Aplicaciones del espacio de Hilbert a la Estadística

Por

FRANCISCO JAVIER URBELZ IBARROLA

Actuario, Catedrático de Estadística (1)

1. INTRODUCCION

En este artículo expondremos brevemente el concepto de espacio de Hilbert y su fundamentación axiomática para dedicarnos, en principio, a las aplicaciones estadísticas en sus dos aspectos: cuando los elementos correspondientes son de naturaleza funcional o estocástica.

En las aplicaciones funcionales trato, en principio, de las funciones ortogonales en general y especialmente dedico atención preferente a las funciones de Fourier para, entre otros, demostrar la relación entre la función característica y su función de densidad.

Siguiendo el mismo método de las funciones ortogonales (generalizadas respecto a una función ponderativa o núcleo) estudiraemos, a título de ejemplo, algunos desarrollos sencillos, como son los de Charlier y de Hermite.

De forma semejante, podrían construirse polinomios ortogonales, como son los de Legendre, Thebychev, Laguerre, Hermite ,etc., que comentaremos brevemente.

La segunda parte de nuestro estudio, preferentemente la dedicamos al espacio estocástico de Hilbert. En esta parte, puntualizaremos la métrica de este espacio que, como se verá, son esperanzas matemáticas.

Es importante la formación de elementos estocásticos ortonormalizados; es decir, construir una base y cualquier elemento de naturaleza estocástica representarlo en esta base. La determinación de las coordenadas en la base elegida, es similar a la de naturaleza funcional. Consecuencia inmediata es la posibilidad de representar una recta o un plano de regresión por medio de la base estocástica ortonormalizada elegida, que tiene la propiedad de ser mínima la varianza residual.

---

(1) Este trabajo está basado en algunos conceptos básicos de mi tesis doctoral de Ciencias Económicas: "Análisis Espectral de Procesos Estocásticos Estacionarios".

2. ESPACIO DE HILBERT

2.1. AXIOMÁTICA.

A finales del siglo pasado se estudiaron las propiedades primarias de matrices. Los métodos del álgebra lineal, estudian dos conceptos de espacios y se aplican a sus transforrmaciones, formas de operadores, proyectores, etc., aunque, como dice Máltsev (*) el objeto del álgebra lineal es el estudio de las "matrices, espacios y formas algebraicas, cuyas teorías están estrechamente vinculadas".

Son inmensas las aplicaciones a la Estadística y a la Econometría y he creído necesario dedicar estas breves páginas en forma elemental al estudio de los conceptos básicos de los espacios de Euclídes y de Hilbert, para después aplicarlos a la Estadística.

En los espacios abstractos se denominan elementos a un conjunto de puntos, *variables estocásticas,* vectores, *funciones* que gozan determinadas propiedades, que se establecen como axiomas o postulados. Particularmente importantes son los espacios euclídeo —de n dimensiones—, y el espacio de Hilbert, de infinitas dimensiones.

Sintetizaremos la axiomática para ambos espacios, representando por H el espacio de Hilbert o euclídeo —n— dimenisones; pero cuando expresamente nos refiramos al euclídeo, representaremos el espacio H_n indicando el subíndice la dimensionalidad.

I.—*Axioma de grupo.*—Dados dos elementos pertenecientes al espacio x, e $y \in H$ está definida la operación adición, con las propiedades de poseer el elemento neutro, el simétrico. (La operación adición es conmutativa y asociativa.)

II.—*Axioma de linealidad.*—Para todo $x \in H$ y para cualquier número complejo está definida la operación producto cx o $c.x.$ que goza con respecto a la adición de ser distributiva a derecha e izquierda y en consecuencia si $x, y \in H, c(x+y)=cx+cy \in H$; $c(dx)=(cd)x \in H$ donde c y d son números escalares cualesquiera. El producto por 0 y 1 son: $0x=0 \quad 1x=1$.

III.—*Axioma métrico.*—Se define una operación interna (también denominada producto interno) con las siguientes propiedades:

(*) A. I. Máltsev: "Fundamentos de Algebra Lineal", Editorial Mir, Moscú, 1972, pág. 12.

Si $x, y \in H$ denominamos producto interno al número complejo (x, y) que cumple las siguientes condiciones:

$$(x, y) = \overline{(y, x)}$$

$$(ax, y) = a(x, y)$$

a un escalar cualquiera.

$$(x+y, z) = (x, z) + (y, z) \qquad [1]$$

$$(ax+by, z) = a(x, z) + b(y, z)$$

corolario de las dos precedentes.

$$(x, ay) = \bar{a}(x, y)$$

igualmente corolario de las dos primeras y también

$$(x, x) \geqslant 0$$

Se denomina longitud de x o norma:

$$//x// = (x, x)^{\frac{1}{2}} \qquad [2]$$

y al producto:

$$(x-y, x-y)^{\frac{1}{2}} = //x-y// \qquad [3]$$

Se llama distancia del elemento x al elemento y; si es cero $x=y$; y mayor que cero cuando $x \neq y$.

IV. *Axiomas complementarios de Hilbert.*—Axiomas de convergencia: Dos axiomas se introducen en este espacio:

a) Propiedad de Cauchy: Si $x_n \in H$ y dado un $\varepsilon > 0$ por pequeño que sea, si para un $n > N$ se verifica:

$$//x_n - x_{n+p}// < \varepsilon \qquad [4]$$

existe un elemento x al que converge x_n

$$\lim_{n \to \infty} //x - x_n// = 0 \qquad [5]$$

Suele sustituirse este axioma por la propiedad denominada espacio completo.

b) Separabilidad.—Cuando hay una sucesión $(x_i \in H)$ la que se aproxima a $x \in H$ con error arbitrariamente pequeño, podemos del espacio H extraer una sucesión de elementos de $x_n \in H$ de forma que:

$$//x - x_n// < \varepsilon \qquad [6]$$

Se denomina a esta propiedad espacio separable cuando el error es arbitrariamente pequeño.

2.2. CONCEPTOS DE ORTOGONALIDAD.

Dados dos elementos $x, y \in H$ se denominan ortogonales cuando:

$$(x, y)=0 \qquad [1]$$

La ortogonalidad se representa también de la forma:

$$\underline{x \mid y} \qquad [1']$$

En el caso particular, si para $x \in H$ el producto escalar es:

$$(x, x)=1 \qquad [2]$$

se denomina al elemento x que está normalizado o tiene módulo unidad.

En un conjunto de puntos: $e_h \in H_n\,(h=1, 2, 3 \ldots n)$ que poseen las propiedades [1] y [2].

$$(e_h, e_k)=\delta_{hk} \quad \begin{array}{ll} h=k & \delta_{hh}=1 \\ h\neq k & \delta_{hk}=0 \end{array} \qquad [3]$$

donde δ_{hk} es símbolo de Kronocker.

Tal conjunto de puntos e_h se denomina: elementos ortonormalizados.

En una sucesión $x_1, x_1, \ldots x_n$ de elementos de H (o de H_n) que cumplen la [1] por los axiomas [1] de 2.1. su producto interno es:

$$//x_1 + x_2 + x_3 \ldots x_n//^2 = //x_1//^2 + //x_2//^2 + //x_3//^2 \ldots //x_n//^2 \qquad [4]$$

Si el espacio es de Hilbert, la expresión [4]

$$\|x\|^2 = \lim_{n\to\infty} \sum_{i=1}^{n} |x_i|^2 \qquad [4']$$

Cuando x es ortogonal a todo elemento de un subconjunto L de H decimos que es ortogonal a L y escribimos $\underline{x \mid L}$. Si L_1 y L_2 son dos subconjuntos, tales que todo elemento de L_1 es ortogonal a todo elemento de L_2 diremos que ambos subconjuntos son ortogonales si cumplen [1] para todos los elementos de los subconjuntos correspondientes.

2.3. COORDENADAS DE LOS ESPACIOS H_n Y H.

Sea el conjunto [3] de 2.2. de elementos ortonormalizados, pudiendo ser n elementos (espacio euclídeo) o un conjunto numerable (espacio de Hilbert).

Todo punto $x \in H_n$ (o $x \in H$) es representable por una combinación lineal.

$$x=x_1e_1+x_2e_2+\ldots+x_ne_n \qquad [1]$$

donde e_i son los elementos básicos [3] de 2.2. ortonormalizados. Si n tiende a infinito el punto $x \in H$. [2]

Multiplicando escalarmente la [1] por e_i tenemos:

$$(x, e_i) = x_i \qquad [3]$$

recordando las [3] de 2.2.

A x_i se denomina la coordenada x_i; y x puede representarse en función de sus coordenadas, por la [1] o una sucesión ordenada:

$$x = (x_1, x_2 \ldots x_i \ldots x_n) \qquad [4]$$

Cuando $x \in H$ la sucesión de las coordenadas se extiende a infinito.

2.4. Proceso de ortogonalización de Gram-Schmidt.

Hemos partido de la hipótesis que conocíamos una base en un espacio H_n o H.

A veces se desconoce; y el método de Schmidt se utiliza en los espacios funcionales para determinar una base.

Sea:

$$v_1, v_2 \ldots v_n \qquad [1]$$

un sistema de elementos pertenecientes a un espacio H.

Se pretende construir un sistema ortonormalizado:

$$(e_1, e_2 \ldots) \qquad [2]$$

de forma que el elemento e_j se exprese en función lineal de los elementos $v_1, v_2 \ldots v_n \ldots$

En principio, formemos los elementos e'_i auxiliares de la forma siguiente y a partir de ellos los elementos e_i que son ortonormalizados:

$$e'_1 = e_1 = \frac{v_1}{(v_1, v_1)\,1/2}; \qquad (e_1, e_1) = 1$$

$$e'_2 = v_2 - (v_2, e_1)\,e_1; \qquad e_2 = \frac{e'_2}{\sqrt{(e'_2, e'_2)}}$$

$$e'_3 = v_3 - (v_3, e_1)\,e_1 - (v_2, e_2)\,e_2; \qquad e_3 = \frac{e'_3}{\sqrt{(e'_3\ e'_3)}}$$

$$e'_n = v_n - \sum_{i=1}^{n-1} (v_n, e_i)\,e_i; \qquad e_n = \frac{e'_n}{\sqrt{(e'_n, e'_n)}} \qquad [3]$$

Los vectores e'_i son auxiliares y de los e_i se comprueba la ortogonalización de manera sencilla. Para:

$$(e'_2, e_1)=(v_2, e_1)-(v_2, e_1)(e_1, e_1)=0$$

por la [3].

Supuesto $j<n$ por inducción,

$$(e'_2, e_j) = (v_n, e_j) - \sum (v_n, e_i)(e_i, e_j) = (v_n, e_j) - (v_n, e_j) = 0 \qquad [4]$$

y la normalización de los vectores o elementos e_i es inmediata a partir de los auxiliares.

2.5. Distancia mínima de $x \in H$ a un subespacio H_n.

El problema planteado es encontrar una descomposición de x en dos: uno perteneciente al espacio H_n y otro $y \notin H_n$ de forma que $\underline{y / H_n}$

$$x = x' + y \qquad [1]$$

El elemento $x' \in H_n$ será combinación lineal de los elementos básicos de este subespacio:

$$x' = \sum_{i=1}^{n} a_i e_i \qquad [2]$$

Resolveremos por mayor sencillez en el caso de ser los elementos $x \in H$ *reales*, en cuyo caso para $\forall\, x, y \in H$ será:

$$(x, y)=(y, x) \qquad [3]$$

de acuerdo con el axioma métrico III de 2.1.

La distancia $//D^2//$ es:

$$// D //^2 = // x - x' //^2 = \left(x - \sum_{i=1}^{n} a_i e_i, x - \sum_{i=1}^{n} a_j e_j\right) =$$

$$= (x, x) - \sum_{i=1}^{n} (x_i, e_i)^2 + \sum_{i=1}^{n} [(x_i, e_i) - a_i]^2 \qquad [4]$$

La distancia $//D//$ no depende de los dos primeros sumandos y sí del término no negativo:

$$\sum_{i=1}^{n} [(x, e_i) - a_i]^2 \qquad [5]$$

Si elegimos:

$$a_i = (x, e_i) = x_i \qquad [6]$$

se anulan todos los términos de [5] y el valor mínimo de [4] se reduce:

$$\text{Min} \, // \, D \, //^2 = (x, x) - \sum_{i=1}^{n} (x, e_i)^2 = (x, x) - \sum_{i=1}^{n} x_i^2 \qquad [7]$$

recordando la [6]).

La [1] la escribiremos sustituyendo a_i por los valores determinados por la [6]:

$$x' = \sum_{i=1}^{n} x_i \, e_i \qquad [8]$$

Si las coordenadas se han elegido según la [6] de la [1] y [7] deducimos para (y, y):

$$(y, y) = (x, x) - \sum_{i=1}^{n} x_i^2 \qquad [9]$$

La distancia x a su aproximación x' combinación lineal $x' \in H_n$ puede hacerse tan pequeña como queramos. Si el conjunto es denso, decimos que x se puede aproximar con expresiones lineales con "error cuadrático" y escribimos:

$$x \stackrel{2}{=} \sum_{i=1}^{n} x_i \, e_i \qquad [10]$$

Cuando $n \to \infty$ y el error tiende a cero, llamamos convergencia cuadrática a la expresión [10].

3. FUNCIONES ORTOGONALES EN EL ESPACIO DE HILBERT

3.1. Generalidades.

La elección de los elementos básicos pertenecientes al espacio de Hilbert, pueden ser funciones según dijimos y subrayamos en 2.1.

En tal caso, se define previamente el campo de variabilidad $C=(a, b)$ que supondremos un intervalo cualquiera.

Llamamos sucesión de funciones ortonormalizadas dentro del intervalo (a, b) a:

$$f_0, f_1, f_2 \ldots f_n \qquad [1]$$

con las condiciones:

$$\int_a^b f_i \bar{f}_j \, dx = \delta_{ij} = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases} \qquad [2]$$

Definamos la operación:

$$(f, g) = \int_a^b f\bar{g}\,dx \qquad [3]$$

para $\forall f, g$ y examinemos si podemos formar un espacio funcional de Hilbert, con los siguientes axiomas:

I. *Axioma de grupo.*—Dadas dos funciones de [1] cualquiera, $f, g \in H$, la suma $f+g \in H$; la diferencia $f-g \in H$; existe un elemento neutro y el simétrico que pertenecen a H.

II. *Axioma de linealidad.*—Dadas dos funciones cualesquiera $f, g \in H$ y los números complejos a y b, $af+bg \in H$.

III. *Axioma métrico.*—Para $\forall f$, g, $h \in H$ con la definición dada en [3] para el *producto escalar*, son inmediatas las *relaciones:*

$$\begin{aligned} &(f, g) = \overline{(g, f)} \\ &(a f, g) = a(f, g) \\ &(f, f) = \int_a^b |f|^2\,dx \end{aligned} \qquad [4]$$

IV. *De convergencia*:

a) *Criterio de Cauchy*:

$$\begin{aligned} &\forall g_n, g_m \in H \\ &\|g_n - g_{n+p}\| \leqslant \varepsilon \end{aligned} \qquad [5]$$

existe un elemento $g \in H$ tal que:

$$\lim_{n\to\infty} \|g - g_n\| < \varepsilon \qquad [6]$$

b) *Separabilidad.*

Se puede elegir una sucesión de funciones pertenecientes a H, de forma que si $g_n \in H$ todo $f \in H$ puede expresarse con error arbitrariamente pequeño.

3.2. Representación por funciones ortogonales.

Si $f \in H$ dada la sucesión de funciones [1], cumpliendo la [2] de 3.1, podemos representar la combinación lineal:

$$g_n = a_0 f_0 + a_1 f_1 + \ldots + a_n f_n \qquad [1]$$

siendo la coordenada:

$$a_k = \int_a^b f \cdot \bar{f}_k \, dx \qquad [2]$$

deducida la distancia mínima de f a la combinación [1] y de forma semejante, como hemos deducido en 2.5.

La distancia de f a la combinación lineal [1] cuando sea menor que ε debe suceder:

$$\| f - g \|^2 = \int_a^b | f - g |^2 \, dx < \varepsilon \qquad [3]$$

y en este caso escribiremos:

$$f \stackrel{2}{=} \sum_{k=1}^{n} a_k f_k = g_n \qquad [4]$$

Si $f(x)$ es una función de cuadrado integrable, el error disminuye al agregar términos en [4] y cuando la distancia tiende a cero estamos en el concepto b) del axioma 4.° de convergencia de 3.1. respecto a ser el espacio funcional separable.

3.3. Convergencia uniforme y cuadrática.

La convergencia uniforme de una función $f(x)$ aproximada por una combinación lineal [1] de 3.2. para $x \in (a, b)$ es:

$$|f(x) - g(x)| < \varepsilon \qquad [1]$$

aproximándose uniformemente a $f(x)$ en el intervalo (a, b).

La convergencia uniforme de $g(x) \to f(x)$ en (a, b) implica la convergencia cuadrática.

$$f(x) \stackrel{2}{=} g(x) \qquad [2]$$

porque la distancia de $f(x)$ a $g(x)$ según [3] de 3.2 será:

$$\int_a^b |f(x) - g(x)|^2 dx < \varepsilon^2 (b - a) < \varepsilon_1 \qquad [3]$$

recordando la [1]; luego:

$$g(x) \stackrel{2}{=} f(x) \qquad [4]$$

3.4. Funciones ortonormalizadas de Fourier.

La sucesión de funciones [1] de 3.1, dentro del campo $-a$, $+a$, puede ser:

$$f_0(x) = \frac{1}{\sqrt{2a}}$$

$$f_1(x) = \frac{1}{\sqrt{a}} \cos \frac{\pi x}{a}; \quad \ldots \quad f_{2n-1}(x) = \frac{1}{\sqrt{a}} \cos \frac{n \pi x}{a} \qquad [1]$$

$$f_2(x) = \frac{1}{\sqrt{a}} \operatorname{sen} \frac{\pi x}{a}; \quad \ldots \quad f_{2n}(x) = \frac{1}{\sqrt{a}} \operatorname{sen} \frac{n \pi x}{a}$$

Esta sucesión de funciones se comprueba sencillamente que es una base ortonormalizada de un espacio funcional por las propiedades de las integrales de senos y cosenos.

La representación de un elemento $f \in H$ por el conjunto de sucesiones ortonormalizadas de Fourier, representando a'_n y b'_n las coordenadas en este espacio, es:

$$f(x) = \frac{a'_0}{\sqrt{2a}} + \frac{1}{\sqrt{a}} \left[a'_1 \cos \frac{\pi x}{a} + b'_1 \operatorname{sen} \frac{\pi x}{a} \right] + \ldots$$

$$\ldots + \frac{1}{\sqrt{a}} \left[a'_n \operatorname{sen} \frac{n \pi x}{a} + b'_n \operatorname{sen} \frac{n \pi x}{a} \right] + \ldots \qquad (n = 1, 2, 3, \ldots) \qquad [2]$$

por la [2] de 3.2. tenemos:

$$\left(f(x), \frac{1}{\sqrt{2a}} \right) = a'_0 = \frac{1}{\sqrt{2a}} \int_{-a}^{a} f(x)\, dx \qquad [3]$$

$$a'_1 = \left(f(x), \frac{\cos(\pi x/a)}{\sqrt{a}} \right) = \int_{-a}^{a} \frac{f(x)}{\sqrt{a}} \cos \frac{\pi x}{a}\, dx = a'_1 = \int_{-a}^{a} f(x) \cos \frac{\pi x}{a}\, dx$$

$$b'_1 = \left(f(x), \frac{\operatorname{sen}(\pi x/a)}{\sqrt{a}} \right) = \frac{1}{\sqrt{a}} \int_{-a}^{a} f(x) \operatorname{sen} \frac{\pi x}{a}\, dx$$

$$\ldots$$

$$a'_n = \frac{1}{\sqrt{a}} \int_{-a}^{a} f(x) \cos \frac{n \pi x}{a}\, dx \qquad [4\,a]$$

$$(n = 1, 2, 3, \ldots)$$

$$b'_n = \frac{1}{\sqrt{a}} \int_{-a}^{a} f(x) \operatorname{sen} \frac{n \pi x}{a}\, dx \qquad [4\,b]$$

$$\ldots$$

3.5. FUNCIÓN CARACTERÍSTICA.

A partir de la [2] en 3.4, podemos deducir la función característica en relación con la función de densidad. Sustituyendo en [2] las coordenadas [4 a] y [4 b], tenemos:

$$f(x) = \frac{1}{2a}\int_{-a}^{a} f(t)\,dt + \sum_{n=1}^{\infty}\frac{1}{a}\int_{-a}^{a} f(t)\left[\cos\frac{n\pi t}{a}\cos\frac{n\pi x}{a} + \operatorname{sen}\frac{n\pi t}{a}\operatorname{sen}\frac{n\pi x}{a}\right]dt =$$

$$= \frac{1}{2a}\int_{-a}^{a} f(t)\,dt + \sum_{n=1}^{\infty}\frac{1}{a}\int_{-a}^{a} f(t)\left[\cos\frac{n\pi(t-x)}{a}\right]dt =$$

$$= \frac{1}{2a}\int_{-a}^{a} f(t)\,dt + \sum_{n=1}^{\infty}\frac{1}{2a}\int_{-a}^{a} f(t)\cos\frac{n\pi(t-x)}{a}\,dt +$$

$$+ \sum_{n=-1}^{-\infty}\frac{1}{2a}\int_{-a}^{a} f(t)\cos\frac{n\pi(t-x)}{a}\,dt = \sum_{-\infty}^{\infty}\frac{1}{2a}\int_{-a}^{a} f(t)\cos\frac{n\pi(t-x)}{a}\,dt \quad [1]$$

por ser la función coseno función par. Precisamente por esta circunstancia y por ser la función seno impar, la [1] es igual:

$$f(x) = \sum_{n=-\infty}^{\infty}\frac{1}{2a}\int_{-a}^{a} f(t)\left[\cos\frac{n\pi(t-x)}{a} + i\operatorname{sen}\frac{n\pi(t-x)}{a}\right]at =$$

$$= \int_{-a}^{a} f(t)\,\frac{1}{2a}\left[\sum e^{\frac{n\pi(t-x)}{a}i}\right]dt \qquad [2]$$

Si hacemos

$$\frac{n\pi}{a} = u \qquad [3]$$

como $n \in N$ cuando varía de n a $n+1$, tenemos que el incremento de u es:

$$\frac{(n+1)\pi}{a} - \frac{n\pi}{a} = \frac{\pi}{a} = \delta u \qquad [4]$$

Llevando este valor a la [2], tenemos:

$$f(x) = \frac{1}{2\pi}\int_{-a}^{a} f(t)\,dt \sum_{u} e^{u(t-x)i}\,\delta u \qquad [2']$$

y cuando $a \to \infty$, [4] tiende a la diferencial du y a la sumatoria [2'] a una integral.

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t)\, dt \int_{-\infty}^{\infty} e^{u(t-x)i}\, du \qquad [5]$$

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-uxi}\, du \int_{-\infty}^{\infty} f(t)\, e^{iut}\, dt \qquad [5']$$

Llamando función característica a la expresión:

$$\varphi(u) = \int_{-\infty}^{\infty} f(t)\, e^{iut}\, dt = \int_{-\infty}^{\infty} f(x)\, e^{iux}\, dx \qquad [6]$$

Sustituyendo en [5'] tenemos que conociendo esta función puede determinarse la función de densidad de probabilidad. Sustituyendo la [6] en la [5'], tenemos:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-uxi}\, \varphi(u)\, du \qquad [7]$$

La reciprocidad de las expresiones [6] y [7] es bien conocida y como pretendíamos hemos demostrado estas fórmulas conociendo las expresiones ortonormalizadas de Fourier.

Para que $f(x)$ sea desarrollable en serie de Fourier [1] debe cumplir las condiciones de Dirichlet: Tener un número finito de discontinuidades en el período; tener un número finito de máximos y mínimos y

$$\int_{-a}^{a} / f(x) /\, dx < \infty$$

4. FUNCIONES ORTOGONALES EN EL ESPACIO DE HILBERT RESPECTO DE UNA FUNCION PONDERATIVA (NUCLEO).

4.1. AXIOMÁTICA.

Este caso puede reducirse al estudiado en 3. Unicamente el Axioma III es donde la operación escalar tiene una significación diferente.

Los Axiomas I, II, III y IV de 3.1. se dan por reproducidos y entonces en el Axioma métrico III se introduce la función ponderativa.

$$h(x) \geqslant 0 \qquad [1]$$

denominada núcleo.

Sea un conjunto de funciones:

$$f_0, f_1, f_2, \ldots, f_n, \ldots \qquad [2]$$

definidas en un intervalo (a, b) la operación escalar $\forall f,\ g \in H$ respecto al núcleo $h(x)$ se define así:

$$(f, g) = \int_a^b f(x),\ \overline{g(x)},\ h(x)\,dx \qquad [3]$$

A veces representamos por una integral de Stieltjes-Lebesgue con función ponderativa $dF(x) \geqslant 0$. La [3] se definiría:

$$(f, g) = \int_a^b f(x)\,\overline{g(x)}\ dF(x) \qquad [3']$$

Si [2] son funciones del espacio funcional de Hilbert respecto del núcleo $h(x)$ o $dF(x)$ (según sea [3] o [3'] se puede reducirlas al espacio funcional estudiado en 3, sólo haciendo la sustitución de las funciones [2] por las nuevas:

$$k_i(x) = f_i(x)\sqrt{h(x)} \qquad [4]$$

habiendo eliminado el núcleo.

4.2. Ortogonormalización y coordenadas.

Decimos tenemos una base ortonormalizada

$$g_1(x),\ g_2(x),\ \ldots \qquad [1]$$

respecto del núcleo $h(x)$ cuando se cumple:

$$(g_i(x),\ g_j(x)) = \int_a^b g_i(x),\ g_j(x) h(x) dx = \delta_{ij} \qquad [2]$$

Una función $f(x) \in H$ puede expresarse por una combinación lineal de la base [1].

$$f(x) = a_1 g_1(x) + a_2\, g_2(x) + \ldots + a_n g_n'(x) + \varepsilon(x) \qquad [3]$$

El término $\varepsilon(x)$ se elige de forma que sea mínima la distancia de $f(x)$ a la expresión:

$$f_n(x) = a_1 g_1(x) + a_2 g_2(x) + \ldots + a_n g_n(x) \qquad [4]$$

La distancia de $f_n(x)$ a $f(x)$ se determina por el producto escalar.

$$// f(x) - f_n(x) //^2 = [f(x) - \Sigma\, a_i g_i(x),\ f(x) - \Sigma\, a_j g_j(x)] \qquad [5]$$

Es evidente y de forma semejante a como demostramos 2.5, esa distancia es mínima cuando los coeficientes de [4] son las ordenadas de $f(x)$ en el espacio de la base ortonormalizada [1], es decir, cuando:

$$a_k = [f(x), g_k(x)] = \int_a^b f(x)\,\overline{g_k(x)}\,h(x)\,dx \qquad [6]$$

a_k es la coordenada k del desarrollo o representación de $f(x)$ por [3].

De [5] deducimos, teniendo en cuenta [2]:

$$//\varepsilon(x)// = //f(x) - f_n(x)// = [f(x), f(x)] - \sum_{k=1}^{n} /a_k/^2 \qquad [7]$$

4.3. Desarrollo de Charlier.

Una aplicación inmediata de lo estudiado en 4.2 es considerar como núcleo la función de densidad normal tipificada, cuyo campo de variabilidad es el eje real.

Formemos las expresiones:

$$\begin{aligned} g_0(x) &= 1 \\ g_1(x) &= P_1(x) \\ g_2(x) &= \frac{P_2(x)}{\sqrt{|\underline{2}}} \\ g_m(x) &= \frac{P_m(x)}{\sqrt{|\underline{m}}} \\ &\dots\dots\dots\dots\dots \end{aligned} \qquad [1]$$

Las expresiones $P_m(x)$ son los polinomios de las derivadas de la districión normal tipificada:

$$f_0(x) = \frac{1}{\sqrt{2\pi}}\, e^{-(x^2/2)} \qquad [2]$$

$$f^{(m}(x) = \frac{1}{\sqrt{2\pi}}\, e^{-(x^2/2)}\, P_m(x) \qquad [3]$$

Se comprueba sencillamente que $\{g_0(x), g_1(x), g_2(x), \dots\}$ [1] es una base ortonormalizada respecto a la función de densidad:

$$h(x) = \frac{1}{\sqrt{2\pi}}\, e^{-(x^2/2)} \qquad [4]$$

Evidentemente para $m>n$

$$[g_m(x), g_n(x)] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)} \frac{P_m(x)}{\sqrt{|m}} \frac{P_n(x)}{\sqrt{|n}} dx =$$

$$= \frac{1}{\sqrt{|m}\sqrt{|n}} \int_{-\infty}^{+\infty} f^{(m}(x) P_n(x)\, dx =$$

$$= \frac{1}{\sqrt{|m}\sqrt{|n}} \left[f^{(m-1}(x) P_n(x) \right]_{-\infty}^{+\infty} - \int_{-\infty}^{-\infty} f^{(m-1}(x) P'_n(x)\, dx =$$

$$= \frac{(-1)^2}{\sqrt{|m}\sqrt{|n}} \left[\int_{-\infty}^{\infty} f^{(m-2}(x) P_n''(x)\, dx \right] = 0 \qquad [5]$$

pues al integrar por partes deducimos en consecuencia, al ser $m>n$ por hipótesis y ser nulos los términos integrados entre $+\infty$ y ser

$$P_n^{(n+1)}(x) = 0 \qquad [6]$$

por ser $P_n(x)$ un polinomio de grado n, deducimos la ortogonalidad de las funciones:

$$[g_m(x), g_m(x)] = \frac{1}{\sqrt{|m}\sqrt{|m}} \left[\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)} P_m(x) P_m(x)\, dx \right] =$$

$$= \frac{1}{\sqrt{(|m)^2}} \left[\int_{-\infty}^{\infty} f^{(m}(x) P_m(x)\, dx \right] = \frac{1}{|m} \left[(-1)^k \int_{-\infty}^{\infty} f^{(m-k}(x) P^{(k}(x)\, dx \right] =$$

$$= \frac{(-1)^m}{|m} \cdot P_m^{(m}(x) \int_{-\infty}^{\infty} f^{(0}(x)\, dx = (-1)^m (-1)^m \frac{|m}{|m} = 1 \qquad [5']$$

por ser $P_m^{m}(x)$ una constante y además es factorial de m con el signo $\pm$ según sea m de naturaleza par o impar.

Puede escribirse:

$$(g_m(x), g_n(x) = \delta_{m,n} \qquad [5'']$$

La base [1] formada como hemos visto, está ortonormalizada con el núcleo [3].

Una función cualquiera $f(x)$ puede desarollarse por medio del sistema ortonormalizado [1] en la forma:

$$f(x) = a_0 g_0(x) + a_1 g_1(x) + a_2 g_2(x) + \ldots + a_m g_m(x) + \ldots \qquad [7]$$

En consecuencia las coordenadas con núcleo [4] se determinan por la expresión [6] de 4.2.

$$a_k = [f\, x), g_k(x)] = \int_{-\infty}^{\infty} f(x) P_k(x) \frac{e^{-(x^2/2)}}{\sqrt{2\pi}} dx \qquad [8]$$

En nuestro caso recordemos la [3] y formemos un desarrollo de la función de densidad $f(x)$ con relación a la función normal $f_0(x)$ y sus derivadas $f^{(m}(x)$ $(m=1, 2, \ldots)$, en forma de combinación lineal.

$$f(x)=a_0 f_0(x)+a_1 f^{(1}(x)+ \ldots +a_k f^{(k}(x)+ \ldots \qquad [9]$$

Multiplicando escalarmente en [9] por $P_k(x) / \underline{|k}$ tenemos:

$$\left(f(x), \frac{P_k(x)}{\underline{|k}}\right)=\int_{-\infty}^{\infty} f(x) \frac{P_k(x)}{\underline{|k}}\, dx = \frac{a_k}{\underline{|k}} \int_{-\infty}^{\infty} f^{(k}(x) P_k(x)$$

Luego:

$$a_k=\int_{-\infty}^{\infty} f(x) \frac{P_k(x)}{\underline{|k}}\, dx \qquad [10]$$

porque todas las integrales se nos anulan por las [5], [5'] y [5''].

Dando valores a $k=0, 1, 2 \ldots$ encontramos los coeficientes del desarrollo [9] denominado de Charlier.

Solamente a título de información daremos los tres primeros. Para ello conozcamos $P_k(x)$ deducimos $P_0(x)=1$ de la derivada de la normal [3]:

$$\begin{aligned} P_1(x)&=-x \\ P_2(x)&=x^2-1 \\ P_3(x)&=-(x^3-3x) \end{aligned} \qquad [11]$$

En consecuencia:

$$\int_{-\infty}^{\infty} f(x)\, dx = a_0 \int_{-\infty}^{\infty} f_0(x)\, dx = 1 \Rightarrow$$

$$a_0 = 1$$

$$a_1 = \int_{-\infty}^{\infty} f(x) \cdot \frac{P_1(x)}{\underline{|1}}\, dx = - \int_{-\infty}^{\infty} x\, f(x)\, dx = - \alpha_1$$

$$a_2 = \frac{1}{\underline{|2}} \int_{-\infty}^{\infty} (x^2-1) f(x)\, dx = \frac{1}{2} [\alpha_2 - 1]$$

$$a_3 = \frac{1}{\underline{|3}} \int_{-\infty}^{\infty} (3x - x^3) f(x)\, dx = \frac{1}{\underline{|3}} [3\alpha_1 - \alpha_3] \qquad [12]$$

. . . .

Estos coeficientes, llevados a la [9] nos dan el desarrollo de Charlier para una función de densidad en relación con la normal.

Indiquemos que para la validez del desarrollo son necesarias las condiciones de continuidad de $f(x)$ y se anule la función y sus derivadas en $+\infty$.

4.4. Otros desarrollos. Desarrollos de Hermite.

La analogía con los de Charlier nace por el hecho que se forman de la normal por derivación, pero de la forma siguiente:

Si

$$f_0(x) = \frac{1}{\sqrt{2\pi}}\, e^{-(x^2/2)} \tag{1}$$

y hacemos el operador derivada:

$$D = \frac{d}{dx} \tag{2}$$

$$D^r = \frac{d}{dx^r} \tag{3}$$

tenemos:

$$(-D)^r f_0(x) = f_0(x) \cdot H_r(x) \tag{4}$$

y en consecuencia:

$$f_0^{(r}(x) = (-1)^r H_r(x) f_0(x) \tag{5}$$

siendo:

$$P_r(x) = (-1)^r H_r(x) \tag{6}$$

los polinomios [3] de 4.3. estudiados en el desarrollo de Charlier (*).

Rey Pastor (**) da una tabla de polinomios ortogonales con diversos núcleos, interesantes y que enunciamos, como son los de Legendre, Tchebychev, Laguerre, Jacobi, etc., así como las funciones generatrices de los polinomios.

5. ESPACIO ESTOCASTICO DE HILBERT

5.1. Axiomática.

En 2.1. postulamos, en general, el espacio de Hilbert y subrayamos diciendo que los elementos pertenecientes al espacio podían ser variables estocásticas.

En el espacio de Hilbert cuando la naturaleza de los elementos sean variables estocásticas, lo llamaremos espacio estocástico de Hilbert.

(*) M. G. Kendall and Stuard: "The Advanced Theory of Statistics", vol. I, pág. 155 y ss., quien demuestra propiedades importantes de estos y otros polinomios.

(**) J. R. Pastor: "Los Problemas Lineales de la Física". Intaet., pág. 32 y siguientes.

Añadiremos, sin embargo, que pueden ser funciones aleatorias o procesos estocásticos, pero no se estudiarán estas cuestiones en este artículo.

I. *Axioma de grupo.*—Si x, y son variables aleatorias $x, y \in H$ deberá suceder que:

$$x+y \in H; \qquad x-y \in H$$

Existe elemento neutro 0, y el simétrico que pertenecen al espacic.

II. *Axioma de linealidad.*—También se verifica el axioma II de 2.1. para combinaciones lineales $ax+by \in H$ y demás propiedades ya indicadas.

III. *Axioma métrico.*—Definimos para el espacio estocástico $\forall_x, y \in H$ la operación interna:

$$(x, y) = \int_{R_2} x\bar{y} d^2F(x, y) = Ex\bar{y} \tag{[1]}$$

El símbolo E es el operador esperanza; y R_2 el campo bidimensional; y el valor conjugado de y cuando las variables aleatorias tomen valores complejos: y $F(x, y)$ la función de distribución conjunta de x e y.

En el caso de ser reales, tenemos que la [1] sería:

$$(y, x) = (x, y) = \int_{R_1} xyd^2F(x, y) = Exy \tag{[1']}$$

Definida la [1'] para variables aleatorias reales, comprobemos cumple el axioma métrico las propiedades siguientes:

$$(x, y) = (y, x) \tag{[2a]}$$

$$(ax, y) = a(y, x) \tag{[2b]}$$

$$(x+y, z) = (x, z) + (y, z) \tag{[2c]}$$

$$(ax+by, z) = a(x, z) + b(y, z) \tag{[2d]}$$

corolario de las precedentes.

$$^a(x, ay) = a(x, y) \tag{[2e]}$$

igualmente corolario de las anteriores.

$$(x, x) \geqslant 0 \tag{[2f]}$$

Demostremos la [2c]:

$$(x+y, z) = E[(x+y)z] = Exz + Eyz \tag{[3]}$$

por la propiedad del operador esperanza E.

IV. *Axiomas complementarios de convergencia.*—Damos por reproducidos los axiomas IV a) y b), siendo los elementos x y x_n variables estocásticas.

5.2. Construcción de una base estocástica ortonormalizada.

En un espacio estocástico H_n —euclídeo de n dimensiones— a partir de unas variables estocásticas podemos construir otras variables estocásticas ligadas por una relación lineal y que tengan la propiedad de estar ortonormalizadas. Cuando $n \to \infty$ formaremos el espacio de Hilbert.

En principio, y sin pérdida de generalidad, supondremos las variables:

$$x_1, x_2, x_3 \dots \qquad [1]$$

están centradas; es decir:

$$Ex_i = 0 \qquad [2]$$

Si no cumpliesen [2] y fuese α_k la media teórica correspondiente a la variable x_k, la nueva variable sería:

$$x_k - \alpha_k \qquad [2']$$

y siempre supondremos las variables aleatorias x_i son centradas sin que perdiéramos generalidad por esta hipótesis adicional.

A partir de las variables centradas

$$x_2, x_3 \dots \qquad [3]$$

construyamos unas variables estocásticas ortonormalizadas y que llamaremos:

$$\xi_2, \xi_3 \dots \qquad [4]$$

Recordemos el procedimiento de ortogonalización de Gram Schmidt expuesto en 2.4. y expresamente sustituiremos en [2] el conjunto $\{v_k\}$ por las correspondientes variables aleatorias $\{x_{k+1}\}$ e igualmente para las variables ortogonales e_k' y e_k por sus correspondientes variables aleatorias ξ_{k+1}' (auxiliar y ortogonal) y ξ_{k+1} ortonormalizadas.

Según hemos indicado, tendremos:

$$\xi_{2'} = \xi_2 = \frac{x_2}{\sqrt{(x_2, x_2)}} \qquad [5]$$

pero:

$$(x_2, x_2) = Ex_2^2 = \sigma_2^2 \qquad [6]$$

según [1] de 5.1. y por ser x_2 centrada; luego la [5] puede escribirse:

$$\xi_2 = \frac{x_2}{\sigma_2} \quad \Rightarrow \quad (\xi_2, \xi_2) = 1 \qquad [5']$$

o sea, la variable tipificada que goza de la propiedad de tener módulo unidad.

Encontremos ξ_3 pero antes introduzcamos la variable auxiliar ξ'_3:

$$\xi'_3 = x_3 - (x_3, \xi_2) \cdot \xi_2 \quad [7]$$

$$\Rightarrow \text{ es } (\xi'_3, \xi_2) = 0 \quad [7']$$

recordando [5'].

Si llamamos:

$$E x_i x_j = \sigma_{ij} = \rho_{ij} \sigma_i \sigma_j \quad [8]$$

donde ρ_{ij} es el coeficiente de correlación entre las variables x_i y x_j. Tenemos:

$$(x_3, \xi_2) = E x_3 \frac{x_2}{\sigma_2} = \frac{\sigma_{23}}{\sigma_2} = \frac{\sigma_{23}}{\sigma_2 \sigma_3} \cdot \sigma_3 = \rho_{23} \sigma_3 \quad [9]$$

Normalicemos la variable auxiliar ξ'_3

$$(\xi'_3, \xi'_3) = [x_3 - (x_3, \xi_2) \xi_2, x_3 - (x_3, \xi_2) \xi_2] = \sigma^2_3 (1 - \rho^2_{23}) \quad [10]$$

Luego:

$$\xi_3 = \frac{\xi'_3}{\sqrt{(\xi'_3, \xi'_3)}} = \frac{x_3 - (x_3, \xi_2) \xi_2}{\sqrt{1 - \rho^2_{23}}} \quad [11]$$

La [11] puede escribirse recordando las [5'], [9] y [10]:

$$\xi_3 = \frac{1}{\sqrt{1 - \rho^2_{23}}} \cdot \frac{x_3}{\sigma_3} - \frac{\rho_{23}}{\sqrt{1 - \rho^2_{23}}} \cdot \frac{x_2}{\sigma_2} \quad [11']$$

Procederíamos continuando formando la base estocástica ortonormalizada siguiendo el método expuesto en 2.4. para nuestro espacio estocástico y la base obtenida es combinación lineal de las variables estocásticas [3].

5.3. Hiperplano de regresión en coordenaras estocásticas ortonormalizadas.

Construida la base $\xi_2, \xi_3 \dots \xi_n$ un elemento x_1 podemos representarlo estocásticamente en esta base de forma:

$$x_1 = b_{12} \xi_2 + b_{23} \xi_3 + \dots + b_{1n} \xi_n + \cdot \eta \quad [1]$$

Si el elemento aleatorio x_1 se proyecta sobre H_n y η es ortogonal a la base $\{\xi_2, \xi_3 \dots \xi_n\}$.

El problema es determinar los parámetros b_{1k} ($h=2, 3 \dots n$) que llamaremos coeficientes de regresión respecto al sistema ortogonalizado. Se deducen sencillamente aplicando los concepto de coordenadas (ya que son las coordenadas de la base estocástica ortogonal) que vimos en 2.5. fórmula [6] y que, en nuestro caso, se reduce a:

$$b_{1k}=(x_1, \xi_k)=Ex_1\xi_k \qquad [2]$$

Dando a k valores, obtenemos todos los coeficientes del hiperplano de regresión.

De la [1] y cuando se han determinado los coeficientes por la fórmula [2] tenemos la varianza residual o error mínimo:

$$(\eta, \eta) = \sigma^2_{1,2,\dots n} = (x_1, x_1) - \sum_{k=2}^{n} b^2_{1k} = \sigma_1{}^2 - \sum_{k=2}^{n} b^2_{1k} \qquad [3]$$

Esta expresión puede compararse con la [9] de 2.5.

Si $x_1 \in H_n$ el error será nulo porque la distancia de x_1 al subespacio H_n (donde se encuentra x_1) es nula. Podemos representar en este caso, por una expresión lineal:

$$x_1 = \sum_{k=1}^{n} b_{1k}\, x_k \qquad [4]$$

A $\sigma^2_{1.2,\dots n}$ se denomina varianza residual y viene expresada en términos de la varianza de x_1 y de las coordenadas sobre el sistema ortogonal elegido, disminuyendo cuando se incrementa la dimensionalidad del espacio pudiendo llegar a ser nulo cuando la proyección de x_1 en el sistema ortogonal estocástico elegido coincida exactamente con la dimensionalidad de x_1 según hemos indicado.

Brevemente completaremos estas ideas y emplearemos las notaciones introducidas por Yule (*) para nuestras regresiones.

La [1] la escribiremos así:

$$x_1 = b_{12.3\dots n}\xi_2 + b_{13.24\dots n}\,\xi_3 + \dots + b_{1n,23\dots(n-1)}\,\xi_n + x_{1,23\dots n} \qquad [5]$$

Llamaremos (como hace Yule) a los primeros subíndices primarios y a los restantes secundarios. El primer subíndice $_1$ en los coeficientes de regresión en base ortonormalizada indica la variable x_1; el segundo subíndice de los coeficientes, indicativo de que es el coeficiente correspondiente a la variable estocástica ortonormalizada, representa su influencia: los coeficientes

(*) G. U. Yule y M. G. Kendall: "Introducción a la Estadística". Aguilar, 1954, pág. 304 y ss.

secundarios representativos de las restantes variables ortonormalizadas que intervienen. Por último, comparando [5] con [1] vemos que:

$$\eta = x_{1.23\dots n} = x_1 - b_{12,3\dots n}\,\xi_2 - \dots - b_{1n.2\dots(n-1)}\,\xi_n \qquad [6]$$

es la desviación o error estocástico.

Multiplicando escalarmente $x_{1.23\dots n}$ por $\xi_h\,(h=2,3\dots n)$ y recordando [2], tenemos:

$$\begin{aligned}(x_{1.23\dots n}, \xi_h) &= E\,x_{1,23\dots n}\,\xi_h = \\ &= E\,(x_1 - \Sigma\, b_{1j.2.3\dots(j-1),(j+1)\dots n}\,\xi_j)\,\xi_h = \\ &= E\,x_1\,\xi_h - b_{1h.23(h-1)(h+1)\dots n} = 0\end{aligned} \qquad [7]$$

según la [6]. Este hecho nos permite escribir:

$$\begin{aligned}\sigma^2_{1.2\dots n} &= (x_{1.23\dots n}, x_{1,23\dots n}) = \\ &= (x_{1,23\dots n}, x_1 - b_{12,3\dots n}\,\xi_2 - b_{13.2\dots n}\,\xi_3 - \dots - b_{1n.2\dots n}\,\xi_n \\ &= E\,x_{1.23\dots(n-1)} - b_{1n.23\dots(n-1)}\,\xi_n)\,x_1 = \\ &= E\,x_{1.23\dots(n-1)}\,x_1 - b_{1n,23\dots(n-1)}\,E\,\xi_n\,x_1 = \\ &= E\,x^2_{1.23\dots(n-1)} - b^2_{1n,23\dots(n-1)} = \sigma^2_{1.23\dots(n-1)} - b^2_{1n.23\dots(n-1)}\end{aligned} \qquad [8]$$

La varianza residual:

$$\sigma^2_{1.2\dots n} = \sigma^2_{1.23\dots(n-1)} - b^2_{1n.23\dots(n-1)} \qquad [8']$$

La relación [8'] puede escribirse:

$$\sigma^2_{1.23\dots n} = \sigma^2_{1.23\dots(n-1)}\left[1 - \frac{b^2_{1n.23\dots(n-1)}}{\sigma^2_{1.2\dots(n-1)}}\right] \qquad [8'']$$

La [8] es la varianza residual de orden $n-2$ y por la [8'] y [8''] vemos la ley de recurrencia para conocido la de orden inferior determinar la siguiente.

La fórmula, deducida por Yule (pág. 307 o. c.) para los coeficientes de regresión, nos permite escribir la importante relación para nuestros coeficientes de regresión en base estocástica ortonormalizada:

$$\rho^2_{1n.23\dots(n-1)} = \frac{b^2_{1n.23\dots(n-1)}}{\sigma^2_{1.23\dots(n-1)}} \qquad [9]$$

pues sustituyendo [9] en [8''] nos da la fórmula deducida por Yule.

$$\sigma^2_{1.23\dots n} = \sigma^2_{1.23\dots(n-1)}\,(1 - \rho^2_{1n\dots(n-1)}) \qquad [10]$$

De [10] deducimos:

$$\begin{aligned} \sigma^2_{1.2} &= \sigma^2_1(1-\rho^2_{12}) \\ \sigma^2_{1.23} &= \sigma^2_{1.2}(1-\rho^2_{13.2}) \\ \sigma^2_{1.234} &= \sigma^2_{1.23}(1-\rho^2_{14,23}) \\ &\cdots\cdots\cdots\cdots \end{aligned} \qquad [11]$$

En consecuencia, multiplicando las [11] hasta la [10] y simplificando, tenemos:

$$\sigma^2_{1.23\ldots n} = \sigma^2_1(1-\rho^2_{12})(1-\rho^2_{13,2})(1-\rho^2_{14,23})\ldots(1-\rho^2_{1n,23\ldots(n-1)}) \qquad [12]$$

En la [9] deducimos:

$$b^2_{1n,23\ldots(n-1)} = \rho^2_{1n,23\ldots(n-1)}\cdot\sigma^2_{1,23\ldots(n-1)} \qquad [13]$$

y sustituyendo la relación [12] en [13] tenemos para nuestros coeficientes de regresión en base ortonormalizada:

$$b^2_{1n,23\ldots(n-1)} = \rho^2_{1n,23\ldots(n-1)}\cdot\sigma^2_1(1-\rho^2_{12})\cdot(1-\rho^2_{13,2})\ldots(1-\rho^2_{1(n-1),23\ldots(n-2)}) \qquad [14]$$

Estos coeficientes $b_{1n,23\ldots(n-1)}$ puede observarse no están influenciados por variables superiore a n, como indica la [14].

Esta ventaja sobre los coeficientes lineales ordinarios de regresión nos permite deducir el siguiente sin modificar los anteriores cuando se ha construido la base ortonormalizada.

5.4. Recta de regresión.

La [1] de 5.3 puede reducirse a la menor expresión lineal posible:

$$x_1 = b_{12}\,\xi_2 + x_{1.2} \qquad [1]$$

y como por la [2] de 5.1, en nuestro caso es:

$$b_{12} = E\,x_1\,\xi_2 = \frac{E\,x_1\,x_2}{\sigma_2} = \frac{E\,x_1\,x_2}{\sigma_1\,\sigma_2}\cdot\sigma_1 = \rho_{12}\,\sigma_1 \qquad [2]$$

La [1] se reduce a:

$$x_1 = \rho_{12}\,\sigma_1\,\xi_2 + x_{1.2} = \rho_{12}\,\sigma_1\frac{x_2}{\sigma_2} + x_{1.2} \qquad [3]$$

recordando las [5] y [8] de 5.3.

Llamando a la proyección sobre ξ_2, recta de regresión de x_1 sobre x_2, con un subíndice secundario para indicar esta combinación lineal:

$$x_{1t} = \rho_{12}\frac{\sigma_1}{\sigma_2}\cdot x_2 \qquad [4]$$

se llama a x_{1t} regresión lineal de x_1 sobre x_2 y puede escribirse:

$$\frac{x_{1t}}{\sigma_1} = \rho_{12}\cdot\frac{x_2}{\sigma_2} \qquad [5]$$

Sencillamente deducimos la varianza residual sin más que en [3] de 5.3 suprimir los subíndices superiores a 2 y sustituyendo b_{12} por el valor deducido en [2] tenemos la conocida fórmula de la varianza residual:

$$\sigma^2_{1.2} = \sigma^2_1 - b^2_{12} = \sigma^2_1 - \sigma^2_1\,\rho^2_{12} = \sigma^2_1\,(1-\rho^2_{12}) \qquad [6]$$

5.5. Plano de regresión.

Por su importancia, dedicaremos estas breves líneal al plano de regresión en coordinadas ortonormalizadas y finalizaremos nuestro artículo.

De las [1] y [5] de 5.3 tenemos para el plano en la base ortonormalizada $\{\xi_2, \xi_3\}$:

$$x_1 = b_{12}\,\xi_2 + b_{13}\,\xi_3 + x_{1.23} \qquad [1]$$

siendo la desviación $x_{1.23}$ ortogonal a ξ_2 y ξ_3.

Las coordenadas b_{12} y b_{13} se deducen de la [2] de 5.3.

$$b_{12} = E\,x_1\,\xi_2; \qquad b_{13} = E\,x_1\,\xi_3 \qquad [2]$$

Siendo $\xi_2 = (x_2/\sigma_2$

$$\xi_3 = \frac{1}{\sqrt{1-\rho^2_{23}}}\cdot\frac{x_3}{\sigma_3} - \frac{\rho_{23}}{\sqrt{1 \quad \rho^2_{23}}}\cdot\frac{x_2}{\sigma_2} \qquad [3]$$

según las [5], [5'] y [11']) estando ortonormalizadas.

La coordenada b_{12} ha sido deducida en 5.4, fórmula [2].

$$b_{12} = \rho_{12}\,\sigma_1 \qquad [4]$$

Sustituyendo en [2] la [3] tenemos:

$$b_{13} = \frac{(\rho_{13} - \rho_{12}\,\rho_{23})}{\sqrt{1-\rho^2_{23}}} \cdot \sigma_1 \qquad [5]$$

Si prescindimos en [1] de la desviación $x_{1.23}$ y llamamos x_{1t} al plano de regresión, tenemos:

$$x_{1t} = b_{12}\,\xi_2 + b_{13}\,\xi_3 \qquad [6]$$

Sustituyendo en [6] las coordenadas expresadas en las fórmulas [5] y las ξ_2 y ξ_3 por la [3], tenemos:

$$x_{1t} = \rho_{12}\,\sigma_1 \frac{x_2}{\sigma_2} + \frac{(\rho_{13} - \rho_{12}\,\rho_{23})}{\sqrt{1-\rho^2_{23}}}\,\sigma_1 \cdot$$

$$\cdot\left(\frac{1}{\sqrt{1-\rho^2_{23}}}\,\frac{x_3}{\sigma_3} - \frac{\rho_{23}}{\sqrt{1-\rho^2_{22}}}\,\frac{x_2}{\sigma_2}\right) \qquad [7]$$

Efectuando operaciones podemos ponerla en forma explícita tipificada:

$$\frac{x_{1t}}{\sigma_1} = \frac{\rho_{12} - \rho_{13}\,\rho_{23}}{1-\rho^2_{23}} \cdot \frac{x_2}{\sigma_2} + \frac{\rho_{13} - \rho_{12}\,\rho_{23}}{1-\rho^2_{23}}\,\frac{x_3}{\sigma_3} \qquad [8]$$

La varianza risidual [3] de 5.3. para nuestro caso es:

$$\sigma^2_{1.23} = \sigma^2_1 - b^2_{12} - b^2_{13} = \sigma^2_1 - \rho^2_{12}\,\sigma^2_1 - \frac{(\rho_{13} - \rho_{13}\,\rho_{23})^2}{1-\rho^2_{23}}\,\sigma^2_1 =$$

$$= \sigma^2_1\,\frac{(1-\rho^2_{12})\,(1-\rho^2_{23}) - (\rho_{13} - \rho_{12}\,\rho_{23})^2}{1-\rho^2_{23}} \qquad [9]$$

Indiquemos que teniendo una base ortonormalizada, el error disminuirá a medida que introduzcamos términos sin necesidad de rehacer los cálculos, como lo hemos hecho para el plano conociendo ya b_{12} calculado previamente en la recta.

BIBLIOGRAFIA

A. I. Maltsey: *Fundamentos de Algebra Lineal,* Edit. Moscú, 1972.

J. Rey Pastor: *Los problemas lineales de la Física.* Intact, 1955.

J. Rey Pastor y Pi Calleja: *Análisis Matemático.* Vol. 3, Kapeluca, Buenos Aires, 1965.

G. U. Yule y Kendall: *Introducción a la Estadística,* Aguilar, 1954.

O. Fernández Baños: *Tratado de Estadística,* Consejo Superior de Investigaciones Científicas, 1945.

Karl Karhunen: *Métodos Lineales en el Cálculo de Probabilidades.* Consejo Superior de Investigaciones Científicas, 1952.

I. M. Gulfand, G. E. Chilov: *Les distributions,* tomo 2, Espaces Fundamentaux, Dunod, 1964.

Béla Sz Nagy: *Analyse Harmonique des Operateurs de l'Espace de Hilbert,* Masson, 1967.

H. Cramer: *Stationary and Related Stochastic Processes.* John Wiley and Son., 1966.

J. L. Doob: *Stochastic Processes,* J. Wiley, 7.ª ed., 1967.

S. K. Berberian: *Introducción al espacio de Hilbert,* Teide, 1967, London.

E. Lukacs: *Characteristic Function,* Charles Griffin, 1960, London.

Anderson: *The Statistical Analysis of Time Series,* Wiley, 1958.