El Espacio de Hilbert y la Teoría general del ajuste

Por **D. Angel Vegas Pérez,**
Catedrático en la Escuela Central de Altos Estudios Mercantiles.—Profesor de Matemáticas para Economistas en la Facultad de Ciencias Económicas.

A) Representaciones analíticas en el espacio de Hilbert.

Las relaciones fundamentales que produce el cálculo vectorial en el Espacio Euclídeo, aplicadas a determinada clase de funciones, otorga el concepto de Espacio de Hilbert.

Sabemos que el producto escalar de dos vectores $\vec{a}$ y $\vec{b}$ cuyas componentes cartesianas son, respectivamente, x_a, y_a, z_a y x_b, y_b, z_b, tiene por representación analítica

$$\vec{a} \cdot \vec{b} = x_a\, x_b + y_a \cdot y_b + z_a \cdot z_b$$

de donde la condición necesaria y suficiente para que los vectores $\vec{a}$ y $\vec{b}$ sean ortogonales, es:

$$x_a\, x_b + y_a\, y_b + z_a \cdot z_b = 0$$

Por otra parte, sabemos que en un espacio de n dimensiones solamente puede haber n vectores independientes, pues cualquier otro vector será una combinación lineal de aquéllos.

Así, si como clásicamente se hace, representamos por $\vec{I}$, $\vec{J}$, $\vec{K}$, tres vectores ortogonales dos a dos y de valor absoluto igual a uno, es les del promedio aritmético, o valor probable, es la llamada de mínimos cuadrados, es decir, que de todos los valores de una variable casual o de los que determinan su distribución, es el promedio aritmético, el que hace que la suma de los cuadrados de las diferencias que de todos los valores de la variable casual y él, sea menor que si dichas diferencias se calculasen respecto a otro valor distinto de dicho promedio.

De aquí que se pueda decir que la función f(x) admite el desarrollo indicado, como una aproximación probable, es decir, la más probable dentro del sistema elegido. Por eso algunos autores dicen que este desarrollo es una aproximación *en media* de la función f(x).

Se llama espacio de Hilbert al conjunto de funciones ortogonales y normalizadas que tienen la propiedad de que cualquier otra función no puede ser ortogonal a todas las del sistema.

La representación de las funciones en Espacio de Hilbert nos dan para el espacio elegido, es decir, para el sistema de funciones que se haya determinado, el valor más probable de la función, operando en la forma vectorial que hemos visto.

Suponiendo que se trata de un espacio de infinitas dimensiones, tendríamos un desarrollo en serie, que suele llamarse de Fourier.

Desde el punto de vista de la representación analítica de las funciones, tiene gran importancia el desarrollo estudiado, ya que juntamente con los desarrollos de Taylor para las funciones que admiten infinitas *derivadas* y los trigonométricos de Fourier para las funciones con discontinuidad de salto finito (según las condiciones estudiadas por Dirichlet), aparecen estos desarrollos más amplios y de incalculable valor para las funciones empíricas.

Entre las funciones ortogonales más clásicas, están las trigonométricas

$$\frac{1}{\sqrt{2\pi}}, \quad \frac{\cos x}{\sqrt{\pi}}, \quad \frac{\cos 2x}{\sqrt{\pi}} \ldots, \quad \frac{\operatorname{sen} x}{\sqrt{\pi}}, \quad \frac{\operatorname{sen} 2x}{\sqrt{\pi}} \ldots$$

los polinomios de Legendre,

$$P_0(x) = 1 \qquad P_1(x) = x_1, \qquad P_2(x) = \frac{3}{2}x^2 - \frac{1}{2} \ldots$$

$$P_n(x) = \frac{1.3 \ldots (2n-1)}{n!}\left[x^n - \frac{n(n-1)}{2(2n-1)}x^{n-2} + \ldots\right.$$

los polinomios de Hermite

$$H_0(x) = 1 \qquad H_1(x) = 2x \,, \qquad H_2(x) = 4x^2 - 2$$
$$H_n(x) = (2x)^n - \frac{n(n-1)}{1!}(2x)^{n-2} + \ldots$$

B) Aplicaciones a la teoría del ajuste.

Una de las cuestiones más importantes del estudio de los fenómenos estadísticos es el de la representación analítica de las funciones de distribución.

El problema de la representación analítica de las relaciones estadísticas, como ya sabemos, se ha resuelto clásicamente, bien por la elección de funciones que tomasen para los correspondientes valores del argumento o variable los mismos valores que aparecen en la serie estadística que se quiere representar, es decir, que la curva que geométricamente representa a la función elegida pasa por todos los puntos del diagrama estadístico (interpolación), bien eligiendo convenientemente una función cuya curva pase entre los puntos del diagrama, "ajustándose" lo más posible (ajuste).

El ajuste es, por su naturaleza, mejor medio para la representación de las series estadísticas que la interpolación, ya que los datos, por su carácter estadístico, estarán influenciados por errores. El ajuste pretende eliminar los errores y obtener el valor más probable para los de la distribución de la serie, tanto para los consignados en la misma serie, como para los comprendidos entre dos de ellos.

La operación de ajuste tiene, pues, dos momentos: elección del sistema de representación, es decir, de las funciones analíticas, y el de cálculo del valor más probable de los parámetros que figuran en dichas funciones.

El segundo momento ha sido resuelto con especial acierto, merced a la célebre teoría de Gaüs, por el procedimiento de mínimos cuadrados. Posteriormente, por el de momentos de Pearson (que tratándose de polinomios coincide con el de mínimos cuadrados); el de las áreas de Cantelli, etc.

Como ya hemos dicho, el procedimiento de mínimos cuadrados nos da el valor probable para los parámetros.

Fué Tchebicheff el que, para obviar el inconveniente que tenía el no poder aprovechar los cálculos efectuados en un ajuste por mínimos cuadrados cuando se trataba de elegir un polinomio de mayor grado que el del anteriormente utilizado, utilizó un sistema de polinomios ortogonales, que pueden ponerse en la forma siguiente:

$$T_r = \begin{vmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1(r-1)} & f_1 \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2(r-1)} & f_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \gamma_{r1} & \gamma_{r2} & \cdots & \gamma_{r(r-1)} & f_r \end{vmatrix}$$

siendo f(x) la mayoría de las veces de forma potencial y $\gamma_{ij} = \sum_{x=1}^{x=n} f_i \cdot f_j$

Si bien Tchebicheff utilizó para determinación de sus polinomios las fracciones continuas, el uso de los determinantes da un procedimiento de elaboración de los polinomios más sencillo.

Una vez construídos los polinomios, la función de ajuste será de la forma:

$$F(x) = \alpha_1 T_1(x) + \alpha_2 T_2(x) + \cdots$$

Como ya hemos indicado, T_r suele ser un polinomio de grado *r;* de aquí que según se tomen más o menos términos en el desarrollo anterior, obtendremos un polinomio de grado mayor o menor.

Estos polinomios otorgan un espacio de funciones ortogonales y los valores que obtenemos para los parámetros α son los más probables.

A esta ventaja doctrinal se une la de poder aprovechar los cálculos hechos cuando se ha efectuado el ajuste para un polinomio de grado inferior y se quiere aproximar más.

En efecto,

$$\alpha_r = \frac{\sum_{i=1}^{i=n} y_r T_r(x_i)}{\sum_{i=1}^{n} [T_r(x_i)]^2}$$

siendo y_r la frecuencia correspondiente a x_r, lo que indica que para el cálculo de α_r no es necesario calcular $\alpha_1, \alpha_2 \cdots \alpha_{r-1}$; luego sirve todo lo anteriormente efectuado.

Habíamos dicho que el primer momento de la operación de ajuste era la elección del sistema representativo.

Debe ser un criterio puramente racional el que presida tan delicado momento.

Para que la función de ajuste cumpla su fin representativo, debe nacer de un estudio previo del mundo en que tiene realidad el fenómeno estadístico que se estudia. Este es el fundamento de las estadísticas concretas en las ciencias de aplicación. Así, por ejemplo, en Biometría nos encontramos con las funciones de supervivencia de Gompertz y Makeham y las más amplias de Quiquet, que recoge a aquéllas. Las primeras, que nacen de suponer las influencias del envejecimiento y del azar, y las segundas, de las relaciones que unen a los individuos de un grupo, relaciones que en términos de Mecánica supondrían pérdida de grados de libertad, y por ello la ecuación de Quiquet, origen de toda la fecunda teoría, supone que la probabilidad de simultánea supervivencia es función de menor número de variables que de individuos integran el grupo que se supone. Todas estas funciones presentan, por su origen racional, parámetros no abstractos, sino de significación concreta, es decir, de satisfactoria interpretación en el ambiente en que se realiza el fenómeno.

Cuando el fenómeno es de carácter general o no admite un estudio que lo encuadre en un marco definido, entonces ha de estudiarse con los recursos de la Estadística en su aspecto más genérico.

Clásicamente, y merced a la teoría de errores de Gaüs, la función que se elegía con carácter más frecuente para la representación de las distribuciones, era la función normal.

A Pearson se debe el interesantísimo estudio de curvas de frecuencias, clasificadas en siete tipos, nacidos de los diferentes valores de los momentos estadísticos de la distribución. Los diferentes tipos de curvas se obtienen de la ecuación diferencial

$$\frac{dy}{dx} = \frac{y\,(x+\alpha)}{ax^2+bx+c}$$

los parámetros α, a, b y c vendrán dados en función de los momentos de diversos órdenes.

Racionalmente pensando, es natural que aparezca en el ajuste de la función estadística la curva normal de probabilidad, ya que en todo fenómeno de carácter colectivo interviene el azar con mayor o menor influencia.

Sabemos que fenómeno estadístico es aquel en que intervienen una infinitud de causas, por lo que no se pueden discriminar y conocer todas, y *por ello es por lo que* sus regularidades son tendenciales, son probables. El teorema de Laplace, que es fundamental en el Cálculo de Probabilidades, enuncia la interesante proposición de que cuando una desviación es suma de infinitas desviaciones elementales de igual influencia, sigue la ley normal.

La demostración de este teorema por los semi-invariantes de Thiele, otorga una expresión de la forma

$$F(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} G(x)$$

Siendo G(x) una serie en que intervienen las derivadas sucesivas de la función normal. Cuando el número de los sumandos desviaciones es infinito, G(x) tiende a la unidad y, por tanto, la ley que sigue la desviación suma es la ley de probabilidad.

Estos razonamientos nos conducen a la conclusión de que el sistema de representación que debe elegirse para una distribución estadística, debe surgir de la ley normal.

En esto consiste la teoría de series de Charlier y Brüns, principalmente.

Charlier propuso tres tipos de series. El tipo A, basado en la ley normal de probabilidad, es de la forma:

$$F(x) = a_0 f_0 + \frac{a_3}{\underline{|3}} f_0^{\mathrm{III}} + \frac{a_4}{\underline{|4}} f_0^{\mathrm{IV}} + \cdots$$

siendo

$$f_0 = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

y, por tanto,

$$f_0^{\mathrm{r}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} H_r$$

siendo H_r un polinomio de grado *r*, llamado de Hermite.

Estos polinomios de Hermite forman un sistema ortogonal, por lo que la determinación de los parámetros a_0, a_3, ..., se hace de la forma conocida, y tendrán los siguientes valores:

$$a_0 = 1 \;,,\; a_3 = -\frac{\mu_3}{\underline{|3}} \quad a_4 = \frac{\mu_4 - 3}{\underline{|4}}$$

siendo μ_3, μ_4, ... los momentos de orden tres, cuatro, etc., tomando por origen el primer momento y por unidad el segundo, es decir, la desviación "standard".

El tipo *B* se basa en la ley de probabilidad de los casos raros o de Poisson:

$$f_0(x) = \frac{e^{-m} m^x}{\underline{|x}}$$

siendo *m* el valor medio de *x* : $m = \mu(x)$

El desarrollo es de la forma

$$F(x) = f_0(x) + b_2 \Delta^2 f_0 + b_3 \Delta^3 f_0 + \cdots$$

siendo

$$\Delta f_0 = f_0(x) - f_0(x-1) = \frac{e^{-m} m^x}{\underline{|x}} - \frac{e^{-m} m^{x-1}}{\underline{|x-1}} =$$

$$= \frac{e^{-m} m^x}{\underline{|x}} \left(1 - \frac{x}{m}\right) = f_0 q_1$$

En general,

$$\Delta^r f_0 = f_0 q_r$$

los polinomios *q* son ortogonales y definen un sistema de representación, cuyos parámetros se calculan de la forma ordinaria, siendo

$$b_2 = \frac{\sigma^2 - m}{2} \text{ (}\sigma\text{ desviación cuadrática).}$$

Este desarrollo es muy apropiado para casos de distribución discontinua.

El tercer tipo de series de Charlier, llamado *C*, es de la forma

$$F(x) = e^{(\Sigma c_r H_r(x))} f_0$$

siendo H_r polinomios de Hermite y t_0 de la ley normal, es decir, el tipo *C* es una exponencial, siendo el exponente el desarrollo según el tipo *A*.

Este *tipo* tiene la ventaja de que jamás la función de distribución se hará negativa, no haciendo preciso una normalización previa, como puede ser conveniente en los tipos anteriores.

Brüns propuso un desarrollo de la forma

$$S(x) = N\left[\frac{1}{2} + \frac{1}{2} S_0(x) + A_3 \frac{S_0^{III}(x)}{2} + A_4 \frac{S_0^{IV}(x)}{2^4} + \right]$$

siendo

$$S_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{x}{2}} dx$$

N es igual al total de las frecuencias.

Se ve que se trata de frecuencias acumuladas, y por ello este desarrollo será conveniente cuando los datos estadísticos vengan en esa forma dados.

Romanovsky también ha estudiado desarrollos en que aparecen polinomios ortogonales, partiendo de las curvas de Pearson.

Todos estos desarrollos cumplen las condiciones que imponíamos a un ajuste racional, pues el sistema representativo se basa en la ley de probabilidad, obteniendo en sucesivas aproximaciones las variaciones que producen causas especiales que imprimen carácter al colectivo y, por otra parte, al definir las funciones obtenidas un Espacio de Hilbert, los parámetros se determinarán por el valor más probable.