

**UTILIZACIÓN DE MÉTRICAS RIEMANNIANAS
EN ANÁLISIS DE DATOS MULTIDIMENSIONALES
Y SU APLICACIÓN A LA BIOLOGÍA**

JOSE M^a OLLER SALA

BARCELONA, 25 de NOVIEMBRE de 1982.

UTILIZACIÓN DE MÉTRICAS RIEMANNIANAS EN
ANÁLISIS DE DATOS MULTIDIMENSIONALES Y
SU APLICACIÓN A LA BIOLOGÍA

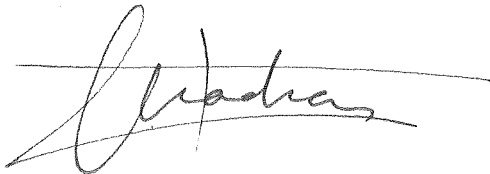
Memoria presentada para optar
al grado de Doctor en Biología
por la Universidad de Barcelona,
por

José M^a Oller Sala

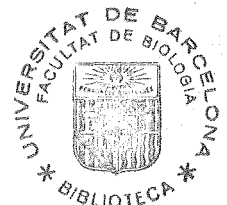


V^oB^o

EL DIRECTOR



Dr. Carlos Ma. Cuadras Avellana
Profesor Agregado de Bioestadística
Facultad de Biología
Universidad de Barcelona



Barcelona, 25 de Noviembre de 1982..



*A mis familiares,
a mis amigos,
a mis maestros.*

Ubi materia, ibi geometria.

J. Kepler.

INDICE

PROLOGO	1
1. DISTANCIAS ESTADISTICAS	3
2. ALGUNOS ASPECTOS DEL ANALISIS TENSORIAL Y GEOMETRIA DIFERENCIAL	28
3. UNA DISTANCIA RIEMANNIANA ENTRE FUNCIONES DE DENSIDAD. RESULTADOS GENERALES	58
4. DISTANCIAS ENTRE DISTRIBUCIONES DISCRETAS	87
5. DISTANCIAS ENTRE DISTRIBUCIONES ABSOLUTAMENTE CONTINUAS ...	144
6. REPRESENTACION EN DIMENSION REDUCIDA	172
7. UNA APLICACION A LA METODOLOGIA ETOLOGICA	191
8. UNA APLICACION AL DIAGNOSTICO AUTOMATIZADO	223
9. UNA ALTERNATIVA AL TEST T DE STUDENT PARA MUESTRAS INDEPENDIENTES	251
10. RESUMEN DE LOS RESULTADOS	274
BIBLIOGRAFIA	279

PROLOGO

Esta Memoria es una aportación al Análisis Multivariante y su aplicación a la Biología. Los numerosos índices de disimilaridad y distancias que se utilizan en Estadística y son aplicados en Genética, Antropología, Ecología, etc., me han motivado a intentar estudiar una distancia entre poblaciones estadísticas paramétricas, de aplicación general, y que posea buenas propiedades matemáticas.

En la primera parte (cap. 1 al 6), se desarrolla esta distancia estadística, definible para una clase muy general de funciones de densidad paramétricas, a través de la matriz de información de Fisher, hallando algunas de sus propiedades básicas y calculándola para ciertas distribuciones de probabilidad concretas.

En la segunda parte (cap. 7 al 9), se aplican algunos de los resultados obtenidos, proponiendo una metodología estadística para el tratamiento de tablas de contingencia multidimensionales, asociadas a experiencias etológicas, ilustrándolo con el estudio de la conducta agonística del lúgano. También se propone un algoritmo utilizable para el diagnóstico de enfermedades, a partir de los resultados de unos análisis, aplicándolo al diagnóstico de ciertas enfermedades hematológicas a través de la interpretación de mielogramas. Finalmente se considera una alternativa al test t de Student para muestras independientes y se ilustra con el estudio de la relación entre la alcohol deshidrogenasa y el tamaño en *Drosophila melanogaster*.



Debo expresar mi agradecimiento al director de dicha Memoria, Dr. Carlos M. Cuadras por sus consejos y orientaciones, a la Dra. Mercedes Durfort por la revisión de una parte del manuscrito, al Dr. Luís Serra por permitirme utilizar datos experimentales de su tesis doctoral, a Juan C. Senar por haber hecho lo mismo con los datos de su tesina. A todo el Departamento de Bioestadística en general y a los profesores Francisco Carmona, José Fortiana y Martín Rios en especial, por la colaboración prestada y finalmente a todos los que me han alentado en la realización de dicha Memoria.

Barcelona, 25 de Noviembre de 1982.

1. DISTANCIAS ESTADÍSTICAS

Resumen:

En el presente capítulo se expone, con caracter introductorio, el concepto de distancia y su utilización en el Análisis de Datos, y en Estadística, citando en particular aplicaciones a diversas areas de la Biología.

Sumario:

1.1. OBJETO.

1.2. PROPIEDADES GENERALES.

1.2.1. Concepto de distancia.

1.2.2. Similaridades y disimilaridades.

1.2.3. Condiciones de euclidianidad.

1.3. DISTANCIAS COMUNMENTE USADAS.

1.4. DISCUSION.

1.1. OBJETO.

Para estudiar de forma cuantitativa las analogías y diferencias entre diversos objetos, que pueden ser individuos, poblaciones, etc., resulta útil introducir el concepto de distancia entre los mismos, el cual es a la vez un concepto intuitivamente claro, por estar muy relacionado con la percepción del espacio físico, y además muy bien formalizado a nivel matemático.

Veamos ahora como podemos utilizar la distancia que hayamos definido para una mejor comprensión de las relaciones entre los distintos objetos. En primer lugar, una distancia es el instrumento adecuado para poder efectuar una clasificación de los objetos, agrupando los más próximos. Estas clasificaciones pueden ser jerárquicas o no jerárquicas. En las clasificaciones no jerárquicas se agrupan los objetos próximos, pero no se pretende definir, a partir de la distancia original, una distancia entre los grupos formados, usualmente porque ello no es posible sin deformar la distancia con que trabajamos. En las clasificaciones jerárquicas si es posible distanciar los grupos formados, distancia que sirve para agrupar a éstos, repitiéndose el proceso las veces necesarias hasta formar un único grupo, con todos los objetos estudiados.

Otra de las posibilidades de la definición de una distancia entre unos objetos, consiste en poder efectuar algún tipo de análisis discriminante. En efecto, si disponemos de unos grupos de objetos bien establecidos y queremos clasificar a un objeto nuevo en uno de los grupos anteriores, podemos hacerlo en base a la distancia definida, pudiéndose seguir varios procedimientos: caracterizar a cada grupo por un "objeto medio" y

asignar el objeto a clasificar al grupo cuyo "objeto medio" sea más próximo, o bien calcular la distancia media del objeto a clasificar al grupo, y asignar al primero, al grupo cuya distancia media sea menor, o si es posible hablar de distancia del objeto al grupo, caso de disponer de una distancia que permita, sin deformarla, establecer una clasificación jerárquica, clasificar al objeto dentro del grupo más próximo.

Una vez introducida una distancia será posible obtener representaciones gráficas de los objetos como puntos en un plano (generalmente), según algún criterio de optimización, de forma que la distancia entre estos puntos sea lo más parecida posible a la distancia originalmente definida entre los objetos, como en el Análisis de Coordenadas Principales o el Multidimensional Scaling, Cuadras (1981), dependiendo, la técnica usada, del tipo de distancia con que trabajemos. Dicha representación gráfica nos ayudará a visualizar las relaciones entre los distintos objetos. También será posible obtener como salida gráfica un dendograma, imagen de la clasificación jerárquica posiblemente establecida.

Después de haber identificado a los objetos como puntos de algún espacio, es también posible buscar interpretaciones causales de la variabilidad obtenida, por direcciones, atendiendo principalmente a las direcciones que absorben mayor variabilidad, en orden a encontrar unas pocas causas que expliquen las analogías y diferencias entre los objetos.

1.2. PROPIEDADES GENERALES.

Vamos a examinar a continuación el problema de la construcción de una distancia. En primer lugar, deberemos caracterizar adecuadamente a los

objetos que queremos distanciar. Sea (Ω, \mathcal{A}, P) un espacio de probabilidad. Con Cuadras (1974) a los elementos de Ω les denominaremos individuos, en sentido estadístico, que puede diferir notablemente del concepto biológico de individuo, mientras que a Ω le denominaremos población, también en sentido estadístico, que difiere con frecuencia del concepto biológico de población: un individuo biológico puede ser una población en el sentido estadístico. Generalmente los objetos a distanciar los podremos identificar con individuos o poblaciones estadísticas. Una vez caracterizados, para poder introducir una distancia, necesitamos disponer de información sobre los mismos, necesitamos de unas variables aleatorias definidas en el espacio de probabilidad, o espacios de probabilidad, con los que trabajamos, y en base a los resultados obtenidos de estas variables aleatorias construiremos una distancia. En este punto hay que hacer notar que la distancia dependerá tanto del objeto en si, como de las variables que hemos definido, esto es, de la información recibida, ya que no es posible la noción de distancia entre dos entes si no disponemos de información sobre las analogías y diferencias entre los mismos.

1.2.1. Concepto de distancia.

A continuación vamos a exponer algunas de las propiedades más importantes que pueden presentar las distancias. Serán necesarias algunas formalizaciones. Sea $E = \{e_1, \dots, e_n\}$ el conjunto de los objetos que deseamos distanciar, nótese que los e_i los podemos identificar con individuos o poblaciones estadísticos. Llamaremos distancia métrica a una aplicación:

$$d: E \times E \longrightarrow \mathbb{R}$$

que verifica:

- 1) $d(e_i, e_j) \geq 0 \quad i, j \in \{1, \dots, n\}$
- 2) $d(e_i, e_j) = 0 \Leftrightarrow i = j$ (definida positiva)
- 3) $d(e_i, e_j) = d(e_j, e_i) \quad i, j \in \{1, \dots, n\}$ (simetría)
- 4) $d(e_i, e_j) \leq d(e_i, e_k) + d(e_k, e_j) \quad i, j, k \in \{1, \dots, n\}$ (desigualdad triangular).

Si la propiedad 2 se sustituye por:

- 5) $i = j \Rightarrow d(e_i, e_j) = 0$ (semidefinida positiva)

entonces llamaremos a d una distancia pseudométrica.

Si se verifica además la propiedad:

- 6) $d(e_i, e_j) \leq \sup \{d(e_i, e_k), d(e_k, e_j)\} \quad \forall i, j, k \in \{1, \dots, n\}$
(desigualdad ultramétrica).

diremos que d es una distancia ultramétrica. Es bien sabido que es preciso disponer de una distancia ultramétrica para obtener una clasificación jerárquica, aunque cuando no dispongamos de ella existen varios algoritmos que construyen una distancia ultramétrica, parecida a la distancia originalmente definida, Cuadras (1980).

Sea (F, π) un espacio euclídeo m -dimensional, siendo π la aplicación producto escalar, entonces sí se verifica:

- 7) $\exists \varphi, \quad \varphi: E \rightarrow F \quad / \quad d(e_i, e_j) = \sqrt{\pi(\varphi(e_i) - \varphi(e_j), \varphi(e_i) - \varphi(e_j))}$
 $\forall i, j \in \{1, \dots, n\}$.

diremos que la distancia métrica definida en E es euclídea. Como cualquier espacio euclídeo de dimensión m es isométrico a \mathbb{R}^m con el producto escalar

usual, la propiedad 7 puede expresarse diciendo que para cada objeto e_i es posible asociarle un elemento \mathbb{R}^m de coordenadas, respecto la base canónica, (x_{i1}, \dots, x_{im}) tal que:

$$d(e_i, e_j) = \sqrt{\sum_{h=1}^m (x_{ih} - x_{jh})^2} \quad \forall i, j \in \{1, \dots, n\}$$

Nótese que si se cumple la propiedad 7, será posible escoger $m \leq n-1$, puesto que n puntos en un espacio euclídeo están incluidos en un subespacio a lo sumo $n-1$ dimensional.

1.2.2. Similaridades y Disimilaridades.

Hay aplicaciones $d: E \times E \rightarrow \mathbb{R}$ que no son métricas ni pseudométricas, pero que pueden también utilizarse para medir las analogías y diferencias entre objetos. A estas aplicaciones las llamaremos disimilaridades y han de verificar, al menos, las propiedades 1, 3 y 5. Gran número de disimilaridades se construyen a partir de similaridades. Veamos seguidamente el concepto de similaridad. Una similaridad es una medida de la analogía entre dos objetos, en base a datos cualitativos, usualmente variables dicotómicas, a dos valores, que podemos designar convencionalmente por + y -. Dados dos objetos se construye la tabla de frecuencias:

		e_j	
		+	-
e_i	+	a_{ij}	b_{ij}
	-	c_{ij}	d_{ij}

con $n = a_{ij} + b_{ij} + c_{ij} + d_{ij}$, donde n es el total de caracteres estudiados, a_{ij} el número de dobles presencias, d_{ij} las dobles ausencias, y b_{ij} y c_{ij} las presencias-ausencias o ausencias-presencias. Formalmente, una similaridad es una aplicación:

$$s: E \times E \longrightarrow \mathbb{R}$$

que cumple:

- 8) $s(e_i, e_j) = f_n(a_{ij}, b_{ij}, c_{ij})$
- 9) f_n es monótona creciente en a_{ij}
- 10) f_n es monótona decreciente en b_{ij} y c_{ij}
- 11) f_n es simétrica en b_{ij} y c_{ij} : $f_n(a_{ij}, b_{ij}, c_{ij}) = f_n(a_{ij}, c_{ij}, b_{ij})$

Puesto que el dominio de definición de f_n es finito, f_n está acotada. Sea α_n el supremo del conjunto de las imágenes de f_n , entonces si definimos:

$$d(e_i, e_j) = \alpha_n - s(e_i, e_j)$$

tendremos que $d(e_i, e_j)$ es una disimilaridad pues verifica evidentemente las propiedades 1, 3 y 5. A partir de una similaridad vemos pues que puede construirse siempre una disimilaridad.

1.2.3. Condiciones de euclidianidad.

Vamos a examinar ahora el problema de determinar cuando se cumple la propiedad 7, es decir, si se trata de una distancia euclídea en el sentido de que existe una configuración de puntos en un espacio euclídeo cuyas interdistancias coincidan con las interdistancias obtenidas con la distancia original. Este problema puede ser analizado de modo diverso, que de--

pende de como haya sido definida la distancia en E.

Consideremos en primer lugar la posibilidad de que hallamos definido una aplicación:

$$\tau : E \longrightarrow H$$

siendo H un espacio normado, es decir un espacio vectorial, generalmente construido sobre el cuerpo real, en el que hay definida una aplicación, denominada norma:

$$H \xrightarrow{\|\cdot\|} \mathbb{R}$$

que verifica:

$$12) \quad \|x\| \geq 0 \quad \forall x \in H \quad \text{y} \quad \|x\| = 0 \Leftrightarrow x = 0$$

$$13) \quad \|x+y\| \leq \|x\| + \|y\| \quad \forall x, y \in H$$

$$14) \quad \|\lambda \cdot x\| = |\lambda| \cdot \|x\| \quad \forall x \in H \quad \forall \lambda \in \mathbb{R}$$

En estas condiciones podemos definir una distancia métrica en E, a partir de la igualdad:

$$d(e_i, e_j) = \|\tau(e_i) - \tau(e_j)\|$$

entonces para saber si se cumple la propiedad 7, puede comprobarse si la norma definida en H puede expresarse como la raíz cuadrada de un producto escalar de un vector consigo mismo:

$$\|x\| = \sqrt{\langle x, x \rangle}$$

hecho que es equivalente a probar, Collatz (1966), que se verifica la regla del paralelogramo a saber:

$$\|x + y\|^2 - \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2)$$

en caso afirmativo se cumplirá evidentemente la propiedad 7; pero en caso negativo no podremos afirmar que no se cumple: se trata de una condición suficiente para la propiedad 7, no necesaria, pues aunque la norma no provenga de un producto escalar, según las imágenes de los e_i que tengamos, $\tau(e_i)$, si puede verificarse la propiedad 7. Consideremos el caso de distanciar sólo dos objetos.

Para investigar condiciones necesarias y suficientes hemos de partir de las distancias entre los objetos, prescindiendo de cómo hayan sido calculadas.

Sea $\delta_{ij} = d(e_i, e_j)$, formemos la matriz $\Delta = (\delta_{ij})$, siendo Δ una matriz de orden n , simétrica, con ceros en la diagonal principal. Entonces, si llamamos I_n a la matriz identidad de orden n , $U = (1, \dots, 1)^t$ un vector columna con unos, y definimos $A = (a_{ij})$ como una matriz cuadrada de orden n , con $a_{ij} = -\frac{1}{2} \delta_{ij}^2$, resulta que si llamamos $H = I_n - \frac{1}{n} U \cdot U^t$ y formamos la matriz $B = H A H$, entonces podemos afirmar que B es semidefinida positiva de rango $m \leq n-1$ si y sólo si se verifica la propiedad 7, Cuadras (1981).

Además es posible obtener explícitamente las coordenadas de las imágenes de los e_i por la aplicación φ , diagonalizando B . Por ser B simétrica, siempre diagonaliza ortogonalmente:

$$B = T D_\lambda T^t = T D_\lambda^{1/2} D_\lambda^{1/2} T^t = (T D_\lambda^{1/2}) (T D_\lambda^{1/2})^t$$

siendo $D_\lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$ $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m > 0$ y T una matriz de orden $n \times m$ cuyas columnas son vectores propios asociados a $\lambda_1 \dots \lambda_m$ res-

pectivamente, y de norma unitaria. Entonces, si llamamos $X = TD_{\lambda}^{1/2}$, matriz de orden $(n \times m)$, resulta que X es una matriz cuyas columnas son vectores propios asociados a los valores propios $\lambda_1 \dots \lambda_m$ y además λ -normalizados. Además, podemos afirmar, Cuadras (1981), que sus filas definen puntos en \mathbb{R}^m tales que sus interdistancias coinciden con las interdistancias entre los $e_i \in E$, es decir las coordenadas de $\varphi(e_i)$, respecto la base canónica de \mathbb{R}^m , son $(x_{i1}, x_{i2}, \dots, x_{im})$. Además, con el cálculo de X , resolvemos también el problema de la representación de los objetos en un subespacio euclídeo de dimensión baja, usualmente de dimensión 2, ya que si queremos hallar un subespacio de \mathbb{R}^m , k -dimensional, tal que al proyectar los $\varphi(e_i)$ en este subespacio, la suma de los cuadrados de las interdistancias entre los puntos proyectados en este subespacio sea máxima, bastará coger las k primeras columnas de la matriz X , que definen las coordenadas de n puntos en un espacio k -dimensional.

Si no se cumple la propiedad 7, entonces B posee valores propios negativos:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0 > \lambda_{r+1} \geq \dots \geq \lambda_m$$

entonces si procedemos de forma análoga, en el cálculo de $D_{\lambda}^{1/2}$ aparecerán números imaginarios puros, y la matriz X resultante será de la forma:

$$X = \begin{pmatrix} x_{11} & \dots & x_{1r} & ix_{1r+1} & \dots & ix_{1m} \\ \vdots & & \vdots & \vdots & & \vdots \\ x_{n1} & \dots & x_{nr} & ix_{nr+1} & \dots & ix_{nm} \end{pmatrix}$$

siendo $i = \sqrt{-1}$. A cada objeto le podemos asociar, análogamente al caso anterior una m -plas de la forma:

$$e_j \rightarrow (x_{j1}, \dots, x_{jr}, ix_{jr+1}, \dots, ix_{jm})$$

en otras palabras no los podemos representar como puntos en un espacio euclídeo de ejes reales, ya que es preciso que alguno de éstos sea imaginario puro. Representaciones en dimensión reducida pueden obtenerse escogiendo, por separado, los ejes reales asociados a los mayores valores propios y los ejes imaginarios asociados a los menores valores propios.

Otra posibilidad consiste en efectuar transformaciones monótonas de la distancia definida inicialmente, hasta obtener otra que sea euclídea (ver capítulo 6).

Hemos visto a grandes rasgos las propiedades generales de las distancias, pasemos a continuación a hacer una revisión de algunas de las más importantes en cuanto a las aplicaciones a la Biología se refiere.

1.3. DISTANCIAS COMUNMENTE USADAS.

Si asignamos a cada objeto e_i , un elemento de \mathbb{R}^m , de la forma siguiente: si e_i es un individuo estadístico, le asociamos $(X_1(e_i), \dots, X_m(e_i)) \in \mathbb{R}^m$ siendo X_1, \dots, X_m variables aleatorias definidas en la población estadística Ω a la que pertenecen los e_i . Si cada objeto e_i es una población estadística, la podemos asociar la m -pla $(E(X_1), \dots, E(X_m)) \in \mathbb{R}^m$, donde las esperanzas de las variables aleatorias X_1, \dots, X_m están restringidas a la población estadística e_i . En estas condiciones podemos definir como distancia entre dos objetos e_i, e_j cuyas coordenadas en \mathbb{R}^m son (x_{i1}, \dots, x_{im}) y (x_{j1}, \dots, x_{jm}) ,

$$d(e_i, e_j) = \sqrt{\sum_{h=1}^m (x_{ih} - x_{jh})^2}$$

Este criterio es usado en el Análisis de Componentes Principales, Pearson (1901), Hotelling (1933), Rao (1964), Cuadras (1981).

Sin embargo, dicha distancia presenta inconvenientes en muchas aplicaciones, uno de los mismos es que si efectuamos un cambio de escala en las medidas de las variables, la distancia cambia de forma no monótona, en el sentido que no se conserva la preordenación asociada a la distancia, Cuadras (1981), Legendre (1979). Este inconveniente puede ser resuelto por el uso de variables tipificadas, es decir, centradas y reducidas. Otro inconveniente es que ignora el hecho de que las variables aleatorias sean o no estocásticamente dependientes, hecho que parece conveniente tener en cuenta, puesto que cuanto mayor sea el grado de dependencia entre dos variables, menos información obtenemos de una cualquiera de ellas, al conocer el valor que ha tomado la otra.

Pueden encontrarse ejemplos de utilización de esta distancia en Williams & Stephenson (1973), Stephenson et al. (1974), aplicados a la ecología marina, y en general en cualquier aplicación del Análisis de Componentes Principales, se diga explícitamente o no, como en el trabajo de Jolicoeur y Mosimann (1960) analizando el tamaño y la forma de *Chrysemys picta marginata*, o en el trabajo de Gittins (1969), donde se estudian diversas comunidades vegetales, en base a la cobertura de distintas especies vegetales.

Una de las primeras distancias que intentó corregir los defectos de la distancia euclídea ordinaria, fue el coeficiente de semejanza racial, desarrollado por Pearson (1926), pensado originariamente para estudios antropológicos. Los objetos a distanciar son poblaciones estadísticas,

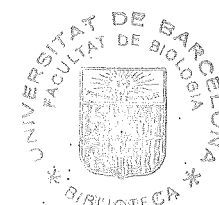
en las que hay definidas X_1, \dots, X_m variables aleatorias. La distancia entre el objeto e_i y el objeto e_j , se define como:

$$d(e_i, e_j) = \sqrt{\frac{1}{m} \sum_{h=1}^m \frac{n_{ih} n_{jh}}{(n_{ih} + n_{jh})} \left(\frac{\bar{x}_{ih} - \bar{x}_{jh}}{s_h} \right)^2}$$

siendo \bar{x}_{ih} y \bar{x}_{jh} los valores medios de la variable h en las poblaciones estadísticas e_i y e_j respectivamente, n_{ih} y n_{jh} los tamaños muestrales usados en el cálculo de las medias \bar{x}_{ih} y \bar{x}_{jh} ; en cuanto a s_h es la desviación típica de la variable h , supuesta común en todas las poblaciones. Esta distancia es invariante frente a cambios de escala de las variables, pero no es invariante frente a transformaciones lineales en general. Además, como la distancia euclídea ordinaria, presenta el problema de prescindir del hecho de que las variables aleatorias observables puedan ser estocásticamente dependientes o no. Una discusión sobre esta distancia puede verse en Rao (1948 b) y Mardia (1977).

Mahalanobis, en su ya clásico trabajo de 1936, desarrolla una distancia entre poblaciones estadísticas, en las que tenemos definidas X_1, \dots, \dots, X_m variables aleatorias que se distribuyen según una distribución normal conjunta, con la hipótesis adicional de que todas las poblaciones poseen la misma matriz de varianzas y covarianzas, Σ . Sean μ_i y μ_j vectores columna, m -dimensionales, cuyas componentes son las esperanzas de las variables aleatorias condicionadas a las poblaciones e_i y e_j . Entonces se define la distancia de Mahalanobis entre e_i y e_j como:

$$d(e_i, e_j) = \sqrt{(\mu_i - \mu_j)^t \Sigma^{-1} (\mu_i - \mu_j)}$$



suponiendo que Σ^{-1} exista. En caso de que Σ sea singular, es decir, que existan relaciones lineales entre las variables, se sustituirá Σ^{-1} por una g-inversa de Σ , Σ^- . Equivale a representar a las poblaciones como puntos en un espacio euclídeo, cuyo producto escalar dependa de las correlaciones entre las variables. En la práctica se sustituyen μ_i y μ_j por sus estimaciones máximo-verosímiles y Σ por una estimación insesgada.

Entre sus principales propiedades destaca la de resultar invariante frente a transformaciones lineales, no singulares, de las variables, en particular por cambios de escala. Además, si llamamos D_m a la distancia de Mahalanobis calculada a partir de X_1, \dots, X_m , D_n a dicha distancia calculada a partir de Y_1, \dots, Y_n , donde dicho conjunto de variables aleatorias es estocásticamente independiente al anterior y D_{n+m} a la distancia de Mahalanobis global, calculada con todas las variables, $X_1, \dots, X_m, Y_1, \dots, Y_n$, entonces se verifica:

$$D_{n+m}^2 = D_n^2 + D_m^2$$

Otra ventaja importante es que el estadístico \hat{D}^2 (distancia estimada de Mahalanobis al cuadrado) está relacionado con la T^2 de Hotelling, Hotelling (1931), ya que si $\mu_1 = \mu_2$, el estadístico:

$$F = \frac{N_1 N_2 (N_1 + N_2 - m - 1)}{m (N_1 + N_2) (N_1 + N_2 - 2)} \hat{D}^2$$

sigue una distribución F con $m-1$ y $N_1 + N_2 - m - 1$ grados de libertad, siendo N_1 y N_2 los tamaños muestrales en las poblaciones e_1 y e_2 respectivamente y m el número de variables estudiadas.

La distancia de Mahalanobis está relacionada con el coeficiente de semejanza racial de Pearson, siendo proporcionales ambas distancias en el caso de que las variables aleatorias con que trabajamos esten incorrelacionadas, Cuadras (1981). Uno de los aspectos más interesantes de esta distancia es que relaciona el concepto de independencia estocástica con el concepto de ortogonalidad, ya que Σ^{-1} es diagonal si y sólo si las variables aleatorias observables son estocásticamente independientes, pues entonces y sólo entonces se anulan todas las covarianzas.

La distancia de Mahalanobis juega un papel fundamental en muchas areas del Análisis Multivariante, principalmente en el Análisis Canónico de Poblaciones y el Análisis Discriminante, por ello ha sido ampliamente utilizada en aplicaciones a la Biología como puede verse en los trabajos de Rao (1948a), estudio antropométrico sobre diversas castas indias, Jolicoeur (1959), análisis de la variación geográfica en *Canis lupus*, Baker, Atchley & McDaniel (1972), estudio de la influencia del cariotipo en la morfometría de *Uroderma bilobatum*, Petitpierre & Cuadras (1977), estudio sistemático a partir de datos biométricos de coleópteros del género *Timarcha* y podemos citar finalmente a Casinos & Ocaña (1979) que efectuan un análisis craneométrico en cetáceos de género *Inia*.

Aunque la distancia de Mahalanobis se definió originalmente para distanciar poblaciones normales multivariantes, con matriz de varianzas-covarianzas común, puede usarse también para otras distribuciones, aunque entonces se pierden algunas de las propiedades, particularmente las que hacen referencia a las distribuciones de las estimaciones de dicha distancia, y también deja de cumplirse la estrecha relación entre independencia estocástica y ortogonalidad.

El tomar la inversa de la matriz de varianzas y covarianzas como la matriz que define el producto escalar en el "espacio de las medias" puede presentar dificultades cuando éstas estén relacionadas funcionalmente con dicha matriz, tal es el caso de trabajar con variables aleatorias que sigan una distribución multinomial, entonces puede tomarse como aproximación de la matriz de varianzas y covarianzas la que se obtiene a partir de las muestras obtenidas en cada una de las dos poblaciones que se comparan, o tomar una matriz de varianzas y covarianzas "promedio", Balakrishnan & Sanghvi (1968), Kurczynski (1970), Krzanowski (1971).

Podemos citar a continuación la métrica ji-cuadrado. Sean H_1, \dots, H_k poblaciones estadísticas y A_1, \dots, A_n caracteres (sucesos) observables en cada población. Si llamamos f_{ij} a la frecuencia absoluta del carácter j en la población i , $f_{i.} = \sum_{j=1}^m f_{ij}$, $f_{.j} = \sum_{i=1}^k f_{ij}$, entonces, puede definirse la distancia entre H_i y H_m como:

$$d^2(H_i, H_m) = \sum_{j=1}^n \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{mj}}{f_{m.}} \right)^2$$

análogamente puede definirse la distancia entre A_j y A_m como:

$$d^2(A_j, A_m) = \sum_{i=1}^k \frac{1}{f_{i.}} \left(\frac{f_{ij}}{f_{.j}} - \frac{f_{im}}{f_{.m}} \right)^2$$

Estas distancias son euclídeas en el sentido de cumplirse la propiedad 7, Lefebvre (1976), y constituyen la base para la formulación del Análisis Factorial de Correspondencias, Benzecri (1973). Han sido ampliamente utilizadas, a través de éste último, en Biología, como puede verse

en Alonso (1975), donde se realiza un estudio sobre la distribución geográfica del polimorfismo cromosómico en *Drosophila subobscura*, Binet et al. (1972), que efectúa un análisis sobre diversas especies de copépodos, pelágicos recogidos en distintas estaciones a lo largo de distintas épocas del año, Fondevila et al. (1981), donde se estudia el proceso de colonización de Europa y Africa por *Drosophila buzzati*, originaria de sud América, etc.

Una distancia definida para distanciar poblaciones estadísticas sobre las que observamos un conjunto de variables aleatorias, X_1, \dots, X_m distribuidas de acuerdo con una multinomial, fue propuesta heurísticamente por Bhattacharyya, Bhattacharyya (1946). Esta distancia está basada en la idea de asociar a cada población un vector en un espacio euclídeo de dimensión m , cuyos cosenos directores al cuadrado coincidan con las distintas probabilidades, p_1, \dots, p_m que definen la distribución multinomial. La distancia entre dos poblaciones vendrá dada por el ángulo entre estos dos vectores, igual a la distancia, euclídea, inducida sobre la superficie de una hiperesfera de radio unidad, entre los dos puntos que definen los dos vectores al cortar la superficie de dicha hiperesfera. La expresión analítica de la distancia entre dos poblaciones multinomiales de parámetros p_1, \dots, p_m y q_1, \dots, q_m respectivamente es:

$$d = \arccos \left(\sum_{i=1}^m \sqrt{p_i q_i} \right)$$

en la práctica se sustituirán los p_i y q_i por sus estimaciones máximo-verosímiles. Dicha distancia ha sido utilizada por Edwards y Cavallisforza (1964), como una distancia genética, distanciendo poblaciones de

seres vivos caracterizados por las frecuencias relativas de distintos tipos de genes o cromosomas, intentando sacar consecuencias sobre la evolución biológica de las mismas. Variantes de dicha distancia se han propuesto, Edwards (1971), Orloci (1967), sustituyendo la medida de la longitud entre los dos puntos sobre la superficie de una hiperesfera, por la longitud de la cuerda que los une. Anteriormente, Rao, Rao (1948a), propuso, también de forma heurística, la extensión de esta distancia a distribuciones absolutamente continuas. Sean f y g las funciones de densidad conjunta de las variables aleatorias X_1, \dots, X_m , entonces definimos las distancias entre ambas como:

$$d = \arccos \left(\int_{\mathbb{R}^m} \sqrt{f \cdot g} \right)$$

A continuación vamos a examinar la distancia de Minkowski. Una vez caracterizados los dos objetos que queremos distanciar, e_i, e_j por las coordenadas: (x_{i1}, \dots, x_{in}) y (x_{j1}, \dots, x_{jn}) , se define la distancia de Minkowski entre ellos como:

$$d_r(e_i, e_j) = \left[\sum_{h=1}^n |x_{ih} - x_{jh}|^r \right]^{1/r}$$

siendo r un número natural arbitrario. Nótese que para $r=2$, la distancia de Minkowski coincide con la distancia euclídea ordinaria. Puede comprobarse que, salvo para este caso, la distancia d_r no puede expresarse en función de la raíz cuadrada de un producto escalar, por no verificar la norma de Minkowski asociada a dicha distancia, la regla del paralelogramo, Cuadras (1981). Examinemos el caso $r=1$. Este caso se conoce con el nombre específico de distancia ciudad:

$$d_1(e_i, e_j) = \sum_{h=1}^n |x_{ih} - x_{jh}|$$

y ha sido utilizada como distancia genética por Prevosti et al. (1975), Fontdevila, et al. (1981).

Tiene también interés examinar el caso límite cuando $r \rightarrow \infty$, distancia de Tchebychef, entonces es fácil comprobar que:

$$d_{\infty}(e_i, e_j) = \max_h |x_{ih} - x_{jh}|$$

Existen muchas variantes de estas distancias, que van desde tomar el promedio, respecto el total de variables:

$$d(e_i, e_j) = \frac{1}{n} \sum_{h=1}^n |x_{ih} - x_{jh}|$$

propuesta por el antropólogo Czekanowski, Czekanowski(1909), hasta la conocida con el nombre de métrica de Cambera, Lance & Williams (1966):

$$d(e_i, e_j) = \sum_{h=1}^n \frac{|x_{ih} - x_{jh}|}{|x_{ih} + x_{jh}|}$$

o bien el coeficiente de divergencia, Clark (1952):

$$d(e_i, e_j) = \sqrt{\frac{1}{n} \sum_{h=1}^n \left(\frac{x_{ih} - x_{jh}}{x_{ih} + x_{jh}} \right)^2}$$

Podemos citar también las distancias o disimilaridades que se construyen a partir de una similaridad. Algunos de los coeficientes de similaridad más conocidos son:

$$S_1 = \frac{a+d}{a+b+c+d} \quad (\text{Sokal y Michener})$$

$$S_2 = \frac{a+d}{a+2b+2c+d} \quad (\text{Roger y Tamimoto})$$

$$S_3 = \frac{2a+2d}{2a+b+c+2d} \quad (\text{Sokal y Sneath})$$

$$S_4 = \frac{a}{a+b+c} \quad (\text{Jaccard})$$

$$S_5 = \frac{a}{a+b+c+d} \quad (\text{Russell y Rao})$$

donde a son las dobles presencias, d las dobles ausencias y b y c las presencias-ausencias. Todas estas similaridades están comprendidas entre 0 y 1, por tanto podemos construir una disimilaridad a partir de:

$$d_i = 1 - S_i$$

Las distancias obtenidas a partir de similaridades no son necesariamente métricas, ni semimétricas.

Citemos también una disimilaridad propuesta cuando se tienen variables cuantitativas y cualitativas X_1, \dots, X_n , Gower (1971)

$$d(e_i, e_j) = \frac{1}{n} \sum_{h=1}^n \frac{1}{R_h} |x_{ih} - x_{jh}|$$

donde R_h es el recorrido de X_h .

Una distancia propuesta por Matusita para distanciar poblaciones estadísticas, Matusita (1966), se desarrolla a partir de las funciones de

densidad de las variables aleatorias definidas en dichas poblaciones.

Sean f_i y f_j las funciones de densidad asociadas a las poblaciones e_i, e_j , entonces se define:

$$d(e_i, e_j) = \sqrt{\int_E (\sqrt{f_i} - \sqrt{f_j})^2 dm}$$

siendo E un espacio sobre el que se han definido f_i y f_j , usualmente \mathbb{R}^n , y m una medida en E.

Un índice de desemejanza entre funciones de distribución es el de Cramer-Von Mises:

$$d = \sqrt{\int_{\mathbb{R}^m} (F-G)^2}$$

siendo F y G funciones de distribución m-variantes. Otro índice es el de Kolmogoroff-Smirnoff:

$$d = \sup_{(X_1 \dots X_m)} |F-G|$$

Una medida de desemejanza entre funciones de densidad es la llamada divergencia:

$$d = \int_{\mathbb{R}^m} \ln \left(\frac{f}{g} \right) (f-g)$$

donde f y g son funciones de densidad m-variantes.

Para distanciar también poblaciones estadísticas, puede considerarse, Rao (1948 b), las funciones de densidad f_1 y f_2 asociadas a las varia--

bles aleatorias X_1, \dots, X_m , condicionadas a las poblaciones e_1 y e_2 y determinar unas regiones mutuamente excluyentes R_1, R_2 tales que la pro babilidad:

$$\int_{R_1} f_2 = \int_{R_2} f_1 = \alpha$$

sea mínima. Estas regiones podrán usarse para clasificar un individuo en una población, $X(\omega) \in R_1$, entonces lo asignaremos a la población e_1 y si $X(\omega) \in R_2$ lo asignaremos a la población e_2 . La probabilidad de mala clasificación será α . Cuanto menor α , más distintos serán las densidades de probabilidad, y por tanto $1-\alpha$ puede ser tomado como una distancia entre las dos poblaciones.

Puede verificarse que es no negativa y verifica la desigualdad triangular. La distancia de Mahalanobis puede expresarse como una función monótona creciente de $1-\alpha$.

Citemos finalmente una distancia aplicada a la genética y desarrollada por Nei (1972), basándose en teoría evolutiva neutralista. Se define la distancia entre dos poblaciones biológicas (y estadísticas a la vez) como:

$$D_{12} = -\ln I_{12}$$

donde $I_{12} = J_{12} / \sqrt{J_1 J_2}$, siendo J_1, J_2 y J_{12} las medias aritméticas, extendidas a todos los locus, de: J_{j1}, J_{j2} , probabilidad de que al extraer al azar dos alelos de la misma población éstos sean iguales, para el locus j , y J_{j12} la misma probabilidad para un alelo extraído de la pobla-

ción 1 y el otro de la población 2. La distancia así obtenida no es una distancia métrica.

1.4. DISCUSION

Hemos visto pues la notable variedad de distancias y disimilaridades existentes de las que hemos citado sólo una parte. Desde un punto de vista estrictamente matemático, el problema de la introducción de una distancia en un conjunto es un problema indeterminado, puesto que es posible desarrollar muchas geometrías todas ellas correctas desde un punto de vista lógico-formal, como quedó claro después de los trabajos de Lobachevski y Bolyai, Sokolnikoff (1971). El problema que se plantea el investigador es ¿que distancia elegir? ya que los resultados, por ejemplo las representaciones gráficas, dependeran de la distancia elegida. Recordemos que estos métodos estan pensados para ayudar a encontrar relaciones entre objetos que pueden no aparecer claras entre la frecuentemente abrumadora cantidad de información experimental: muchos objetos a estudiar y muchas variables. La posibilidad de elegir para cada caso particular una distancia u otra, implica el riesgo de que ésta se escoja de forma que los resultados sean, de una u otra forma, mas convenientes para el investigador, en el sentido de encajar mejor con las ideas que se suelen tener preconcebidas, o aunque no fuese así, elegir aquella que permite en cada situación sacar más conclusiones. En el primer caso de alguna forma invalidamos el método, puesto que en esencia no nos aporta nada nuevo. En el segundo, aunque quizá sea defendido por algunos desde un punto de vista pragmático, es un procedimiento estéticamente insatisfactorio y podemos pensar además que dicho proceder conduce a posibles artefactos, que no corresponden a ninguna realidad objetiva.

Mientras que en la descripción del espacio físico, tal como se nos presenta a nuestros sentidos, el problema de elegir una geometría u otra es un problema que puede plantearse a nivel experimental, Einstein (1921), el problema planteado en nuestro caso no es posible resolverlo experimentalmente, puesto que tratamos de introducir una distancia entre objetos sobre los que obtenemos cierta información experimental, resultados de unas variables, que en general carecerán de una significación geométrica inmediata en el universo aparentemente tridimensional que percibimos.

Una posibilidad consiste en el desarrollo de geometrías construidas teniendo en cuenta solamente la naturaleza probabilística de las variables observables, prescindiendo de su naturaleza física concreta. En otras palabras, cada población estadística la caracterizamos por la función de distribución de probabilidad de las variables aleatorias observables y posteriormente definiríamos una distancia entre estas funciones de distribución, prescindiendo siempre de la naturaleza física de las poblaciones estadísticas que distanciamos. Esta postura metodológica estaría basada en último término en la idea que la geometría adecuada no depende solamente de los objetos físicos en sí, sino que depende en gran medida de la información obtenida de los mismos, información que queda reflejada en la función de distribución de las variables aleatorias observables, en este sentido algunos científicos, como H. Poincaré, Rashevsky (1969), han llegado a sugerir que la percepción de la tridimensionalidad del espacio es fruto de nuestra propia naturaleza biológica, que procesa y filtra la información sensorial. Este planteamiento es más bello y elegante desde un punto de vista matemático. Ciertamente los presupuestos filosóficos influyen consciente o inconscientemente en el desarrollo del trabajo de un científico, Waddington (1969).

Sin embargo, este método podría también justificarse a nivel científico-experimental en base a los resultados obtenidos si éstos nos ayudan a una mejor comprensión de la realidad.

Con todo el problema no estará unívocamente determinado, desde el momento que es posible la construcción de varias geométricas asociadas a funciones de densidad de probabilidad, todas ellas correctas desde un punto de vista lógico-formal. La elección de una de ellas no será un problema fácil. Parece razonable elegir la geometría en función de criterios de sencillez, belleza, elegancia y fecundidad matemática.

En esta memoria vamos a centrarnos en el estudio de distancias entre funciones de densidad paramétricas. Cada población estadística vendrá caracterizada como un punto de un espacio paramétrico. En este espacio paramétrico definiremos un campo tensorial covariante de segundo orden y simétrico, que nos permitirá definir una métrica en el mismo. Haremos pues, uso de la geometría desarrollada por Riemann, como una generalización de la geometría intrínseca de superficies en el espacio que desarrolló Gauss. Para ello necesitamos, como herramientas de trabajo algunos elementos de análisis tensorial desarrollados por el propio Riemann, Christoffel, Ricci, Levi-Civita, etc., así como elementos de cálculo variacional y geometría diferencial. El siguiente capítulo proporciona una revisión de los conceptos más importantes de estas disciplinas, para una correcta comprensión del tema.

2. ALGUNOS ASPECTOS DEL ANALISIS TENSORIAL Y GEOMETRIA DIFERENCIAL

Resumen:

En el presente capítulo se exponen, a grandes rasgos y sin ánimo de hacer una exposición completa del tema, algunas de las técnicas matemáticas que se utilizan a lo largo de los siguientes capítulos. En este sentido debe considerarse como un capítulo metodológico, que proporciona las herramientas de trabajo necesarias para el desarrollo del tema.

Sumario:

- 2.1. VARIEDAD. TRANSFORMACION ADMISIBLE.
- 2.2. TENSORES.
- 2.3. ALGEBRA TENSORIAL.
- 2.4. GEOMETRIA EN UNA VARIEDAD.
- 2.5. SIMBOLOS DE CHRISTOFFEL. DERIVADA COVARIANTE.
- 2.6. CARACTERIZACION DE LA EUCLIDIANIDAD. CURVATURA RIEMANNIANA.
- 2.7. CALCULO DE GEODESICAS.

2.1. VARIEDAD. TRANSFORMACION ADMISIBLE

Este capítulo se inicia con el concepto de variedad. Un conjunto M , $M \subset \mathbb{R}^n$, diremos que es una variedad k -dimensional, en \mathbb{R}^n , si y sólo si para todo punto $x \in \mathbb{R}^n$ existe un conjunto abierto U , $U \subset \mathbb{R}^n$, que contiene a x , un conjunto abierto V , $V \subset \mathbb{R}^n$, y un difeomorfismo (función diferenciable con inversa diferenciable) $h: U \rightarrow V$ tal que:

$$h(U \cap M) = V \cap (\mathbb{R}^k \times \{0\}) = \{x \in V / x^{k+1} = \dots = x^n = 0\}$$

Existe un teorema de caracterización de las variedades que resulta de suma importancia:

Un subconjunto $M \subset \mathbb{R}^n$ es una variedad k -dimensional si y sólo si para cada punto $x \in M$ existe un conjunto abierto U , $U \subset \mathbb{R}^n$, que contiene a x , un conjunto abierto $W \subset \mathbb{R}^k$ y una función inyectiva diferenciable $f: W \rightarrow \mathbb{R}^n$ tal que $f(W) = M \cap U$ y $f'(y)$ (matriz jacobiana de f en y) tiene rango k para cada $y \in W$. A la función f se la denomina sistema de coordenadas.

La anterior formalización nos garantiza que a cada punto de la variedad k -dimensional M , respecto de un sistema de coordenadas f , le podemos asociar k números, x^1, \dots, x^k , que serán las coordenadas de dicho punto bajo este sistema de coordenadas. Nótese que si cambiamos de sistema de coordenadas, caracterizaremos al mismo punto de M por otros números y^1, \dots, y^k . Propiedades geométricas tales como la distancia entre dos puntos de la variedad, independientemente de como la definamos, no deben depender del sistema de coordenadas, en otras palabras la distan-

cia entre dos puntos debe permanecer invariante frente a cualquier transformación de coordenadas. Más información del concepto de variedad puede ser hallada en Hicks (1974), Spivak (1970). Veamos a continuación la clase de transformaciones de coordenadas que vamos a considerar.

Una transformación de coordenadas en \mathbb{R}^k es una aplicación T de \mathbb{R}^k en si mismo, $T: \mathbb{R}^k \rightarrow \mathbb{R}^k$. Vamos a imponer la condición que T admita inversa, T^{-1} . Podremos simbolizar T por el conjunto de k funciones:

$$y^i = y^i(x^1, \dots, x^k) \quad i=1, \dots, k$$

y T^{-1} por:

$$x^i = x^i(y^1, \dots, y^k) \quad i=1, \dots, k$$

Exigiremos además que la transformación T sea diferenciable con continuidad, es decir, las derivadas parciales $\partial y^j / \partial x^i$ sean continuas. En estas condiciones el rango de la matriz jacobiana es k en todo su dominio ya que el determinante jacobiano:

$$J = \left| \frac{\partial y^i}{\partial x^j} \right| \neq 0$$

A esta clase de transformaciones (de coordenadas) las denominaremos transformaciones admisibles. Nótese que una transformación de coordenadas con estas características es localmente lineal.

El conjunto de las transformaciones admisibles de coordenadas, en \mathbb{R}^k , con la ley de composición de aplicaciones, tiene estructura de grupo:

la operación es cerrada puesto que la composición de dos transformaciones admisibles es una transformación admisible, la ley de composición de aplicaciones cumple con la propiedad asociativa, la identidad es una transformación admisible y es el elemento neutro de dicha operación y cada transformación admisible admite una transformación inversa admisible. Dicho grupo es obviamente no conmutativo.

El hecho que las transformaciones admisibles tengan estructura de grupo nos justifica el tomar como punto de partida un sistema de coordenadas adecuado, puesto que cada sistema de coordenadas puede obtenerse de uno particular mediante una conveniente transformación admisible.

2.2. TENSORES

A continuación introduciremos el concepto de tensor, en su sentido más amplio, debido a Weyl y Veblen, Sokolnikoff (1971). Consideremos una transformación admisible de coordenadas:

$$T: y^i = y^i(x^1, \dots, x^n) \quad i=1, \dots, n$$

y un conjunto f_i de m funciones continuas:

$$f_i(x^1, \dots, x^n) \quad i=1, \dots, m$$

definidas en una variedad n -dimensional. La transformación T induce una

transformación G en el conjunto de las $f_i(x^1, \dots, x^n)$, transformandolas en $g_i(y^1, \dots, y^n)$.

Si el conjunto de las transformaciones inducidas, con la ley de composición de aplicaciones, es isomorfo al grupo de transformaciones admisibles, diremos que el conjunto de las f_i representan las componentes de un tensor A en el sistema de coordenadas $x, (x^1, \dots, x^n)$, siendo el tensor A la totalidad de conjuntos de funciones f_i, g_i , etc., relacionados entre sí por una transformación G conveniente, inducida por T . Recordemos que para que el grupo de transformaciones inducidas sea isomorfo al grupo de transformaciones admisibles, debe verificarse que si T_1 y T_2 son transformaciones admisibles, y G_1 y G_2 las transformaciones inducidas por ambas, entonces la transformación inducida por $T_2 \circ T_1$ es $G_2 \circ G_1$.

Nos centraremos a continuación en el estudio de los invariantes, tensores covariantes y tensores contravariantes.

Sea ϕ una función que a cada punto P de una variedad n -dimensional le asigna un escalar $\phi(P)$. El valor $\phi(P)$ depende únicamente de P , no del sistema de coordenadas con el que trabajemos. Llamaremos a ϕ función escalar, campo escalar o simplemente escalar.

Dado un sistema concreto de coordenadas, $\phi(P)$ tomara una forma concreta:

$$f(x^1, \dots, x^n)$$

y si introducimos un nuevo sistema de coordenadas mediante la transformación admisible:

$$T: \quad x^i = x^i(y^1, \dots, y^n) \quad i=1, \dots, n$$

la forma funcional de $\phi(P)$ en el nuevo sistema de referencia es:

$$f(x^1(y^1, \dots, y^n), \dots, x^n(y^1, \dots, y^n)) \equiv g(y^1, \dots, y^n) \quad (1)$$

puesto que ya hemos dicho que el escalar que asignamos a cada punto no depende del sistema de coordenadas.

Podemos decir que $f(x^1, \dots, x^n)$ es la componente de la función escalar $\phi(P)$ en el sistema de coordenadas X , mientras que $g(y^1, \dots, y^n)$ es la componente de la misma función escalar en el sistema de coordenadas Y . La transformación que ha inducido T , definida por (1), la denominaremos transformación por invariancia:

$$G^0: f(x(y)) = g(y) \quad (2)$$

y es fácil comprobar que el conjunto de transformaciones inducidas, con la ley de composición de aplicaciones, es isomorfo al grupo de transformaciones admisibles, por tanto ϕ es un tensor, que denominaremos tensor de orden 0, ó invariante.

Veamos a continuación como se transforman las derivadas parciales de una función escalar, al efectuar una transformación de coordenadas. Sea ϕ una función escalar, que bajo un sistema de coordenadas X toma la forma funcional:

$$f(x^1, \dots, x^n)$$

y sea T una transformación admisible $T: x^i = x^i(y^1, \dots, y^n) \quad i=1, \dots, n$.

Aplicando la regla de la cadena, a partir de (2) resultará:

$$G^1: \frac{\partial g}{\partial y^i} = \sum_{h=1}^n \frac{\partial f}{\partial x^h} \frac{\partial x^h}{\partial y^i} \quad (3)$$

y usando el convenio de sumación de los índices repetidos, para el resto del capítulo, queda:

$$G^1: \frac{\partial g}{\partial y^i} = \frac{\partial f}{\partial x^h} \frac{\partial x^h}{\partial y^i} \quad (4)$$

a la transformación inducida por T , G^1 , definida en (3), la denominaremos transformación por covariancia y puede demostrarse que el conjunto de dichas transformaciones inducidas en el conjunto de funciones de la forma $\frac{\partial f}{\partial x^1}, \dots, \frac{\partial f}{\partial x^n}$, con la ley de composición de aplicaciones, es isomorfo al grupo de transformaciones admisibles. Las derivadas parciales de una función escalar son pues las componentes de un tensor.

Generalizando, a un conjunto de n funciones definidas en una variedad n -dimensional, $A_1(x^1, \dots, x^n), \dots, A_n(x^1, \dots, x^n)$ que al efectuar una transformación de coordenadas:

$$x^i = x^i(\bar{x}^1, \dots, \bar{x}^n) \quad i=1, \dots, n \quad (5)$$

se transforma en $\bar{A}_1(\bar{x}^1, \dots, \bar{x}^n), \dots, \bar{A}_n(\bar{x}^1, \dots, \bar{x}^n)$ de acuerdo con:

$$\bar{A}_\mu = \frac{\partial x^v}{\partial \bar{x}^\mu} A_\nu \quad \mu=1, \dots, n \quad (6)$$

diremos que representan las componentes de un tensor covariante de primer orden.

En general, a un conjunto de n^r funciones definidas en una variedad n -dimensional, $A_{\mu_1 \dots \mu_r}(x^1, \dots, x^n)$ $\mu_1, \dots, \mu_r=1, \dots, n$, si al efectuar la transformación (5) se transforman en $\bar{A}_{\mu_1 \dots \mu_r}(\bar{x}^1, \dots, \bar{x}^n)$ de acuerdo con:

$$\bar{A}_{\mu_1 \dots \mu_r} = \frac{\partial x^{v_1}}{\partial \bar{x}^{\mu_1}} \dots \frac{\partial x^{v_r}}{\partial \bar{x}^{\mu_r}} A_{v_1 \dots v_r} \quad (7)$$

diremos que representan las componentes de un tensor covariante de orden r .

Veamos a continuación como se transforman el conjunto de formas diferenciales dx^1, \dots, dx^n , al efectuar la transformación (5).

$$G^2: \quad d\bar{x}^\beta = \frac{\partial \bar{x}^\beta}{\partial x^\alpha} dx^\alpha \quad (8)$$



a la transformación G^2 definida en (8) la denominaremos transformación por contravariancia. Además, el conjunto de dichas transformaciones inducidas en el conjunto de las dx^1, \dots, dx^n , con la ley de composición de aplicaciones es isomorfo al grupo de transformaciones admisibles: las formas diferenciales dx^1, \dots, dx^n son las componentes de un tensor.

Generalizando, a un conjunto de n funciones definidas en una variedad n -dimensional, $A^1(x^1, \dots, x^n), \dots, A^n(x^1, \dots, x^n)$, que al efectuar transformaciones (5) se convierten en $\bar{A}^1(\bar{x}^1, \dots, \bar{x}^n), \dots, \bar{A}^n(\bar{x}^1, \dots, \bar{x}^n)$ de acuerdo con:

$$\bar{A}^\mu = \frac{\partial \bar{x}^\mu}{\partial x^\nu} A^\nu \quad (9)$$

diremos que representan las componentes de un tensor contravariante de primer orden. Nótese en el uso de superíndices para distinguir el caso contravariante del covariante.

En general, a un conjunto de n^r funciones $A^{\mu_1 \dots \mu_r}(x^1, \dots, x^n)$ $\mu_1, \dots, \mu_r = 1, \dots, n$ definidas en una variedad n -dimensional, si al efectuar la transformación (5) se convierten en $\bar{A}^{\mu_1 \dots \mu_r}(\bar{x}^1, \dots, \bar{x}^n)$ de acuerdo con:

$$\bar{A}^{\mu_1 \dots \mu_r} = \frac{\partial \bar{x}^{\mu_1}}{\partial x^{\nu_1}} \dots \frac{\partial \bar{x}^{\mu_r}}{\partial x^{\nu_r}} A^{\nu_1 \dots \nu_r} \quad (10)$$

diremos que representan las componentes de un tensor contravariante de orden r .

También podremos hablar de tensores mixtos con algunos índices covariantes y otros índices contravariantes. El conjunto de n^{p+q} funciones definidas sobre una variedad n -dimensional, $A_{\lambda_1 \dots \lambda_q}^{\mu_1 \dots \mu_p}(x^1, \dots, x^n)$, $\mu_1, \dots, \mu_p, \lambda_1, \dots, \lambda_q = 1, \dots, n$, si al efectuar la transformación (5) se transforman en $\bar{A}_{\lambda_1 \dots \lambda_q}^{\bar{\mu}_1 \dots \bar{\mu}_p}(\bar{x}^1, \dots, \bar{x}^n)$, de acuerdo con:

$$\bar{A}_{\lambda_1 \dots \lambda_q}^{\bar{\mu}_1 \dots \bar{\mu}_p} = \frac{\partial \bar{x}^{\bar{\mu}_1}}{\partial x^{\alpha_1}} \dots \frac{\partial \bar{x}^{\bar{\mu}_p}}{\partial x^{\alpha_p}} \cdot \frac{\partial x^{\beta_1}}{\partial \bar{x}^{\lambda_1}} \dots \frac{\partial x^{\beta_q}}{\partial \bar{x}^{\lambda_q}} A_{\beta_1 \dots \beta_q}^{\alpha_1 \dots \alpha_p} \quad (11)$$

diremos que representan las componentes de un tensor mixto de orden $p+q$, p veces contravariante y q veces covariante.

Los tensores covariantes, contravariantes y mixtos, pueden ser introducidos también como campos de formas multilineales definidos en la variedad, es decir, funciones que asignan a cada punto de la variedad una forma multilineal definida en el espacio tangente a dicho punto, o a su dual, según se trate de tensores covariantes o contravariantes, Hicks (1974).

2.3. ALGEBRA TENSORIAL

Visto el concepto de tensor, pasemos a revisar las propiedades básicas del Algebra tensorial. Es posible definir la suma de tensores que tengan el mismo número de índices covariantes y contravariantes:

si $A_{\lambda_1 \dots \lambda_q}^{\mu_1 \dots \mu_p}$ y $B_{\lambda_1 \dots \lambda_q}^{\mu_1 \dots \mu_p}$ son las componentes de dos tensores mixtos de orden $p+q$, $C_{\lambda_1 \dots \lambda_q}^{\mu_1 \dots \mu_p}$ definido por:

$$C_{\lambda_1 \dots \lambda_q}^{\mu_1 \dots \mu_p} = A_{\lambda_1 \dots \lambda_q}^{\mu_1 \dots \mu_p} + B_{\lambda_1 \dots \lambda_q}^{\mu_1 \dots \mu_p} \quad (12)$$

$$\mu_1, \dots, \mu_p, \lambda_1, \dots, \lambda_q = 1, \dots, n$$

son las componentes de otro tensor mixto de orden $p+q$, con el mismo número de índices contravariantes p , y covariantes q . Esta operación con fiere al conjunto de los tensores mixtos, con p índices contravariantes y q índices covariantes, la estructura de grupo conmutativo.

También es posible definir una operación externa, escalar por tensor, de forma que si $A_{\lambda_1 \dots \lambda_q}^{\mu_1 \dots \mu_p}$ son las componentes de un tensor mixto de orden $p+q$, con p índices contravariantes y q covariantes, y α es un escalar, entonces $B_{\lambda_1 \dots \lambda_q}^{\mu_1 \dots \mu_p}$ definido por:

$$B_{\lambda_1 \dots \lambda_q}^{\mu_1 \dots \mu_p} = \alpha A_{\lambda_1 \dots \lambda_q}^{\mu_1 \dots \mu_p} \quad (13)$$

$$\mu_1 \dots \mu_p, \lambda_1 \dots \lambda_q = 1, \dots, n$$

son las componentes de un tensor de orden $p+q$, con el mismo número de índices contravariantes p y covariantes q , que el anterior, como es fácil comprobar.

Con las dos operaciones citadas, el conjunto de tensores mixtos de orden $p+q$, con p índices contravariantes y q covariantes, tiene estructura de espacio vectorial.

Hay una operación que conecta dichos espacios vectoriales entre sí.