MATEMÁTICAS

25 de noviembre de 2024

Índice general

Parte I

Análisis matemático

Capítulo 1

Límites y continuidad

1.1. Introducción

Supongamos al lector ya familiarizado con el concepto de límites tal como es introducido en el cálculo elemental donde es corriente presentar varios tipos de límites. Por ejemplo, el *límite de una sucesión* de números reales $\{x_n\}$, que simbolizamos cuando escribimos

$$\lim_{n\to\infty} x_n = A$$

significa que para cada número $\varepsilon > 0$ existe un entero $N > 0$ tal que

$$|x_n - A| < \varepsilon \quad \text{siempre que } n \geq N$$

Este límite pretende transmitir la idea intuitiva de que x_n puede estar suficientemente próximo a A en el supuesto de que n sea suficientemente grande. También se da el *límite de una función*, indicado por medio de la notación

$$\lim_{x\to p} f(x) = A$$

que significa que para cada $\varepsilon > 0$ existe otro número $\delta > 0$ tal que

$$|f(x) - A| < \varepsilon \quad \text{siempre que } 0 < |x - p| < \delta$$

Esta definición expresa la idea de que $f(x)$ puede conseguirse tan próximo a A como queramos, siempre que x se tome lo suficientemente próximo a p, sin llegar a ser p.

Las aplicaciones del cálculo a los problemas geométricos y físicos del espacio tridimensional y a las funciones de varias variables nos obligan a extender estos conceptos a $\mathbb{R}^n$. Es tan necesario como fácil dar un paso más e introducir los límites en el marco más general de los espacios métricos. Esto simplifica la teoría puesto que elimina restricciones innecesarias y al mismo tiempo cubre casi todos los aspectos necesarios del análisis.

Primeramente discutiremos los límites de las sucesiones de puntos de un espacio métrico y después discutiremos los límites de funciones y el concepto de continuidad.

1.2. Sucesiones convergentes en un espacio métrico

Definición 1.1. Una sucesión $\{x_n\}$ de puntos de un espacio métrico (S, d) es convergente si existe un punto p de S que satisfaga la siguiente propiedad:

Para todo $\varepsilon > 0$ existe un entero N tal que

$$d(x_n, p) < \varepsilon \quad \text{siempre que } n \geq N$$

Diremos también que $\{x_n\}$ converge hacia p y escribiremos $x_n \to p$ cuando $n \to \infty$, o simplemente $x_n \to p$. Si no existe un tal número p de S, se dice que la sucesión $\{x_n\}$ es divergente.

La definición de convergencia implica que

$$x_n \to p \quad \text{si, y sólo si, } d(x_n, p) \to 0$$

La convergencia de la sucesión $\{d(x_n, p)\}$ hacia 0 se realiza en el espacio euclídeo $\mathbb{R}$.

Ejemplo 1.1. En el espacio euclídeo $\mathbb{R}$, una sucesión $\{x_n\}$ se llama *creciente* si $x_n \leq x_{n+1}$ para todo n. Si una sucesión creciente está acotada superiormente (esto es, si $x_n \leq M$ para un $M > 0$ y para todo n), entonces $\{x_n\}$ converge hacia el supremo de su recorrido, $\sup\{x_1, x_2, \dots\}$. Análogamente, $\{x_n\}$ se llama *decreciente* si $x_{n+1} \leq x_n$ para todo n. Cada sucesión decreciente acotada inferiormente converge hacia el ínfimo de su recorrido. Por ejemplo, $\{1/n\}$ converge a 0.

Ejemplo 1.2. Si $\{a_n\}$ y $\{b_n\}$ son sucesiones reales que convergen hacia 0, entonces $\{a_n + b_n\}$ también converge hacia 0. Si $0 \leq c_n \leq a_n$ para todo n y si $\{a_n\}$ converge hacia 0, entonces $\{c_n\}$ también converge hacia 0. Estas propiedades elementales de las sucesiones $\mathbb{R}$ pueden ser útiles para simplificar algunas de las demostraciones concernientes a límites de un espacio métrico general.

Ejemplo 1.3. En el plano complejo $\mathbb{C}$, sea

$$z_n = 1 + \frac{1}{n^2} + \left(2 - \frac{1}{n}\right) i$$

Entonces $\{z_n\}$ converge hacia $1 + 2i$ puesto que

$$d(z_n, 1+2i)^2 = |z_n - (1+2i)|^2 = \left|\frac{1}{n^2} - \frac{1}{n} i\right|^2 = \frac{1}{n^4} + \frac{1}{n^2} \to 0 \text{ cuando } n \to \infty$$

luego $d(z_n, 1+2i) \to 0$.

Teorema 1.1. *Una sucesión* $\{x_n\}$ *en un espacio métrico* (S, d) *puede converger hace un punto de* S*, a lo sumo.*

Demostración. Supongamos que $x_n \to p$ y que $x_n \to q$. Probaremos que $p = q$. En virtud de la desigualdad triangular se tiene

$$0 \leq d(p, q) \leq d(p, x_n) + d(x_n, q)$$

Como $d(p, x_n) \to 0$ y $d(x_n, q) \to 0$ se tiene que $d(p, q) = 0$, luego $p = q$. $\square$

Si una sucesión $\{x_n\}$ converge, el único punto hacia el que converge se llama *límite* de la sucesión y se designa por medio de $\text{lím}\, x_n$ o por medio de $\text{lím}_{n\to\infty} x_n$.

Parte II

Cálculo vectorial

Capítulo 2

Transformaciones lineales y matrices

2.1. Representación matricial de las transformaciones lineales

Ya que es posible obtener distintas representaciones matriciales de una transformación lineal dada mediante la elección de bases distintas, parece natural intentar elegir bases de modo que la matriz resultante tenga una forma lo más sencilla posible. El teorema que sigue prueba que podemos hacer todos los elementos 0 excepto los de la diagonal que va desde el vértice superior izquierdo al inferior derecho. A lo largo de esa diagonal habrá una hilera de unos seguidos de ceros, siendo el número de unos igual al rango de la transformación. Una matriz (t_{ik}) con todos los elementos $t_{ik} = 0$ cuando $i \neq k$ se llama **matriz diagonal**.

Teorema 2.1. *Sean V y W espacios lineales de dimensión finita, con $\dim V = n$ y $\dim W = m$. Supongamos que $T \in \mathscr{L}(V,W)$ y que $r = \dim T(V)$ representa el rango de T. Existen entonces una base $(e_1, \ldots, e_n)$ para V y otra $(w_1, \ldots, w_m)$ para W tales que*

$$T(e_i) = w_i \quad \text{para } i = 1, 2, \ldots, r \tag{2.1}$$

y

$$T(e_i) = 0 \quad \text{para } i = r+1, \ldots, n \tag{2.2}$$

Por consiguiente, la matriz (t_{ik}) de T relativa a esas bases tiene todos los elementos cero excepto los r elementos de la diagonal que valen

$$t_{11} = t_{22} = \cdots = t_{rr} = 1$$

Parte III

Análisis de datos multivariantes

Capítulo 3

Álgebra matricial

3.1. Definiciones básicas

Un conjunto de n números reales $\mathbf{x}$ puede representarse como un punto en el espacio de n dimensiones $\mathbb{R}^n$. Definiremos el vector $\mathbf{x}$ como el segmento orientado que une el origen de coordenadas con el punto $\mathbf{x}$. La orientación es importante, porque no es lo mismo el vector $\mathbf{x}$ que el $-\mathbf{x}$. Con esta correspondencia, a cada punto del espacio en $\mathbb{R}^n$ le asociamos un vector. En adelante, representaremos un vector mediante $\mathbf{x}$, para diferenciarlo del escalar x, y llamaremos $\mathbb{R}^n$ al espacio de todos los vectores de n coordenadas o componentes. En particular, un conjunto de números con todos los valores iguales se representará por un vector **constante**, que es aquél con todas sus coordenadas iguales. Un vector constante es de la forma $c\mathbf{1}$, donde c es cualquier constante y $\mathbf{1}$ es el vector con todas sus coordenadas iguales a la unidad.

En Estadística podemos asociar a los valores de una variable en n elementos un vector en $\mathbb{R}^n$, cuyo componente i-ésimo es el valor de la variable en el elemento i. Por ejemplo, si medimos las edades de tres personas en una clase y obtenemos los valores 20, 19 y 21 años, esta muestra se representa por el vector tridimensional

$$\mathbf{x} = \begin{pmatrix} 20 \\ 19 \\ 21 \end{pmatrix}$$

La **suma** (o **diferencia**) de dos vectores $\mathbf{x}$ e $\mathbf{y}$, ambos en $\mathbb{R}^n$, se define como un nuevo vector con componentes iguales a la suma (diferencia) de los componentes de los sumandos:

$$\mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

Es inmediato comprobar que la suma de vectores es asociativa, $\mathbf{x}+(\mathbf{y}+\mathbf{z}) = (\mathbf{x}+\mathbf{y})+\mathbf{z}$

y conmutativa, $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.

La suma de dos vectores corresponde a la idea intuitva de trasladar un vector al extremo del otro y construir la línea que va desde el origen del primero al extremo del segundo. La operación suma (resta) de dos vectores da lugar a otro vector y estadísticamente corresponde a generar una nueva variable como suma (resta) de otras dos anteriores. Por ejemplo, si $\mathbf{x}$ representa el número de trabajadores varones en un conjunto de empresas e $\mathbf{y}$ el número de trabajadoras, la variable $\mathbf{x} + \mathbf{y}$ representa el número total de trabajadores y la variable $\mathbf{x} - \mathbf{y}$ la diferencia entre hombres y mujeres de cada empresa.

El producto de una constante por un vector, es un nuevo vector con componentes los del vector inicial multiplicados por la constante.

$$\mathbf{z} = k\mathbf{x} = \begin{pmatrix} kx_1 \\ \vdots \\ kx_n \end{pmatrix}$$

Multiplicar por una constante equivale a un cambio en las unidades de medición. Por ejemplo, si en lugar de medir el número de trabajadores en unidades (variable $\mathbf{x}$) lo hacemos en centenas (variable $\mathbf{z}$), entonces la variable $\mathbf{z}$ es igual a $\mathbf{x}/100$.

Llamaremos **vector transpuesto $\mathbf{x}'$**, de otro $\mathbf{x}$, a un vector con las mismas componentes, pero escritas ahora en fila:

$$\mathbf{x}' = (x_1, \ldots, x_n)$$

Al transponer un vector columna se obtiene un vector fila. Generalmente los vectores fila se utilizan para describir los valores de p variables distintas en un mismo elemento de una población.

El **producto escalar o interno** de dos vectores $\mathbf{x}$, $\mathbf{y}$, ambos en $\mathbb{R}^n$, que escribiremos $\mathbf{x}'\mathbf{y}$ o $\mathbf{y}'\mathbf{x}$, es el escalar obtenido al sumar los productos de sus componentes.

$$\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x} = \sum_{i=1}^{n} x_i y_i$$

Se llamará **norma (cuadrática)** o longitud de un vector $\mathbf{x}$, a la raíz cuadrada positiva del producto escalar $\mathbf{x}'\mathbf{x}$. Se escribe $\|\mathbf{x}\|$:

$$\|x\| = \sqrt{\mathbf{x}'\mathbf{x}} = \sqrt{x_1^2 + \cdots + x_n^2}$$

La norma es la longitud del segmento que une el origen con el punto $\mathbf{x}$.

El producto escalar puede calcularse también como el producto de las normas de los vectores por el coseno del ángulo que forman. Para ilustrar este concepto consideremos los vectores

$$\mathbf{x} = \begin{pmatrix} a \\ 0 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} a \\ c \end{pmatrix}$$

Observemos que el producto escalar es $\mathbf{x}'\mathbf{y} = a^2$ y que este mismo resultado se obtiene multiplicando la norma de ambos vectores, $\|\mathbf{x}\| = a$, $\|\mathbf{y}\| = \sqrt{a^2 + c^2}$ por el coseno del ángulo θ que forman, dado por $\cos\theta = a/\sqrt{a^2 + c^2}$. Observemos que el producto escalar puede también expresarse como el producto de la norma de un vector por la proyección del otro sobre él. Si uno de los vectores tiene norma uno, el producto escalar es directamente la proyección del otro vector sobre él.

Se demuestra en general que:

$$|\mathbf{x}'\mathbf{y}| \leq \|\mathbf{x}\| \, \|\mathbf{y}\|$$

que se conoce como la **desigualdad de Cauchy-Schwarz**. Esta desigualdad permite definir el **ángulo** entre dos vecotres $\mathbf{x}$ e $\mathbf{y}$ de $\mathbb{R}^n$ (no nulos) por la relación:

$$\cos\theta = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{x}\| \, \|\mathbf{y}\|}$$

Si dos variables tienen media cero, el coseno del ángulo que forman es su **coeficiente de correlación**.

Dos vectores son **ortogonales**, o perpendiculares, si y sólo si, su producto escalar es cero. Por la definición de ángulo

$$\mathbf{x}'\mathbf{y} = \|\mathbf{x}\| \, \|\mathbf{y}\| \cos\theta$$

siendo θ el ángulo que forman los vectores. Si $\theta = 90°$ el coseno es cero y también lo será su producto escalar.

El producto escalar tiene una clara interpretación estadística. Para describir una variable tomamos su media. Para describir un vector podemos tomar su proyección sobre el vector constante. El vector constante de norma unidad en dimensión n es $\frac{1}{\sqrt{n}}\mathbf{1}$, y la proyección de $\mathbf{x}$ sobre este vector es $\frac{1}{\sqrt{n}}\mathbf{1}'\mathbf{x} = \sum x_i/\sqrt{n} = \bar{x}\sqrt{n}$. El vector constante resultante de esta proyección es

$$\bar{x}\sqrt{n}\left(\frac{1}{\sqrt{n}}\right)\mathbf{1} = \bar{x}\mathbf{1}$$

Por tanto, la media es el escalar que define el vector obtenido al proyectar el vector de datos sobre la dirección constante. También puede interpretarse como la norma estandarizada del vector obtenido al proyectar los datos en la dirección del vector constante, donde para estandarizar la norma de un vector dividiremos siempre por $\sqrt{n}$, siendo n la dimensión del espacio.

La variabilidad de los datos se mide por la **desviación típica**, que es la distancia estandarizada entre el vector de datos y el vector constante. La proyección del vector de datos sobre la dirección del vector constante produce el vector $\bar{x}\mathbf{1}$, y la norma del vector diferencia, $\mathbf{x} - \bar{x}\mathbf{1}$, mide la distancia entre el vector de datos y el vector constante. La norma estandarizada, dividiendo por la raíz de la dimensión del espacio es:

$$\frac{1}{\sqrt{n}}\,\|x - \bar{x}\mathbf{1}\| = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} = \sigma_{\mathbf{x}}$$

La medida de dependencia lineal entre dos variables, $\mathbf{x}$, $\mathbf{y}$ es la **covarianza**. La covarianza es el producto escalar promedio de los dos vectores medidos en desviaciones a la media, o tomando sus diferencias respecto a la proyección sobre el vector constante. Si promediamos el producto escalar de estos vectores

$$\frac{1}{n}(\mathbf{x} - \bar{x}\mathbf{1})'(\mathbf{y} - \bar{y}) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n} = \mathrm{Cov}(\mathbf{x}, \mathbf{y})$$

se obtiene la covarianza. Para variables con media cero, el producto escalar de los dos vectores, dividido por n, es directamente la covarianza.

Para variables estandarizadas (de media cero y desviación típica uno) la covarianza es el **coeficiente de correlación**:

$$\rho = \frac{\mathrm{Cov}(\mathbf{x}, \mathbf{y})}{\sigma_{\mathbf{x}}\sigma_{\mathbf{y}}} = \mathrm{Cov}(\mathbf{x}, \mathbf{y})$$

Para vectores media cero, el coeficiente de correlación es el coseno del ángulo entre los vectores que las representan:

$$\rho = \frac{\mathrm{Cov}(\mathbf{x}, \mathbf{y})}{\sigma_{\mathbf{x}}\sigma_{y}} = \frac{\frac{\mathbf{x}'\mathbf{y}}{n}}{\frac{\|x\|}{\sqrt{n}}\frac{\|y\|}{\sqrt{n}}} = \frac{\mathbf{x}'\mathbf{y}}{\|x\|\,\|y\|} = \cos\theta$$

que es la interpretación geométrica del coeficiente de correlación. La implicación estadística de la ortogonalidad es la incorrelación. Si dos variables tienen media cero y son ortogonales, es decir, los vectores que las caracterizan forman un ángulo de 90 grados, $\rho = 0$, las variables están incorreladas.

Bibliografía

[Apostol, 1985] Apostol, T. M. (1985). *Calculus. Volumen 2, segunda edición.* Editorial Reverté S.A.

[Peña, 2002] Peña, D. (2002). *Análisis de Datos Multivariantes.* Editorial McGrawHill.