuenta las restricciones que acabamos de apuntar y que son intrínsecas a los modelos con diferencias individuales, resultan las ecuaciones de verosimilitud que veremos a continuación. Para ello, es necesario enunciar el siguiente resultado previo.

**Lema 3.3.18** *En la situación anterior, consideremos  $x_i$  y  $x_j$  dos puntos representantes de los estímulos  $i$  y  $j$  y supongamos que la configuración se encuentra determinada en dimensión  $K$ . Así pues será*

$$x_i = (x_{i1}, \dots, x_{im}, \dots, x_{iK})$$

$$x_j = (x_{j1}, \dots, x_{jm}, \dots, x_{jK})$$

Consideremos la distancia euclídea ponderada entre esos dos puntos que notamos por:

$$d_{ijr}^* = \sqrt{\sum_{m=1}^K w_{rm} (x_{im} - x_{jm})^2}$$

Entonces,

$$\frac{\partial d_{ijr}^*}{\partial x_{pq}} = 0, \quad \text{si } p \neq i, j$$

$$\text{Si } p = i, \quad \frac{\partial d_{ijr}^*}{\partial x_{pq}} = \frac{(x_{pq} - x_{jq})}{d_{ijr}^*}$$

$$\text{Si } p = j, \quad \frac{\partial d_{ijr}^*}{\partial x_{pq}} = \frac{(x_{pq} - x_{iq})}{d_{ijr}^*}$$

Además se tiene que,

$$\frac{\partial d_{ijrt}^*}{\partial w_{sq}} = 0, \quad \text{si } r \neq s$$

$$\text{Si } r = s, \quad \frac{\partial d_{ijrt}^*}{\partial w_{sq}} = \frac{(x_{iq} - x_{jq})^2}{2d_{ijrt}^*}$$

Veamos a continuación como queda en este modelo la estimación de la matriz de configuración  $X$ .

**Teorema 3.3.19** Consideremos en el modelo euclídeo ponderado, heteroscedástico, lognormal triparamétrico, con diferencias individuales, una representación en un espacio de configuración conjunta de dimensión  $K$ . Notemos por  $x_i$  y  $x_j$  a los puntos de dicho espacio, representantes de los estímulos  $i$  y  $j$  respectivamente,

$$x_i = (x_{i1}, \dots, x_{im}, \dots, x_{iK})$$

$$x_j = (x_{j1}, \dots, x_{jm}, \dots, x_{jK})$$

Entonces, el estimador por máxima verosimilitud de la matriz de configuración  $X$ , sujeto a las restricciones impuestas por la ecuación,

$$\sum_i x_{im} = 0; \quad m = 1, \dots, K$$

vendrá dado por el sistema de ecuaciones implícitas

$$x_{pq} = \frac{\sum_j t_{pjq} x_{jq} - \lambda_q}{\sum_j t_{pjq}}; \quad p = 1, \dots, n. \quad y \quad q = 1, \dots, K.$$

donde, para cada  $p, j$  y  $q$ ,

$$t_{pjq} = \frac{1}{(d_{pj}^*)^2} \sum_r \sum_t \frac{w_{rq}}{\sigma_{pjrt}^2 (d_{pjr}^*)^2} \left[ \log \left( \frac{d_{pjrt} - \theta_{pjr}}{d_{pjr}^*} \right) + \log \left( \frac{d_{jprt} - \theta_{jpr}}{d_{pjr}^*} \right) \right]$$

y donde los multiplicadores de Lagrange vienen dados a través de la expresión,

$$\lambda_q = \frac{\sum_p \sum_j u_{pq} t_{pjq} x_{jq}}{\sum_p u_{pq}}$$

$$\text{siendo, } u_{pq} = \frac{1}{\sum_j t_{pjq}}$$

DEMOSTRACIÓN.

En este caso, la función a maximizar es:

$$Q = \log L + \sum_m \lambda_m \left( \sum_i x_{im} \right)$$

Por tanto, tendremos que:

$$\frac{\partial Q}{\partial x_{pq}} = \sum_i \sum_j \sum_r \sum_t \frac{1}{\sigma_{ijrt}^2} \log \left( \frac{d_{ijrt} - \theta_{ijr}}{d_{ijr}^*} \right) \frac{1}{d_{ijr}^*} \frac{\partial}{\partial x_{pq}} d_{ijr}^* + \lambda_q = 0$$

de donde aplicando el lema anterior y reajustando índices se obtiene,

$$\sum_j \sum_r \sum_t \frac{w_{rq}}{\sigma_{pjrt}^2 (d_{pjr}^*)^2} (x_{pq} - x_{jq}) \left[ \log \left( \frac{d_{pjrt} - \theta_{pjr}}{d_{pjr}^*} \right) + \log \left( \frac{d_{jprt} - \theta_{jpr}}{d_{pjr}^*} \right) \right] + \lambda_q = 0$$

Si llamamos

$$t_{pjq} = \frac{1}{(d_{pjr}^*)^2} \sum_r \sum_t \frac{w_{rq}}{\sigma_{pjrt}^2 (d_{pjr}^*)^2} \left[ \log \left( \frac{d_{pjrt} - \theta_{pjr}}{d_{pjr}^*} \right) + \log \left( \frac{d_{jprt} - \theta_{jpr}}{d_{pjr}^*} \right) \right]$$

tendremos que

$$x_{pq} \sum_j t_{pjq} = \sum_j t_{pjq} x_{jq} - \lambda_q$$

Por otro lado tenemos que

$$\frac{\partial Q}{\partial \lambda_q} = \sum_p x_{pq} = 0$$

Al resolver esas ecuaciones de verosimilitud, despejando  $x_{pq}$  en la primera de ellas se obtiene,

$$x_{pq} = \frac{\sum_j t_{pjq} x_{jq} - \lambda_q}{\sum_j t_{pjq}}$$

Si sumamos en  $p$  en esa ecuación y considerando la definición de  $u_{pq}$  dada en el enunciado, se tendrá finalmente el valor de  $\lambda_q$  deseado.

C.Q.D.

Veamos a continuación la estimación de los elementos de la matriz de pesos  $W$  sujeta a las restricciones impuestas.

**Teorema 3.3.20** Consideremos en el modelo euclídeo ponderado, heteroscedástico, lognormal triparamétrico, con diferencias individuales, una representación en un espacio de configuración conjunta de dimensión  $K$ . Notemos por  $x_i$  y  $x_j$  a los puntos de dicho espacio, representantes de los estímulos  $i$  y  $j$  respectivamente,

$$x_i = (x_{i1}, \dots, x_{im}, \dots, x_{iK})$$

$$x_j = (x_{j1}, \dots, x_{jm}, \dots, x_{jK})$$

Entonces, el estimador por máxima verosimilitud de la matriz de pesos  $W$ , sujeto a las restricciones impuestas por la ecuación,

$$\sum_{m=1}^K w_{rm}^2 = K, \quad \forall r = 1, \dots, R$$

$$\sum_{r=1}^R w_{rm}^2 = R, \quad \forall m = 1, \dots, K$$

viene dado por el sistema de ecuaciones de verosimilitud:

$$w_{sq} = \frac{1}{\lambda_s + \mu_q} \sum_i \sum_j \sum_t \frac{1}{\sigma_{ijst}^2} \log \left( \frac{d_{ijst} - \theta_{ijs}}{d_{ijs}^*} \right) \frac{(x_{iq} - x_{jq})^2}{d_{ijs}^*}$$

donde los multiplicadores de Lagrange vienen dados por

$$\lambda_s = \frac{1}{K} \left[ \sum_i \sum_j \sum_t \frac{1}{\sigma_{ijst}^*} \log \left( \frac{d_{ijst} - \theta_{ijs}}{d_{ijs}^*} \right) - \sum_q \mu_q w_{sq}^2 \right]$$

y

$$\mu_q = \frac{1}{R} \left[ \sum_i \sum_j \sum_s \sum_t \frac{1}{\sigma_{ijst}^2} \log \left( \frac{d_{ijst} - \theta_{ijs}}{d_{ijs}^*} \right) \frac{w_{sq} (x_{iq} - x_{jq})^2}{(d_{ijs}^*)^2} - \sum_s \lambda_s w_{sq}^2 \right]$$

DEMOSTRACIÓN.

En esta ocasión, se trata de maximizar la función,

$$Q(w_{sq}) = \log L - \frac{1}{4} \sum_r \lambda_r \left( \sum_m w_{rm}^2 - K \right) - \frac{1}{4} \sum_m \mu_m \left( \sum_r w_{rm}^2 - R \right)$$

Por tanto, derivando respecto de  $w_{sq}$ , queda;

$$\frac{\partial}{\partial w_{sq}} \log L = \sum_i \sum_j \sum_r \sum_t \frac{1}{\sigma_{ijrt}^2} \log \left( \frac{d_{ijrt} - \theta_{ijr}}{d_{ijr}^*} \right) \frac{1}{d_{ijr}^*} \frac{\partial}{\partial w_{sq}} d_{ijr}^* = 0$$

En virtud del lema anterior, se tiene que

$$\frac{1}{2} \sum_i \sum_j \sum_t \frac{1}{\sigma_{ijst}^2} \log \left( \frac{d_{ijst} - \theta_{ijs}}{d_{ijs}^*} \right) \frac{(x_{iq} - x_{jq})^2}{(d_{ijs}^*)^2}$$

Además, las restantes derivadas parciales quedan:

$$\frac{\partial}{\partial w_{sq}} = \frac{1}{2} \lambda_s w_{sq}$$

$$\frac{\partial}{\partial w_{sq}} = \frac{1}{2} \mu_q w_{sq}$$

Por tanto, las ecuaciones de verosimilitud pueden ponerse como:

$$\frac{\partial Q}{\partial w_{sq}} = \sum_i \sum_j \sum_t \frac{1}{\sigma_{ijst}^2} \log \left( \frac{d_{ijst} - \theta_{ijs}}{d_{ijs}^*} \right) \frac{(x_{iq} - x_{jq})^2}{(d_{ijs}^*)^2} - \lambda_s w_{sq} - \mu_q w_{sq} = 0$$

$$\frac{\partial Q}{\partial \lambda_s} = \sum_q w_{sq}^2 = K$$

$$\frac{\partial Q}{\partial \mu_q} = \sum_s w_{sq}^2 = R$$

Multiplicando en la primera ecuación por  $w_{sq}$  queda,

$$\sum_i \sum_j \sum_t \frac{1}{\sigma_{ijst}^2} \log \left( \frac{d_{ijst} - \theta_{ijs}}{d_{ijs}^*} \right) \frac{w_{sq} (x_{iq} - x_{jq})^2}{(d_{ijs}^*)^2} = (\lambda_s + \mu_q) w_{sq}^2$$

Despejando en esa expresión el valor  $w_{sq}$  se tiene que

$$w_{sq} = \frac{1}{\lambda_s + \mu_q} \sum_i \sum_j \sum_t \frac{1}{\sigma_{ijst}^2} \log \left( \frac{d_{ijst} - \theta_{ijs}}{d_{ijs}^*} \right) \frac{(x_{iq} - x_{jq})^2}{d_{ijs}^*}$$

Sumando ahora la expresión anterior para  $q$  y aplicando la segunda ecuación de verosimilitud, se deduce que

$$\sum_i \sum_j \sum_t \frac{1}{\sigma_{ijst}^2} \log \left( \frac{d_{ijst} - \theta_{ijs}}{d_{ijs}^*} \right) = \lambda_s K + \sum_q \mu_q w_{sq}^2$$

y despejando ahí  $\lambda_s$  se tiene que

$$\lambda_s = \frac{1}{K} \left[ \sum_i \sum_j \sum_t \frac{1}{\sigma_{ijst}^2} \log \left( \frac{d_{ijst} - \theta_{ijs}}{d_{ijs}^*} \right) - \sum_q \mu_q w_{sq}^2 \right]$$

Volviendo nuevamente a la primera ecuación de verosimilitud, multiplicada por  $w_{sq}$  y sumando en dicha expresión en el índice  $s$  se consigue

$$\sum_i \sum_j \sum_s \sum_t \frac{1}{\sigma_{ijst}^2} \log \left( \frac{d_{ijst} - \theta_{ijs}}{d_{ijs}^*} \right) \frac{w_{sq}(x_{iq} - x_{jq})^2}{(d_{ijs}^*)^2} = \sum_s \lambda_s w_{sq}^2 + \mu_q R$$

y despejando en esa expresión  $\mu_q$  queda finalmente

$$\mu_q = \frac{1}{R} \left[ \sum_i \sum_j \sum_s \sum_t \frac{1}{\sigma_{ijst}^2} \log \left( \frac{d_{ijst} - \theta_{ijs}}{d_{ijs}^*} \right) \frac{w_{sq}(x_{iq} - x_{jq})^2}{(d_{ijs}^*)^2} - \sum_s \lambda_s w_{sq}^2 \right]$$

C.Q.D.

Finalmente, veamos la expresión resultante de la estimación por máxima verosimilitud del parámetro de localización  $\theta_{pqs}$ . Para la estimación de este parámetro y teniendo en cuenta los problemas que la estimación por máxima verosimilitud plantea en el caso de la distribución lognormal triparamétrica,

nos decidimos igualmente por el método de estimación por máxima verosimilitud local indicado por Cohen (Crown & Shimizu, [1988]). Así pues será considerado en esta ocasión el modelo dado por:

$$D_{ijr} \sim \Lambda(\theta_{ijr}, \log(d_{ijr}^*, \sigma_{ijr}))$$

$$\log L(\theta_{ijr}, \log(d_{ijr}^*, \sigma_{ijr})) =$$

$$-\frac{1}{2} \sum_i \sum_j \sum_r \sum_t \frac{1}{\sigma_{ijr}^2} \log^2 \left( \frac{d_{ijrt} - \theta_{ijr}}{d_{ijr}^*} \right) - \sum_i \sum_j \sum_r \sum_t \log(\sigma_{ijr})$$

$$- \sum_i \sum_j \sum_r \sum_t \log(d_{ijrt} - \theta_{ijr}) - M \log(\text{sqrt}2\pi)$$

**Teorema 3.3.21** *El estimador por máxima verosimilitud del parámetro de localización en el modelo euclídeo ponderado, lognormal triparamétrico y con diferencias individuales, viene determinado por la ecuación:*

$$\left[ \sum_t \frac{1}{d_{pqst} - \theta_{pqs}} \right] \left[ \sum_t (d_{pqst} - \theta_{pqs}) - \sum_t \log^2(d_{pqst} - \theta_{pqs}) + \right.$$

$$\left. + \frac{1}{T} \left( \sum_t \log(d_{pqst} - \theta_{pqs}) \right)^2 \right] - T \sum_t \frac{1}{d_{pqst} - \theta_{pqs}} \log(d_{pqst} - \theta_{pqs}) = 0$$

DEMOSTRACIÓN.

En esta ocasión, la log-verosimilitud asociada al modelo viene determinada por:

$$\log L(\{\theta_{ijr}\}, \{\log(d_{ijr}^*), \{\sigma_{ijr}\}) =$$

$$-\frac{1}{2} \sum_i \sum_j \sum_r \sum_t \frac{1}{\sigma_{ijr}^2} \log^2 \left( \frac{d_{ijrt} - \theta_{ijr}}{d_{ijr}^*} \right) - \sum_i \sum_j \sum_r \sum_t \log(\sigma_{ijr})$$

$$- \sum_i \sum_j \sum_r \sum_t \log(d_{ijrt} - \theta_{ijr}) - M \log(\text{sqrt}2\pi)$$

El procedimiento ideado por Cohen consiste en obtener en primer lugar las ecuaciones de verosimilitud en base a los parámetros del modelo,  $\log(d_{pqs}^*)$ ,  $\sigma_{pqs}$  y  $\theta_{pqs}$ , que serán:

$$\frac{\partial \log L}{\partial \log(d_{pqs}^*)} = \frac{1}{\sigma_{pqs}^2} \sum_t \log \left( \frac{d_{pqst} - \theta_{pqs}}{d_{pqs}^*} \right) = 0$$

$$\frac{\partial \log L}{\partial \sigma_{pqs}} = \frac{T}{\sigma_{pqs}} + \frac{1}{\sigma_{pqs}^3} \sum_t \log^2 \left( \frac{d_{pqst} - \theta_{pqs}}{d_{pqs}^*} \right) = 0$$

$$\frac{\partial \log L}{\partial \theta_{pqs}} = \frac{1}{\sigma_{pqs}^2} \sum_t \frac{1}{d_{pqst} - \theta_{pqs}} \log^2 \left( \frac{d_{pqst} - \theta_{pqs}}{d_{pqs}^*} \right) + \sum_t \frac{1}{d_{pqst} - \theta_{pqs}} = 0$$

De modo análogo a Cohen, (Crown & Shimizu, [1988]), eliminamos en las ecuaciones anteriores  $\log(d_{pqs}^*)$  y  $\sigma_{pqs}^2$ , de la forma:

$$\log(d_{pqs}^*) = \frac{1}{T} \sum_t \log(d_{pqst} - \theta_{pqs})$$

$$\sigma_{pqs}^2 = \frac{1}{T} \sum_t \log^2(d_{pqst} - \theta_{pqs}) - \left[ \frac{1}{T} \sum_t \log(d_{pqst} - \theta_{pqs}) \right]^2$$

Se obtiene finalmente la ecuación,

$$\left[ \sum_t \frac{1}{d_{pqst} - \theta_{pqs}} \right] \left[ \sum_t (d_{pqst} - \theta_{pqs}) - \sum_t \log^2(d_{pqst} - \theta_{pqs}) + \right.$$

$$\left. + \frac{1}{T} \left( \sum_t \log(d_{pqst} - \theta_{pqs}) \right)^2 \right] - T \sum_t \frac{1}{d_{pqst} - \theta_{pqs}} \log(d_{pqst} - \theta_{pqs}) = 0$$

C.Q.D.

Calculando el estimador por máxima verosimilitud local de  $\theta_{pqs}$  en la ecuación anterior, mediante procedimientos numéricos como los que a continuación describiremos y sustituyendo dicho estimador en las restantes ecuaciones de verosimilitud. asociadas a los parámetros, que la estimación en el modelo log-normal triparamétrico se reduce a la del modelo biparamétrico al considerar fijado el estimador del parámetro de localización y por tanto, estimar los restantes parámetros como si  $\theta_{pqs}$  fuese conocido.

### 3.3.5 Estudio de la resolución numérica de las ecuaciones de verosimilitud.

Resulta bastante frecuente que sea difícil resolver explícitamente de las ecuaciones de verosimilitud, incluso en aquellos casos en los que se cumplen todas las condiciones de regularidad clásicas y existe una única solución, como ocurre por ejemplo con la familia exponencial. Además, si la familia de distribuciones que se considera no es del tipo exponencial y cabe la posibilidad de que existan varias raíces de las ecuaciones de verosimilitud, resulta bastante difícil localizar un máximo absoluto. Fisher [1929] fue el primero que apuntó el uso de iteraciones sucesivas para la resolución de la ecuación de verosimilitud. Argumentó que, en algunos casos, basta con ejecutar un solo ciclo de iteraciones para la obtención de una buena aproximación. Posteriormente otros autores han descrito que son necesarios varios ciclos para obtener una convergencia razonable en la mayoría de los casos. Veamos algunas notas sobre aproximaciones a la solución del problema a través de la utilización del método de Newton-Raphson y de algunas variantes del mismo.

#### 1. El método de Newton-Raphson.

Este método está basado en el desarrollo en serie de la ecuación de verosimilitud en torno a la raíz de la ecuación que notaremos por  $\hat{\Theta}$ . Supongamos una muestra de tamaño  $n$  de una v.a.  $X$ . Así pues,

$$0 = \frac{\partial}{\partial \Theta} L(\hat{\Theta}; X) = \frac{\partial}{\partial \Theta} L(\Theta_1; X) + (\hat{\Theta} - \Theta_1) \frac{\partial^2}{\partial \Theta^2} L(\Theta_1 + \nu(\hat{\Theta} - \Theta_1); X)$$

para algún valor  $0 \leq \nu \leq 1$ , donde  $\hat{\Theta}$  es una raíz de la ecuación de verosimilitud y  $\Theta_1$  es una solución inicial. Si tomamos  $\nu = 0$  se obtiene una aproximación para  $\hat{\Theta}$ , llamada

$$\Theta_2 = \Theta_1 - \frac{\frac{\partial}{\partial \Theta} L(\Theta_1; X)}{\frac{\partial^2}{\partial \Theta^2} L(\Theta_1; X)}$$

El valor de  $\Theta_2$  puede sustituirse en la expresión anterior por  $\Theta_1$  para obtener un nuevo valor  $\Theta_3$  y así sucesivamente. Generalmente, comenzando con una solución inicial  $\Theta_1$  se genera una sucesión de valores  $\{\Theta_t; t \geq 1\}$ , cuyo término general puede escribirse

$$\Theta_{t+1} = \Theta_t - \frac{\frac{\partial}{\partial \Theta} L(\Theta_t; X)}{\frac{\partial^2}{\partial \Theta^2} L(\Theta_t; X)}; \quad t = 1, 2, \dots$$

Si la solución inicial  $\Theta_1$  se elige suficientemente próxima a la verdadera solución o raíz de la ecuación,  $\hat{\Theta}$  y si  $\frac{\partial^2}{\partial \Theta^2} L(\Theta_t; X)$ ,  $t = 1, 2, \dots$ , está acotada lejos del cero, puede afirmarse que la sucesión de valores aproximados generada por la expresión anterior, converge a la raíz deseada  $\hat{\Theta}$ . La sucesión anterior depende de los valores muestrales  $X_1, \dots, X_n$ . Es conocido que si se elige un valor inicial  $\Theta_1$  que sea un estimador consistente de  $\Theta$ , bastará con una iteración para obtener el mejor estimador asintóticamente normal de  $\Theta$ . No obstante, en situaciones muestrales pequeñas, la sucesión  $\{\Theta_t; t \geq 1\}$  puede mostrar irregularidades, debido a los valores muestrales particulares obtenidos en el experimento. En relación con la eliminación de irregularidades en la sucesión de aproximación, se han propuesto clásicamente dos variantes del método de Newton-Raphson: el método de la aproximación de Newton Raphson de derivadas fijas y el método de las puntuaciones o de tanteo de Fisher.

## 2. El método de la aproximación a Newton con Derivadas fijas.

En esta variante del método de Newton-Raphson mediante el empleo de derivadas fijas, cada término del tipo,

$$\frac{\partial^2}{\partial \Theta^2} L(\Theta_t; X)$$

es reemplazado por el término,

$$-\frac{n}{a_t}$$

donde la sucesión  $\{a_t : t \geq 1\}$  es una sucesión de constantes y  $n$  es el tamaño muestral. Así pues, se obtiene la sucesión:

$$\Theta_{t+1} = \Theta_t + \frac{a_t}{n} \frac{\partial}{\partial \Theta} L(\Theta_t; X); \quad t = 1, 2, \dots$$

Por tanto, una secuencia de aproximación, para una sucesión de valores bien elegidos  $\{a_t\}_t$ , debe ser más estable que la sucesión que se proponía en el método de Newton-Raphson y converge hacia el valor  $\hat{\Theta}$  de forma más regular. No obstante, en muchos casos se ha observado que si la función de log-verosimilitud entra en el entorno de un máximo local, esta última sucesión converge normalmente a dicho máximo local. En el caso especial de tomar  $a_t = n$ ,  $\forall t$ , se obtiene la sucesión

$$\Theta_{t+1} = \Theta_t + \frac{\partial}{\partial \Theta} L(\Theta_t; X); \quad t = 1, 2, \dots$$

la cual, ha sido utilizada por muchos investigadores.

### 3. El método de las puntuaciones o del tanteo (Scoring) de Fisher.

El método de las puntuaciones es un caso particular de la aproximación de Newton-Raphson por derivadas fijas introducido por Fisher [1925]. La sucesión especial de valores  $\{a_t\}_t$  sugerida por Fisher fue,

$$\frac{1}{I(\Theta_t)}$$

donde  $I(\Theta)$  es la función de información de Fisher (la cual se supone positiva para todo valor de  $\Theta$ ) y donde  $\Theta_t$  es el valor de la aproximación después de la  $k$ -ésima iteración. De esa forma, el método de las puntuaciones de Fisher genera la sucesión:

$$\Theta_{t+1} = \Theta_t + \frac{1}{nI(\Theta_t)} \frac{\partial}{\partial \Theta} L(\Theta_t; X)$$

Para la  $k$ -ésima iteración se define,

$$\zeta_t = \frac{\Theta_{t-1}^{(1)} \frac{\partial}{\partial \Theta} L(\Theta_{t-1}^{(2)}; X) - \Theta_{t-1}^{(2)} \frac{\partial}{\partial \Theta} L(\Theta_{t-1}^{(1)}; X)}{L(\Theta_{t-1}^{(2)}; X) - L(\Theta_{t-1}^{(1)}; X)}$$

donde

$$\Theta_t^{(1)} = \zeta_t, \quad \Theta_t^{(2)} = \Theta_{t-1}^{(2)} \quad \text{si} \quad \frac{\partial}{\partial \Theta} L(\zeta_t; X) > 0$$

$$\Theta_t^{(1)} = \Theta_{t-1}^{(1)}, \quad \Theta_t^{(2)} = \zeta_t \quad \text{si} \quad \frac{\partial}{\partial \Theta} L(\zeta_t; X) < 0$$

Mediante este método se consigue la convergencia a un mínimo local, aunque hay que suponer que  $\frac{\partial}{\partial \Theta} L(\Theta; X)$  es continua en  $\Theta$ , para cada,  $X$ . El inconveniente primordial de este método es que no resulta obvio como elegir los dos valores iniciales  $\Theta_1^{(1)}, \Theta_2^{(2)}$ .

### Un método iterativo para el modelo euclídeo homocedástico.

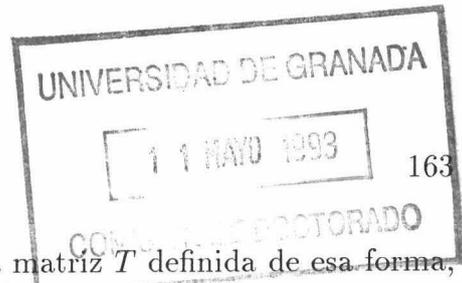
Una vez vistos los métodos generales por los que se aborda comúnmente el problema de la resolución de la ecuación de verosimilitud, veamos la aplicación que en nuestro caso puede elegirse para el modelo euclídeo, homocedástico, lognormal biparamétrico. El método que vamos a exponer puede verse en su forma general en Ramsay [1975].

Se trata de resolver numéricamente el sistema de ecuaciones

$$x_{pq} \sum_j t_{pj} = \sum_j x_{jq} t_{pj}; \quad p = 1, \dots, n. \quad \text{y} \quad q = 1, \dots, K.$$

donde, para cada  $p$  y cada  $j$ ,

$$t_{pj} = \frac{1}{(d_{pj}^*)^2} \sum_r \left[ \log \left( \frac{d_{pjr}}{d_{pj}^*} \right) + \log \left( \frac{d_{jpr}}{d_{pj}^*} \right) \right]$$



José Fernando Vera Vera.

Puede observarse en primer lugar, que la matriz  $T$  definida de esa forma, es simétrica e independiente de  $\sigma^2$ . Esa definición de los elementos  $t_{pj}$ , no especifica los elementos diagonales de  $T$  y aunque la definición es correcta cualesquiera que sean los elementos diagonales, resulta imprescindible para la convergencia del método iterativo propuesto por Ramsay, que los elementos de la diagonal de  $T$ , sean mayores que los elementos no diagonales. Consideremos ahora una nueva matriz  $C$  que se define de la forma

$$c_{pj} = \frac{t_{pj}}{\sum_i t_{pi}}$$

Operando en el sistema de ecuaciones de verosimilitud, queda

$$x_{pq} \sum_i t_{pi} = \sum_j x_{jq} t_{pj}$$

luego

$$x_{pq} = \sum_j x_{jq} \frac{t_{pj}}{\sum_i t_{pi}} = \sum_j x_{jq} c_{pj}$$

Así pues, se tiene que  $x_{pq} = \sum_j c_{pj} x_{jq}$  o matricialmente

$$X_{n \times K} = C_{n \times n} X_{n \times K}$$

De ahí se deduce que:

- Cada coordenada estimada  $x_{pq}$  es un promedio ponderado de todas las coordenadas de  $X$  en la  $q$ -ésima dimensión ( $q$ -ésima columna de  $X$ ).
- Los elementos de la diagonal de  $T$  ( $t_{pp}$  a través de los  $c_{pp}$ ), asignan pesos a cada coordenada estimada en dicho promedio ( $x_{pq}$  en  $\sum_j c_{pj} x_{jq}$ ).
- El hecho de que la diagonal de  $T$  sea dominante sobre los restantes elementos de  $T$  (equivalentemente para  $C$ ), indicará que la coordenada asociada a ese peso dominará sobre las restantes coordenadas de ese promedio.

Por tanto, se plantea la resolución de la ecuación implícita  $X = CX$ . Para ello, siguiendo el método de Ramsay, se considera la ecuación implícita general dada por

$$X^{(\nu)} = \Phi X^{(\nu-1)} + (1 - \Phi)C^{(\nu-1)}X^{(\nu-1)}$$

en donde ahora,  $\Phi$  es el parámetro de aceleración que se elige para determinar la velocidad de convergencia del algoritmo. La descripción detallada de la estimación  $\Phi$  y de su actualización iterativa, puede verse en el trabajo de Ramsay[1975] y comentado para este caso en Ramsay[1977]. En líneas generales puede decirse que existen dos cantidades sin determinar, que son  $\Phi$  y las  $t_{pp}$ . Para ver el desarrollo del proceso iterativo, vamos a reordenar la ecuación en términos de un algoritmo ascendente.

Tenemos

$$\begin{aligned} x_{pq}^{(\nu)} &= \Phi x_{pq}^{(\nu-1)} + (1 - \Phi) \sum_j x_{jq}^{(\nu-1)} c_{pj}^{(\nu-1)} = \\ &= x_{pq}^{(\nu-1)} + \Phi x_{pq}^{(\nu-1)} - x_{pq}^{(\nu-1)} + (1 - \Phi) \sum_j x_{jq}^{(\nu-1)} c_{pj}^{(\nu-1)} = \\ &= x_{pq}^{(\nu-1)} + (\Phi - 1)x_{pq}^{(\nu-1)} - (\Phi - 1) \sum_j x_{jq}^{(\nu-1)} c_{pj}^{(\nu-1)} = \\ &= x_{pq}^{(\nu-1)} + (\Phi - 1) \left[ x_{pq}^{(\nu-1)} - \sum_j x_{jq}^{(\nu-1)} \frac{t_{pj}^{(\nu-1)}}{\sum_i t_{pj}^{(\nu-1)}} \right] = \\ &= x_{pq}^{(\nu-1)} + (\Phi - 1) \left[ \frac{1}{\sum_j t_{pj}^{(\nu-1)}} \sum_j x_{pq}^{(\nu-1)} t_{pj}^{(\nu-1)} - \sum_j x_{jq}^{(\nu-1)} t_{pj}^{(\nu-1)} \frac{1}{\sum_j t_{pj}^{(\nu-1)}} \right] = \\ &= x_{pq}^{(\nu-1)} + \frac{\Phi - 1}{\sum_j t_{pj}^{(\nu-1)}} \left[ \sum_j (x_{pq}^{(\nu-1)} - x_{jq}^{(\nu-1)}) t_{pj}^{(\nu-1)} \right] = \\ &= x_{pq}^{(\nu-1)} + \frac{\sigma^2(\Phi - 1)}{t_{pp}^{(\nu-1)} \sum_{j \neq p} t_{pj}^{(\nu-1)}} \left[ \frac{1}{\sigma^2} \sum_j (x_{pq}^{(\nu-1)} - x_{jq}^{(\nu-1)}) t_{pj}^{(\nu-1)} \right] = \end{aligned}$$

$$= x_{pq}^{(\nu-1)} + \frac{\sigma^2(\Phi - 1)}{t_{pp}^{(\nu-1)} + \sum_{j \neq p} t_{pj}^{(\nu-1)}} \frac{\partial \log L}{\partial x_{pq}^{(\nu-1)}}$$

De ahí se deduce que si el parámetro de aceleración cumple la restricción  $\Phi > 1$  entonces los elementos de la diagonal de  $T$  deben verificar que  $t_{pp} > -\sum_{j \neq p} t_{pj}$  para que haya convergencia.

Hay que añadir a este respecto en primer lugar, que el procedimiento puede describirse igualmente para el caso triparamétrico, aunque no se ha considerado en este trabajo ya que como se pone de manifiesto en Ramsay [1982], el método en Ramsay [1975], no puede utilizarse cuando se introducen nuevos parámetros en el modelo, ya que la convergencia en esos casos se vuelve tediosa y en algunas ocasiones, difícil de obtener.

### 3.3.6 Desarrollo computacional del modelo.

La resolución de las ecuaciones implícitas que determinan la obtención del estimador que maximiza la verosimilitud, viene dada a través de un procedimiento numérico, por ser de todo punto imposible la obtención de los estimadores de forma explícita. Los problemas computacionales de MDS provienen, entre otros, de los siguientes aspectos: Por un lado la imposibilidad de obtener una solución explícita de las coordenadas, que conlleve las distancias óptimas y por otro, la gran cantidad de parámetros que deben ser estimados.

Para resolver esos problemas, hemos utilizado un procedimiento similar al descrito por Takane et al. [1977], tal y como propone Ramsay [1982] para el cálculo de su modelo, que consiste en un procedimiento de actualización compuesto, formado por unas iteraciones principales, dentro de cada una de las cuales se actualizan una serie de parámetros a través de una serie de iteraciones secundarias independientes. La resolución de este tipo de procesos iterativos se realiza en base a técnicas derivadas del procedimiento de Newton-Raphson para la resolución de sistemas de ecuaciones no lineales. Concretamente en nuestro modelo se utiliza el método Scoring de Fisher, el cual ofrece la ventaja de que cuando no existe convergencia cuadrática en el método clásico de Newton-Raphson (lo cual no puede comprobarse en nuestro modelo), ofrece garantías de una convergencia rápida bajo una buena aproximación inicial. En primer lugar se puede elegir estimar  $\theta_{pqs}$  a través de un método similar al de Cohen o bien actualizar el valor del parámetro dentro del proceso de

iteración. El método actúa de forma que en cada iteración principal aparecen varias etapas denominadas secundarias y para la estimación en cada una de esas etapas, se utiliza de nuevo Newton-Raphson, y manteniendo fijos los demás parámetros, actualiza el parámetro en cuestión.

A pesar de que el método Scoring es de los más adecuados para la resolución de sistemas de ecuaciones de verosimilitud tal y como se pone de manifiesto en Zacks [1971], en muchas ocasiones representa un problema calcular el Jacobiano de la función *Score* o Hessiano de la función de log-verosimilitud. Dicha matriz juega un papel fundamental en el método y en algunas ocasiones su obtención presenta graves inconvenientes. Por eso hemos utilizado también como procedimiento de actualización de parámetros, el *método del gradiente*. Así, el método Scoring requiere gran cantidad de memoria y de tiempo para el cálculo de la matriz de información de Fisher, por lo que en algunas ocasiones resulta necesario utilizar el método del gradiente el cual, en general, requiere menos iteraciones y menor costo por iteración además de menos memoria.

El método del paso descendente o del gradiente es un procedimiento de minimización de funciones (maximizar  $\log(L)$  equivale a minimizar  $-\log(L)$ ) que actualiza sucesivamente los parámetros mediante el desplazamiento de éstos en la dirección del vector gradiente negativo de la función en dichos parámetros. Para controlar la razón de cambio de la función a minimizar, se utiliza la norma del vector gradiente. Así pues, se consideran a la hora de controlar la razón del cambio los siguientes valores:

1. **Longitud del gradiente.** Si el gradiente viene dado por los elementos  $g_1, \dots, g_N$ , con  $N$  el número de parámetros a actualizar, la longitud del gradiente no es más que la norma del vector gradiente.

$$LG = \sqrt{\sum_h g_h^2}$$

2. **Longitud relativa del gradiente.** Si los parámetros son  $\Theta_1, \dots, \Theta_N$ , la longitud relativa del gradiente viene dada por

$$LRG = \sqrt{\frac{\left(\sum_h g_h^2\right) \left(\sum_h \Theta_h^2\right)}{N}}$$

Esta medida tiene la ventaja práctica de no depender de la escala de medida ni del número de parámetros como la longitud simple del gradiente. Por tanto cuando nos refiramos a longitud del gradiente lo haremos respecto a éste valor.

Para aplicar el procedimiento iterativo se han agrupado los parámetros en bloques de forma que su actualización sigue un proceso iterativo similar al descrito por Takane et al [1977] en ALSCAL y que ha sido utilizado también por Ramsay. Este procedimiento, utiliza iteraciones principales y secundarias, que actualizan los parámetros, previamente clasificados en bloques.

La *iteración principal* es un ciclo en el que se actualizan los parámetros de cada bloque. Dentro de cada uno de esos ciclos, para actualizar cada bloque de parámetros se utiliza uno de los procedimientos iterativos antes descritos de modo que asociado a cada uno de estos parámetros existirá una serie de iteraciones que constituyen la *iteración secundaria*. Veamos detalladamente el proceso describiendo las etapas dentro de cada iteración principal.

1. **Actualización de la matriz de configuración  $X$**  (y el parámetro de localización  $\theta_{ijr}$ , si no se ha hecho en un principio mediante el procedimiento similar al descrito por Cohen).

La matriz de configuración inicial,  $X^{(0)}$  vendrá dada por el método que se quiere contrastar, si éste fuese el caso, o por el procedimiento métrico clásico de Torgerson. En la  $h$ -ésima iteración principal intervienen:

$$W_s^{(h-1)}, \sigma_s^{(h-1)}, \gamma_{pq}^{(h-1)} \text{ y } \tau_u^{(h-1)},$$

donde cada subíndice varía en su conjunto correspondiente.

2. **Actualización de las matrices de ponderaciones  $W_s$ .**

En la iteración principal inicial se toma  $W_s^{(0)}$  igual a la matriz identidad, para todo individuo  $s$ . En la  $h$ -ésima iteración principal intervienen:

$$X^{(h)}, W_s^{(h-1)}, \sigma_s^{(h-1)}, \gamma_{pq}^{(h-1)} \text{ y } \tau_u^{(h-1)}.$$

3. **Actualización de las componentes de variabilidad entre pares y de los errores típicos de los estímulos:**  $\gamma_{pq}$  y  $\alpha_p$ , (si no son calculados a posteriori).

En la iteración principal inicial,  $\gamma_{pq}^{(0)} = 1$  y  $\alpha_p^{(0)} = 1$ . En la  $h$ -ésima iteración principal intervienen:

$$X^{(h)}, W_s^{(h)}, \sigma_s^{(h-1)}, \gamma_{pq}^{(h-1)}, \alpha_p^{(h-1)} \text{ y } \tau_u^{(h-1)}.$$

4. **Actualización de la variabilidad asociada a cada réplica  $\tau_u$ .**

En la iteración principal inicial,  $\tau_u^{(0)} = 1$ . En la  $h$ -ésima iteración principal intervienen:

$$X^{(h)}, W_s^{(h)}, \sigma_s^{(h-1)}, \gamma_{pq}^{(h)} \text{ y } \tau_u^{(h-1)}.$$

5. **Actualización de la variabilidad individual  $\sigma_s$ .**

En la  $h$ -ésima iteración se toman los valores:

$$X^{(h)}, W_s^{(h)}, \gamma_{pq}^{(h)} \text{ y } \tau_u^{(h)}.$$

Un aspecto fundamental de todo procedimiento iterativo es determinar cuando debe finalizar. Este hecho introduce el problema de definir el *criterio de convergencia* del método. La convergencia será alcanzada cuando en cualquier proceso iterativo, principal o secundario, se satisfaga una condición numérica. Algunas de las más importantes consisten en terminar el proceso iterativo cuando los parámetros que se están actualizando, no varíen más de una cierta cantidad, entre una iteración y la siguiente. Alternativamente otro criterio consiste en examinar la función de log-verosimilitud de forma que se termina el proceso si dicha función no aumenta más de una cierta cantidad entre dos iteraciones consecutivas.

En nuestro modelo el criterio de convergencia está basado en el desarrollo de la función de log-verosimilitud en lugar de los parámetros. Veamos los criterios usados en cada proceso iterativo.

- **Convergencia para las iteraciones principales.**

Se controla la variación de la log-verosimilitud de una iteración a otra, de forma que mediante un contraste de  $\chi^2$ , puede determinarse la iteración final. En este caso se calcula la diferencia entre dos valores de la log-verosimilitud y se compara el doble de esa diferencia con valores de la variable  $\chi^2$ .

No es necesaria una gran exactitud para aplicar este contraste y como regla general es suficiente conocer una variable  $\chi^2$  con algo más de cinco grados de libertad (casi siempre es el caso en MDS). Así pues, el proceso finaliza cuando la log-verosimilitud no cambia más de 0.05, de una iteración a la siguiente.

Por razones del análisis, puede requerirse finalizar para valores menores que rechaza este criterio, por lo que también se tiene en cuenta la longitud del gradiente, de modo que se finalizará el proceso cuando dicha longitud tienda a cero.

- **Convergencia para las iteraciones de la matriz de configuración.**

Puesto que la log-verosimilitud depende de las coordenadas  $x_{im}$  a través de la suma de los cuadrados de los errores, bastará con minimizar dicha suma para maximizar la log-verosimilitud. Así si llamamos

$$SEC = \sum_r \sum_t \sum_i \sum_j [\log(d_{ijrt}) - \log(d_{ijr}^*)]^2$$

fácilmente puede comprobarse que el criterio para  $SEC$  es equivalente al criterio del cambio en 0.05 en la log-verosimilitud. Por tanto, se usa en este caso el criterio de que el valor

$$\frac{SEC^{(h-1)} - SEC^{(h)}}{SEC^{h-1}}$$

sea menor que una cantidad suficientemente pequeña, por ejemplo 0.001.

Para el parámetro de localización el criterio que se sigue es análogo al de la iteración principal.

- **Convergencia para las iteraciones de las matrices de la métrica.**

La log-verosimilitud depende de las matrices  $W_r$ , para el sujeto  $r$ , a través de la suma de los errores al cuadrado para ese sujeto. Se utiliza por tanto el criterio de un cambio análogo al de la log-verosimilitud en las iteraciones principales, no mayor a 0.05 en la suma de los cuadrados de los errores.

- **Convergencia para las componentes de variabilidad entre pares de estímulos, errores típicos de los estímulos y componentes de réplicas.**

El proceso es análogo a los anteriores.

- **Convergencia para las componentes de variabilidad asociadas a los sujetos.**

Su resolución es exacta y no necesita de un proceso iterativo.

Finalmente hay que añadir que para la resolución del proceso, se ha construido un programa en lenguaje PDS Basic cuyo listado así como ejemplos de comparación y aplicación se incluyen en el anexo a esta memoria.

### 3.3.7 Comparación de modelos y contraste de hipótesis.

Es suficientemente conocido que los estimadores por máxima verosimilitud poseen propiedades muy deseables cuando los tamaños muestrales son grandes y las hipótesis estadísticas están justificadas. Bajo esas circunstancias, la log-verosimilitud puede utilizarse para contrastar hipótesis referentes a qué modelo, entre dos dados, proporciona mejor ajuste a los datos.

Esto puede realizarse cuando uno de los modelos es un caso especial del otro en el sentido de que el más simple se deriva del otro mediante alguna restricción de los parámetros del modelo más general. Un caso muy importante al respecto es la comparación entre un modelo de distancia en  $M - 1$  y en  $M$  dimensiones. Este último se reduce al anterior sin más que considerar cero las  $M$ -ésimas coordenadas. De igual forma, otro caso puede aplicarse respecto a la matriz identidad  $I$  que es un caso especial del modelo con diferencias individuales y métrica diagonal, en el que puede considerarse que todos los elementos de cada matriz  $W_r$  son iguales a uno.

Notemos la verosimilitud para ese modelo más general por  $L_\Omega$  y para el modelo más simple o caso especial por  $L_\omega$ . Entonces, la cantidad

$$\chi^2 = 2(L_\Omega - L_\omega)$$

suministra un criterio de comparación suficiente, en virtud del contraste asintótico de la razón de verosimilitud. Así pues, cuando el tamaño muestral es grande y bajo la hipótesis nula de que el modelo simple describe la población, este criterio se distribuye según una  $\chi^2$ . Los grados de libertad de esa distribución son simplemente la diferencia entre el número de parámetros para el modelo general y el número de parámetros del modelo más simple. Consultando el valor de la  $\chi^2$  asociado se determina si el valor calculado es significativo o no al nivel deseado. Si se considera significativo, indicará que el modelo más general añade información sustancial en el ajuste de los datos.

De entre los diferentes contrastes uno a destacar es el contraste de dimensionalidad. Se trata de decidir qué dimensión es la más adecuada para la representación de los datos. En base al criterio de la  $\chi^2$ , se evalúa la log-verosimilitud resultante en dimensión  $K - 1$  y en dimensión  $K$ . Para describir este contraste vamos a centrar nuestra atención en el modelo sin diferencias individuales homocedástico. Para dicho modelo, supuestos  $n$  estímulos y sin réplicas, en dimensión  $K$  hay  $n \times K$  parámetros referentes a la matriz de configuración  $X$  y un parámetro más relativo a la componente de la varianza  $\sigma$ . Puesto que existe para este modelo una invarianza tanto rotacional como traslacional, surge la necesidad de imponer restricciones que disminuyan los grados de libertad. En ese sentido, las restricciones impuestas son,

$$\frac{K(K-1)}{2} \text{ restricciones rotacionales } \sum_i^n x_{in}x_{im} = 0; \quad n \neq m = 1, \dots, K.$$

$$K \text{ restricciones traslacionales } \sum_i^n x_{im} = 0; \quad m = 1, \dots, K.$$

Por tanto, el número de parámetros independientes será el número total de parámetros del modelo en dimensión  $K$  menos el número de restricciones,

$$\left( nK - \left[ \frac{K(K-1)}{2} + K \right] \right) + 1 = \left( nK - \frac{K(K-1)}{2} \right) + 1$$

Así pues, si se quiere contrastar la hipótesis nula de que la dimensión es  $K - 1$ , frente a que es  $K$ , el número de grados de libertad de la distribución  $\chi^2$  asociada viene dado por

$$\left( nK - \frac{K(K-1)}{2} \right) + 1 - \left( n(K-1) - \frac{(K-1)K}{2} \right) - 1 = n - K$$

Luego si se decide un nivel de significación para el contraste de 0.05, se concluye que la dimensión del modelo es  $K$  si se tiene que

$$2(\log L_K - \log L_{K-1}) > \chi_{0.05}^2(n - K)$$

De la misma forma se procede cuando se desea determinar en base al valor de la log-verosimilitud, si se termina el proceso en la iteración  $h - 1$  o en la iteración  $h$ .

Por último, para la comparación entre dos modelos en la situación de uno simple y otro más general, hay que hacer consideraciones adicionales. La log-verosimilitud no puede interpretarse por sí misma. Ésta puede ser negativa o positiva y su tamaño depende del número de observaciones y del número de parámetros del modelo. No obstante, bajo ciertas circunstancias, dos log-verosimilitudes pueden ser comparadas, como es el caso que nos ocupa.

Como se acaba de comentar, el ajuste de criterios como el de la log-verosimilitud y el de mínimos cuadrados adolece del problema de que sus valores dependen del número de parámetros estimados, de forma que cuantos más parámetros se introducen en el modelo, mayor será el valor de la log-verosimilitud asociado. Por tanto, dos log-verosimilitudes pueden ser comparadas sólo cuando el modelo que produce la primera es un caso particular o un subconjunto del otro modelo. Además, la log-verosimilitud debe ser corregida en relación al número de parámetros del modelo, de forma que dos modelos puedan compararse a nivel descriptivo, incluso cuando posean un número diferente de parámetros.

El estadístico AIC es una de esas modificaciones, descrita por Akaike [1974]. Si  $L$  es la log-verosimilitud, entonces el estadístico AIC se define

$$AIC = -2L + 2NP,$$

donde  $NP$  es el número de parámetros independientes que son estimados.

Otra modificación de la log-verosimilitud que persigue el mismo fin es la propuesta por Schwarz [1978] mediante el estadístico BIC que se define

$$BIC = -2L + NP \log NTO$$

donde  $NTO$  es el número total de observaciones.

Cuanto menores sean estos estadísticos mejor será el modelo de ajuste, ya que ambos están relacionados con  $-\log L$ . El modelo que tenga asociado el menor valor de cualquiera de estos estadísticos será preferido al otro. El estadístico BIC es generalmente más conservativo en el sentido de que es mejor para aquellos modelos que poseen un pequeño número de parámetros, aunque ambos llegan a las mismas conclusiones. No obstante, el uso de ambos estadísticos es exclusivamente descriptivo y no en el sentido de un contraste de hipótesis.

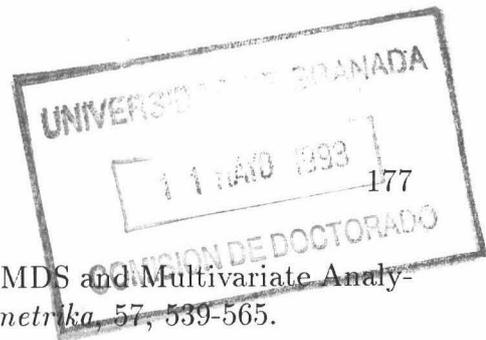
## Bibliografía

- [1] Abrahamowicz, M. & Ramsay, J. O. (1992).- Multicategorical Spline Model for Item Response Theory. *Psychometrika*, 57, 5-27.
- [2] Akaike, H. (1987).- Factor Analysis and AIC. *Psychometrika*, 52, 317-332.
- [3] Arabie, P., Carroll, J. D. & DeSarbo, W. S. (1987).- Three-Way Scaling And Clustering. *Sage Publications, Inc.*
- [4] Arabie, P. (1991).- Was Euclid an Unnecessarily Sophisticated Psychologist?. *Psychometrika*, 56, 567-587.
- [5] Besse, Ph. & Ramsay, J. O. (1986).- Principal Components Analysis of Sampled Functions. *Psychometrika*, 51, 285-311.
- [6] Bijleveld, C. C. J. H. & de Leeuw, J. (1991).- Fitting Longitudinal Reduced-Rank Regression Models by Alternating Least Squares. *Psychometrika*, 56, 433-447.
- [7] Bissett, R. & Schneider, B. (1991).- Spatial and Conjoint Models Based on Pairwise Comparisons of Dissimilarities and Combined Effects: Complete and Incomplete Designs. *Psychometrika*, 56, 685-698.
- [8] Bloxom, B. (1978).- Constrained MDS in N Spaces. *Psychometrika*, 43, 397-408.
- [9] Bozdogan, H. (1987).- Model Selection and Akaike's Information Criterion (AIC): The General Theory and its Analytical Extensions. *Psychometrika*, 52, 345-370.

- [10] Böckenholt, U. & Böckenholt, I. (1991).- Constrained Latnet Class Analysis: Simultaneous Classification and Scaling of Discrete Choice Data. *Psychometrika*, 56, 699-716.
- [11] Brady, H. E. (1985).- Statistical Consistency and Hypothesis Testing for Nonmetric MDS. *Psychometrika*, 50, 509-537.
- [12] Carroll, J. D., & Chang, J. J. (1970).- Analysis of individual differences in multidimensional scaling via  $N$ -way generalization of Eckart-Young decomposition. *Psychometrika*, 35, 283-320.
- [13] Carroll, J. D., Pruzansky, S. & Kruskal, J. B. (1980).- Candelic: A general Approach to MDS Analysis of Many-Way Arrays with Linear Constraints on Parameters. *Psychometrika*, 45, 3-24.
- [14] Chan, S. (1991).- On the Calculations of the Maximum Likelihood Estimates for the Polynomial Spline Regression Models with Unknown Knots and AR(1) Errors. *Commun. Statist.- Theory Meth.*, 20, 4, 1199-1209.
- [15] Corter, J. E. & Tversky, A. (1986).- Extended Similarity Trees. *Psychometrika*, 51, 429-451.
- [16] Crown, E. L. & Shimizu, K. (1988).- Lognormal Distribution Theory and Applications. *Marcel Dekker Ed.*
- [17] De Leeuw, J. & Heiser, W. (1980).- MDS with Restrictions on the Configuration. *Multivariate Analysis, Vol. 5* (P. R. Krishnaiah, Ed.). Amsterdam, North-Holland.
- [18] De Leeuw, J. (1984).- Differentiability of Kruskal's Stress at a Local Minimum. *Psychometrika*, 49, 111-113.
- [19] De Leeuw, J. & Bettonvil, B. (1986).- An Upper bound for Stress. *Psychometrika*, 51, 149-153.
- [20] DeSarbo, W. S., Howard, D. J. & Jedidi, K. (1991).- Multiclus: A New Method for Simultaneously Performing MDS and Cluster Analysis. *Psychometrika*, 56, 121-136.

- [21] DeSarbo, W. S., Jedidi, K. & Steckel, J. H. (1991).- A Stochastic MDS Procedure for the Empirical Determination of Convex Indifference Curves for Preference/Choice Analysis. *Psychometrika*, 56, 279-301.
- [22] DeSarbo, W. S., Johnson, M. D., Manrai, L. A. & Edwards, E. A. (1992).- Tscale: A New MDS Procedure Based on Tversky's Contrast Model. *Psychometrika*, 57, 43-69.
- [23] Folks, J. L. & Chhikara, R. S. (1978).- The Inverse Gaussian Distribution and its Statistical Application-A Review (with Discussion), *J. R. Statist. Soc. B*, 40, 263-289.
- [24] Green, R. S. & Bentler, P. M. (1979).- Improving the Efficiency and Effectiveness of Interactively Selected MDS Data Designs. *Psychometrika*, 44, 115-119.
- [25] Greenacre, M. J. & Browne, M. W. (1986).- An Efficient Alternating Least-Squares Algorithm to Perform Multidimensional Unfolding. *Psychometrika*, 51, 241-250.
- [26] Heiser, W. J. (1991).- A Generalized Majorization Method for Least Squares MDS of Pseudodistances that May Be Negative. *Psychometrika*, 56, 7-27.
- [27] Jedidi, K. & DeSarbo, W. S. (1991).- A Stochastic MDS Procedure for the Spatial Representation of Three-Mode, Three-Way Pick Any/j Data. *Psychometrika*, 56, 471-494.
- [28] Jennrich, R. J. (1986).- A Gauss-Newton Algorithm for Exploratory Factor Analysis. *Psychometrika*, 51, 277-284.
- [29] Jones, C. L. (1983).- A Note on the Use of Directional Statistics in Weighted Euclidean Distances MDS Models. *Psychometrika*, 48, 473-476.
- [30] Kamakura, W. A. (1991).- Estimating Flexible Distributions of Ideal-Points with External Analysis of Preferences. *Psychometrika*, 56, 419-431.
- [31] Kiers, H. A. & Krijnen, W. P. (1991).- An Efficient Algorithm for Parafac of Three-Way Data with Large Numbers of Observation Units. *Psychometrika*, 56, 147-152.

- [32] Kiers, H. A. (1991).- Hierarchical Relations Among Three-Way Methods. *Psychometrika*, 56, 449-470.
- [33] Kiers, H. A., Kroonenberg, P. M. & ten Berge, J. M. F. (1992).- An Efficient Algorithm for TUCKALS3 on Data with Large Numbers of Observations Units. *Psychometrika*, 57, 415-422.
- [34] Koschat, M. A. & Swayne, D. F. (1991).- A Weighted Procrustes Criterion. *Psychometrika*, 56, 229-239.
- [35] Kruskal, J. B. & Wish, M. (1978).- Multidimensional Scaling. *Sage Publications, Inc.*
- [36] Langeheine, R. (1982).- Statistical Evaluation of Measures of Measures of Fit in the Lingoes-Borg Procrustean Individual Differences Scaling. *Psychometrika*, 47, 427-442.
- [37] Lee, S. Y. & Bentler, P. M. (1980).- Functional relations in multidimensional scaling. *British Journal of Mathematical and Statistical Psychology*, 33, 142-150.
- [38] Lingoes, C. J. (1980).- Testing Regional Hypotheses in Multidimensional Scaling. *Data Analysis and Informatics, E. Diday et al. (eds.), North-Holland*, 191-205.
- [39] Liou, M. (1991).- Assessing Statistical Accuracy in Ability Estimation: A Bootstrap Approach. *Psychometrika*, 56, 55-67.
- [40] Mardia, K. V., Kent, J. T. & Bibby, J. M. (1980).- Multivariate Analysis. *Academic Press*.
- [41] Mathar, R. (1990).- MDS with Constraints on the Configuration. *Journal of Multivariate Analysis*, 33, 151-156.
- [42] McCullagh, P. (1980).- Regression Models for Ordinal Data. *J. R. Statist. Soc. B*, 42, 109-142.
- [43] McDonald, R. P. (1983).- Alternative Weights and Invariant Parameters in Optimal Scaling. *Psychometrika*, 48, 377-391.



José Fernando Vera Vera.

- [44] Meulman, J. J. (1992).- The Integration of MDS and Multivariate Analysis with Optimal Transformations. *Psychometrika*, 57, 539-565.
- [45] Miyano, H. & Inukai, Y. (1982).- Sequential Estimation in MDS. *Psychometrika*, 47, 321-336.
- [46] Mullen, K. & Ennis, D. M. (1991).- A simple Multivariate Probabilistic Model for Preferential and Triadic Choices. *Psychometrika*, 56, 69-75.
- [47] Nishisato, S. & Sheu, W. (1984).- A Note on Dual Scaling of Successive Categories Data. *Psychometrika*, 49, 493-500.
- [48] Peay, E. R. (1988).- Multidimensional Rotation and Scaling of Configurations to Optimal Agreement. *Psychometrika*, 53, 199-208.
- [49] Ramsay, J. O. (1975).- Solving Implicit Equations in Psychometric Data Analysis. *Psychometrika*, 40, 337-360.
- [50] Ramsay, J. O. (1977).- Maximum Likelihood Estimation in Multidimensional Scaling. *Psychometrika*, 42, 241-266.
- [51] Ramsay, J. O. (1978).- Confidence Regions for MDS Analysis. *Psychometrika*, 43, 145-160.
- [52] Ramsay, J. O. (1980a).- Some Small Sample Result for Maximum Likelihood Estimation in MDS. *Psychometrika*, 45, 141-146.
- [53] Ramsay, J. O. (1980b).- The Joint Analysis of Direct Ratings, Pairwise Preferences, and Dissimilarities. *Psychometrika*, 45, 149-165.
- [54] Ramsay, J. O. (1982).- Some Statistical Approaches to MDS Data. *J. R. Statist. Soc. A*, 145, 285-312.
- [55] Ramsay, J. O. (1986).- A Proc Matrix Program for Preference-Dissimilarity MDS. *Psychometrika*, 51, 163-170.
- [56] Ramsay, J. O. & Dalzell, C. J. (1991).- Some Tools for Functional Data Analysis. *J.R. Statist. Soc. B*, 3, 539-572.
- [57] Ramsay, J. O. (1991).- Kernel Smoothing Approaches to Nonparametric Item Characteristic Curve Estimation. *Psychometrika*, 56, 611-630.

- [58] Saito, T. & Otsu, T. (1988).- A Method of Optimal Scaling for Multivariate Ordinal Data and its Extensions. *Psychometrika*, 53, 5-25.
- [59] Schiffman, S. S., Reynolds, M. L. & Young, F. W. (1981).- Introduction to Multidimensional Scaling. Theory, Methods and Applications. *Academic Press, Inc.*
- [60] Schoenberg, I. J., (1935).- Remarks to Maurice Fréchet's article "Sur la définition axiomatique d'une classe d'espaces distanciés vectoriellement applicable sur l'espace de Hilbert". *Ann. Math.*, 36, 724-732.
- [61] Schwarz, G. (1978).- Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.
- [62] Storms, G & Delbeke, L. (1992).- The Irrelevance of Distributional Assumptions on Reaction Times in MDS of Same/Different Judgment Tasks. *Psychometrika*, 57, 599-614.
- [63] Takane, Y., Young, F. W. & de Leeuw, J.(1977).- Nonmetric Individual Differences MDS: An alternating Least Squares Method with Optimal Scaling Features. *Psychometrika*, 42, 7-67.
- [64] Takane, Y. (1981).- Multidimensional Successive Categories Scaling: A Maximum Likelihood Method. *Psychometrika*, 46, 9-28.
- [65] Takane, Y. & Carroll, J. D. (1981).- Nonmetric Maximum Likelihood MDS from Directional Rankings of Similarities. *Psychometrika*, 46, 389-406.
- [66] Takane, Y. & Sergent, J. (1983).- MDS Models for Reaction Times and Same-Different Judgements. *Psychometrika*, 48, 393-423.
- [67] Ten Berge, J. M. F. & Kiers, H. A. (1991).- A Numerical Approach to the Approximate and the Exact Minimum Rank of a Covariance Matrix. *Psychometrika*, 56, 309-315.
- [68] Ten Berge, J. M. F. & Kiers, H. A. (1991).- Some Clarifications of the CANDECOP Algorithm Applied to INDSCAL. *Psychometrika*, 56, 317-326.

- [69] Thisted, R. A. (1988).- Elements of Statistical Computing. Numerical Computation. *Chapman & Hall*.
- [70] Tversky, A., Rinott, Y. & Newman, C. M. (1983).- Nearest Neighbor Analysis of Point Processes: Applications to Multidimensional Scaling. *Journal of Mathematical Psychology*, 27, 235-250.
- [71] Van der Burg, E. & de Leeuw, J. (1988).- Homogeneity Analysis With k Sets of Variables: An Alternating Least Squares Method With Optimal Scaling Features. *Psychometrika*, 53, 177-197.
- [72] Weeks, D. G. & Bentler, P. M. (1982).- Restricted MDS Models for Asymmetric Proximities. *Psychometrika*, 47, 201-208.
- [73] Winsberg, S. L. & Ramsay, J. O. (1980).- Monotonic transformations to additivity using splines. *Biometrika*, 67, 669-674.
- [74] Winsberg, S. L. & Ramsay, J. O. (1981).- Analysis of Pairwise Preference Data Using Integrated B-Splines. *Psychometrika*, 46, 171-186.
- [75] Winsberg, S. L. & Ramsay, J. O. (1983).- Monote Spline Transformations for Dimension Reduction. *Psychometrika*, 48, 575-595.
- [76] Winsberg, S. L., Carroll, J. D. & Cohen, H. S. (1984).- Confidence Regions for Indscal Using the Jackknife and Bootstrap Techniques. *Psychometrika*, 49, 475-491.
- [77] Winsberg, S. & Carroll, J. D. (1989).- A Cuasi-Nonmetric Method for Multidimensional Scaling of Multiway Data via a Restricted case of an Extended INDSCAL model. *Multiway Data Analysis, North-Holland*, 405-414.
- [78] Winsberg, S. & Carroll, J. D. (1989).- A cuasi-Nonmetric Method for Multidimensional Scaling via an Extended Euclidean Model. *Psychometrika*, 54, 217-229.
- [79] Wright, I. W. & Wegman, E. J. (1980).- Isotonic, Convex and Related Splines. *Ann. Statist.*, 8, 1023-1035.

- [80] Yen, W. M., Burket, G.R. & Sykes, R.C. (1991).- Nonunique Solutions to the Likelihood Equation for the Three-Parameter Logistic Model. *Psychometrika*, 56, 39-54.
- [81] Young, G. & Householder, A. S., (1938).- Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 3, 19-22.
- [82] Zacks, S. (1971).- The Theory of Statistical Inference. *John Wiley & Sons, Inc.*
- [83] Zinnes, J. L. & MacKay, D. B. (1983).- Probabilistic MDS: Complete and Incomplete Data. *Psychometrika*, 48, 27-48.