

## **MATRIZ DE PROYECCION Y DISEÑOS OPTIMOS PARA REGRESION**

**Edith SEIER\***

*La Matriz de Proyección  $H$ , de amplio uso en la delegación de observaciones potencialmente influyentes cuando se tiene un conjunto de datos destinados a estimar un modelo de Regresión, se utiliza en este trabajo para analizar los Criterios de Optimalidad del Diseño Experimental para el caso particular de los modelos de Regresión. Esto permite definir de una manera muy sencilla un tipo de diseños que son  $G$ -óptimos y a partir de allí obtener una serie de consecuencias útiles en la práctica.*

---

\* Facultad de Ciencias Matemáticas  
Universidad Nacional Mayor de  
San Marcos

## 1. Introducción.

Desde hace varios años y debido a que el Trabajo computacional que supone la estimación y docimacia de parámetros de un modelo de Regresión Múltiple se ha simplificado enormemente gracias a la popularización de las computadoras y los respectivos programas, los estadísticos han centrado su interés en la observación y análisis más cuidadoso de los datos, el modelo estimado y sus implicancias. Uno de los aspectos que ha recibido considerable interés es el llamado Análisis de Influencia o Sensibilidad, el cual analiza las consecuencias en el modelo de diferentes tipos de perturbaciones que pueden afectar a sus elementos. Entre las perturbaciones más estudiadas están la omisión de una o varias observaciones perturbaciones en algún regresor o perturbaciones de la distribución asumida para el error. La importancia del análisis de influencia cobra importancia ante el surgimiento de métodos de estimación alternativos a los tradicionales y que se conocen con el nombre de *Regre-*

*sión Resistente y Regresión Robusta* debido a que el análisis de influencia nos permite detectar las situaciones en que es necesario aplicar estos métodos alternativos, por lo general bastante laboriosos.

Para desarrollar herramientas a fin de realizar el análisis de influencia hay que delimitar el ámbito de trabajo en lo que se refiere a tres aspectos: el modelo, el método de estimación y el tipo de perturbación. Al respecto, el tema más desarrollado en la literatura y de más fácil tratamiento es el análisis de influencia para el estimador mínimo cuadrático (que también es estimador máximo verosímil y mejor estimador lineal insesgado)

$$\hat{\beta} = [X'X]^{-1}X'Y$$

para el modelo lineal de regresión múltiple

$$Y = X \beta + \varepsilon \text{ con } \varepsilon \sim N(0, \sigma^2 I)$$

cuando la perturbación es la omisión de una de las observaciones  $(y_i, x_{1i}, x_{2i}, \dots, x_{ri})$ . En ese análisis la matriz

$H = X[X'X]^{-1}$  ocupa un lugar preponderante.

Por otro lado, en los últimos 30 años se ha desarrollado una teoría de Diseño Experimental para modelos lineales en general (de regresión y análisis de varianza) donde el diseño se concibe con una distribución de probabilidad definida sobre la región experimental y en la que se definen diversos criterios de optimalidad. Los aportes de *J.C.Kiefer* a esta teoría han sido considerables<sup>(1)</sup>. Sin embargo no siempre los diseños logrados con la teoría aproximada son factibles de ser llevados a la práctica. La versión discreta (distribución que asigna una probabilidad  $1/n$  a cada uno de  $n$  puntos no necesariamente diferentes) de estos diseños es lo que se conoce como diseños exactos, los cuales sí son siempre posibles de realizar.

---

(1) El tercer volumen de "Collected Papers de J.C. Kiefer", publicado luego de su muerte, está íntegramente dedicado a Diseño Experimental y consta de más de 700 páginas.

En este trabajo, que tuvo como motivación original la obtención de un diseño óptimo para un caso práctico dentro del ámbito de la ingeniería electrónica (*Seier & Romberger* [1990]) utilizamos la matriz  $H$  y sus propiedades para lograr una mejor comprensión de lo que proponen los criterios de optimalidad en el caso específico de los modelos de regresión y los diseños exactos. En la sección 2 se hace un breve resumen del papel de la matriz  $H$  y sus propiedades y en la sección 3 se hace lo propio con los criterios de optimalidad. En la sección 4 se utiliza la noción de influencia potencial, no en el sentido que es usual para datos ya obtenidos, sino para diseñar experimentos de modo que no hayan datos potencialmente influyentes. El resultado de esto es muy alentador ya que se logra demostrar que los diseños así definidos cumplen la propiedad de ser  $G$ -óptimos. En la sección 4 se exploran algunas de las consecuencias del resultado principal de la sección 3 útiles en el trabajo práctico de diseñar experimentos

para obtener datos a fin de estimar posteriormente modelos de regresión.

## 2. La Matriz de Proyección

Es un resultado muy conocido en Regresión que para el modelo lineal de regresión múltiple

$$Y = X \beta + \epsilon \quad \text{con} \quad \epsilon \sim N(0, \sigma^2 I)$$

mediante el método mínimo cuadrático (que no requiere la hipótesis específica de la normalidad de los errores), mediante el método máximo verosímil, o tratando de buscar el mejor estimador lineal insesgado se llega al estimador

$$\hat{\beta} = [X'X]^{-1}X'Y$$

y que

$$E(\hat{\beta}) = \beta$$

$$\text{Var}(\hat{\beta}) = \sigma^2 [X'X]^{-1}$$

De esto se desprende que

$$\hat{Y} = X \hat{\beta} = X[X'X]^{-1}X'Y$$

lo cual se puede escribir como

$$\hat{Y} = H Y \quad (1)$$

donde  $H = X[X'X]^{-1}X'$ , por lo que

$$\begin{aligned} \text{Var}(\hat{Y}) &= H \text{Var}(Y) H' \\ &= \sigma^2 H \end{aligned}$$

debido a que  $\text{Var}(Y) = \sigma^2 I$  y a que la matriz  $H$  es simétrica e idempotente.

A la matriz  $H$  se le conoce con el nombre de matriz "sombrero" o "matriz de proyección". A pesar que varias de las propiedades de  $X[X'X]^{-1}X'$  aparecen ya en textos de fechas anteriores, la introducción de  $H$  (llamada  $P$  por otros autores) con un nombre propio y captando una mayor atención se encuentra en *Hoaglin & Welsh [1978]*. Esta matriz tiene entre otras las siguientes propiedades fácilmente demostrables:

- i) es simétrica
- ii) es idempotente
- iii) es de dimensión  $n \times n$ , su rango y su

traza son ambos iguales a  $p$ .

iv) el valor promedio de los  $h_{ii}$  es  $p/n$

v) es invariante bajo transformaciones lineales siempre y cuando el modelo contenga el término constante  $\beta_0$ .

Debido a la expresión (1) cada valor estimado  $\hat{y}_i$  puede pensarse como una combinación lineal de los valores observados  $y_1, y_2, \dots, y_n$  donde los pesos  $h_{ij}$  con que intervienen cada uno de los  $y_j$ ,  $j = 1, \dots, n$  están exclusivamente determinados por los valores de las  $x_j$  por tanto esa estructura de pesos será la misma para cualquier otro modelo en que otra variable respuesta pretenda ser explicada por las mismas variables  $x_1, x_2, \dots, x_r$  mediante un modelo de similar estructura.

Merecen especial interés los elementos de la diagonal, es decir los pesos  $h_{ii}$  con que intervienen cada  $y_i$  en la construcción  $\hat{y}_i$ . Este peso es también el factor por el que queda multiplicado  $\sigma^2$  para formar  $Var(\hat{y}_i)$ .



Cuando hay una sola variable  $x$  y el modelo es una recta, al graficar los pesos  $h_{ii}$  vs.  $x_i$  se insinúa el gráfico de una parábola. Si ya no pensamos en los  $n$  valores aislados de  $x$  sino en que puede tomar cualquier valor en un intervalo dado,  $h$  sería una función cuadrática de  $x$  cuyo gráfico es una parábola con vértice  $(\bar{x}, \frac{1}{n})$  y por tanto dos valores de  $x$  a los que corresponde el mismo valor de  $h$  estarían unidos por un segmento de recta paralelo al eje  $x$ . Si hay dos regresores  $x_1, x_2$  el gráfico de  $h$  como función de  $x_1, x_2$  es un paraboloides elíptico y los cortes efectuados con planos paralelos al plano  $x_1, x_2$  darán origen a elipses que unen los puntos correspondientes a observaciones con valores similares de  $h$ , elipses cuyo eje principal coincide con la recta de regresión que explica  $x_2$  en función de  $x_1$ .

Los valores  $h_{ii}$  se utilizan como indicadores de la influencia potencial de los datos, es decir pesos  $h_{ii}$  relativamente altos indican que existe la posibilidad de

que la presencia de la  $i$ -ésima observación haga que el modelo sea considerablemente diferente de lo que sería si tal observación no existiera, debido a la posición de los valores de las variables  $x_1, x_2, \dots, x_r$ . Que tal cosa suceda, es decir que el dato realmente ejerza una influencia decisiva, depende por supuesto no sólo de la influencia potencial sino también de cuan concordante sea la relación entre  $y$  y  $x_1, x_2, \dots, x_r$  para esa observación con la relación entre esas variables para las otras observaciones. La influencia propiamente dicha se cuantifica mediante la estadística llamada "distancia de Cook" definida como

$$D_i = \frac{(\hat{\beta} - \hat{\beta}_{-i})' [X'X] (\hat{\beta} - \hat{\beta}_{-i})}{p s^2}$$

donde  $\hat{\beta}_{-i}$  denota el parámetro estimado omitiendo la  $i$ -ésima observación. La definición de  $D_i$  aparentemente involucra el trabajo de volver a realizar la estimación omitiendo cada uno de los  $n$  datos, uno a

la vez, pero es posible evitar este trabajo al llegar a la expresión

$$D_i = \left[ \frac{e_i}{1-h_{ii}} \right] \frac{h_{ii}}{p s^2}$$

que se obtiene utilizando el resultado del Teorema de *Sherman-Morrison-Woodbury* para escribir  $\hat{\beta}_{-i}$  en función de  $\hat{\beta}$  (ver por ejemplo Myers [1986]).

Esta forma de escribir la distancia de Cook, además de utilizarse comúnmente para su cálculo a partir de la estimación del modelo con todos los datos, permite apreciar la composición de  $D_i$  y facilitar la interpretación de los valores que resulten.

Así podemos identificar en  $D_i$  tres partes o aportes:

- influencia potencial, mediante  $h_{ii}$
- estandarización con fines de comparación mediante  $ps^2$
- discrepancia con respecto a los demás, a través de  $e_i/(1-h_{ii})$  (llamado residuo de predicción,  $e_{-i}$ ).

La influencia no depende pues sólo de los valores de  $x_1, x_2, \dots, x_r$  sino también de la relación entre  $y$  y  $x_1, x_2, \dots, x_r$ , por tanto en un diseño experimental donde imponemos los valores de  $x_1, x_2, \dots, x_r$  pero observamos el resultante valor de  $y$ , podemos controlar la influencia potencial más no así la influencia efectiva.

### **3. Los Criterios de Optimalidad en el Diseño Experimental**

Al diseñar y luego realizar un experimento o una serie de experimentos con idea de obtener datos para luego estimar un cierto modelo, se denomina *región experimental* al conjunto de valores posibles de las variables  $x_1, \dots, x_r$  y las combinaciones de ellos. La matriz  $X$  recibe el nombre de *matriz de diseño* y se denomina *matriz de información* a la matriz  $X'X$  de dimensión  $pxp$ . Un *diseño exacto* es una distribución de probabilidad que asigna una probabilidad  $1/n$  a cada uno de  $n$  puntos, no necesariamente diferentes de la región experimental. La teoría del diseño exacto se

puede considerar la versión discreta de la teoría general o aproximada en la cual se trabaja con una medida  $\xi$  definida sobre la región experimental  $X$  y donde la atención está centrada en la matriz de información  $M(\xi)$ . Para la teoría aproximada existe abundante literatura, rigurosa desde el punto de vista matemático, pero en la práctica un diseño exacto no siempre es realizable debido a que la proporción de observaciones a tomarse en determinado punto de la región experimental puede ser un número irracional.

Otro término que utilizaremos es el de *cápsula de las variables independientes* y que se refiere al conjunto convexo más pequeño que contiene a todos los puntos de diseño (una definición alternativa sería la intersección de los conjuntos convexos que contienen a todos los puntos del diseño) y que abreviaremos con *CVI*. Aquí trabajaremos con matrices de diseño o lo que es equivalente, con los puntos ya seleccionados de la región experimental donde se van a llevar a cabo los experimentos.

Las definiciones de los criterios de optimalidad los daremos ya particularizados para el caso de los diseños exactos.

Una parte considerable de la teoría del Diseño Experimental está dedicada a los criterios de optimalidad. Un resumen sobre ellos así como abundantes referencias bibliográficas para el caso de la teoría aproximada se encuentra en Atkinson(1988).

Los criterios de optimalidad se pueden dividir en dos clases, aquellos que se preocupan por la varianza de los estimadores de los parámetros ( $D, A$  y  $E$  optimalidad) y aquellos que fijan su interés en la varianza de los estimadores de la variable respuesta ( $V$  y  $G$  optimalidad).

**$D$ -Optimalidad.-** Un diseño se denomina  $D$ -óptimo si maximiza el valor de  $|X'X|$ . Se dice que el objetivo de la  $D$ -optimalidad es minimizar la varianza global (varianzas y covarianzas) de  $\hat{\beta}$  y minimizar por tanto el volumen del elipsoide descrito al realizar estimaciones por intervalos de los

parámetros. Es un hecho conocido que dada una región experimental rectangular, ese determinante se maximiza cuando los experimentos se localizan en los vértices de la región experimental. La  $D$ -optimalidad se preocupa pues por la cobertura de la región experimental por parte de  $CVI$ . Existen algoritmos para la generación de diseños  $D$ -óptimos en la práctica.

Algo importante acerca del criterio de  $D$ -Optimalidad es que es invariante bajo transformaciones lineales en la región experimental siempre y cuando el modelo incluya el término constante  $\hat{\beta}_0$ . Esto significa que si encontramos un diseño  $D$ -óptimo para una región experimental y un cierto modelo, la disposición relativa de los puntos dentro de la región experimental será la misma cuando la región sufre un alargamiento, contracción o desplazamiento por efecto de haberse aplicado transformaciones lineales a las variables.

**A-Optimalidad.-** Este criterio busca minimizar la varianza promedio de los estima-

dores de los parámetros. No toma en cuenta las covarianzas de los estimadores. Utilizando la notación  $C = (X'X)^{-1}$ , la varianza promedio de los  $\hat{\beta}_j$  será:

$$\frac{1}{p} \sum_{j=1}^p \text{Var} (\hat{\beta}_j) = \frac{\sigma^2}{p} \sum_{j=1}^p c_{jj} = \frac{\sigma^2}{p} \text{tr}(X'X)^{-1}$$

Asumiendo la varianza  $\sigma$  fija, el criterio de *A*-optimalidad busca minimizar la traza de  $(X'X)^{-1}$ . En John (1972) se demuestra que para minimizar esa traza los experimentos deben realizarse en puntos equidistantes del centro.

**E-Optimalidad.-** Este criterio busca minimizar la varianza de combinaciones lineales de los estimadores de los parámetros. No discutiremos por considerarlo más apropiado para los modelos de análisis de varianza.

**V-Optimalidad.-** De acuerdo a ese criterio se debe minimizar el promedio de las varianzas de las estimaciones de la variable respuesta



$$(1/n) \sum_{i=1}^n \text{Var}(\hat{y}_i)$$

Veamos qué forma toma este promedio para el caso de los modelos de regresión :

$$(1/n) \sum_{i=1}^n \text{Var}(\hat{y}_i) = (\sigma^2/n) \sum_{i=1}^n h_{ii} = \frac{p \sigma^2}{n}$$

De esa expresión deducimos que en el caso de los modelos de regresión el criterio de  $V$ -optimalidad no nos indica en cuales puntos de la región experimental conviene realizar los experimentos pero nos sugiere que el número de experimentos debe ser grande en relación al número de parámetros. Este es un primer resultado logrado en forma muy sencilla gracias a las propiedades de la matriz  $H$ .

**G-Optimalidad.**- Un diseño se dice  $G$ -óptimo si minimiza el valor de

$$\max \text{Var}(\hat{y}_i)$$

para  $i = 1, \dots, n$ . El criterio de  $G$ -optimalidad busca minimizar la máxima varianza del

estimador de la variable respuesta. Entre todos los diseños para un determinado modelo, será  $G$ -óptimo aquel diseño para el que

$$\max_i \text{Var} (\hat{y}_i) = \min_x \max_i \text{Var} (\hat{y}_i)$$

La construcción de diseños  $G$ -óptimos no es un problema comunmente atacado en la literatura. Cuando se trabaja en la teoría aproximada existe el célebre Teorema de la Equivalencia General de *Kiefer & Wolfowitz* (1960), razón por la cual con frecuencia cuando se trabaja en la teoría aproximada se construyen diseños  $D$ -óptimos y por el Teorema de la Equivalencia General se espera que dichos diseños sean también  $G$ -óptimos. Desafortunadamente para los diseños exactos dicha equivalencia entre  $G$ -óptimalidad y  $D$ -Optimalidad, no se cumple en general.

#### 4. Diseños justos y $G$ -optimalidad

Al diseñar un experimento el Estadístico decide las condiciones en que se rea-

lizará éste, de común acuerdo con el especialista en el campo de la experimentación. En otras palabras, los valores de  $x_1, x_2, \dots, x_r$  no son observados sino impuestos y por tanto la influencia potencial de las futuras observaciones puede ser decidida en forma arbitraria por el Estadístico. Imaginemos que, como medida preventiva para evitar en lo posible la existencia de datos influyentes, exigimos que todas las observaciones tengan la misma influencia potencial, es decir que  $h_{ii}$  el mismo valor para todo  $i$ . Como  $\sum_{i=1}^n h_{ii} = p$  entonces  $h_{ii} = p/n$  para todo  $i = 1, \dots, n$ . A esta condición impuesta sobre la matriz  $H$  le hemos llamado "condición -h" y al diseño resultante podemos llamarlo un "diseño justo". Si tal cosa sucede, como los  $h_{ii}$  son positivos (por tratarse de una matriz idempotente los elementos de la diagonal de  $H$  pueden ser pensados como sumas de cuadrados) y la suma  $\sum_{i=1}^n h_{ii}$  tiene un valor fijo, si pretendemos hacer uno de

los  $h_{ii}$  más pequeño, otro tendrá que volverse más grande. En otras palabras, no podremos conseguir otra matriz  $H$  correspondiente a otro diseño donde el máximo elemento de la diagonal sea menor que para nuestro diseño "justo". Pero precisamente ésa es la definición de los diseños que satisfacen la  $G$ -Optimalidad. A continuación formalizamos este hecho :

**Definición.-** Una matriz de diseño se dice que satisface la condición  $h$  si  $h_{ii}$  es constante para todos los valores de  $i$ .

**Definición.-** Para un modelo de regresión lineal con  $p$  parámetros y  $n$  observaciones, asumiendo que  $\sigma^2$  es constante, se dice que una matriz de diseño  $X_0$  es  $G$ -óptima en sentido local si

$$\max_i h_{0ii} = \min_X \max_i h_{ii}$$

(Usamos la expresión "en sentido local" simplemente para enfatizar que la optimalidad se refiere al tipo de modelo elegido y para el número  $n$  de observaciones; se

trata de un mayor rigor en la escritura antes que una pérdida de generalidad.

**Lema.-** Si una matriz de diseño satisface la condición  $h$  entonces es  $G$ -óptima en sentido local.

**Demostración.-** - Como  $X_0$  satisface la condición  $h$

$$h_{0ii} = p/n$$

Si suponemos  $X_0$  no es  $G$ -óptima esto significa que hay otra matriz de diseño  $X^*$  tal que para algún  $j$

$$h_{jj}^* = \max_i h_{ii}^* < p/n$$

Luego

$$\begin{aligned} p = \text{tr}(H^*) &= \sum_{i=1}^n h_{ii}^* \leq n \max_i h_{ii}^* \\ &\leq n h_{jj}^* < p \end{aligned}$$

lo cual no es posible. Luego si  $X_0$  satisface la condición  $h$  entonces es una matriz de diseño  $G$ -óptima en sentido local.

## 5. Consecuencias.

### i) Formas de lograr diseños $G$ -óptimos en la práctica

Los valores de  $h_{ii}$  están relacionados con la distancia de un determinado punto  $(X_{1i}, X_{2i}, \dots, X_{ri})$  del diseño con respecto al centro de gravedad definido por todos los puntos del diseño  $(i = 1, \dots, n)$  y que no es necesariamente el centro de la región experimental. Luego imponer la condición- $h$  equivale a elegir puntos equidistantes del centro de gravedad.

En el caso del modelo

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

para construir un diseño que satisfaga la condición  $h$  podemos trazar un polígono regular de  $n$  lados en el plano  $x_1 x_2$  y tomar las observaciones en los vértices de dicho polígono. El número de lados del polígono depende del número de experimentos que deseamos realizar.

Por supuesto no hay una sólo manera de imponer la condición  $h$  para este modelo y los diferentes valores de  $n$ . **Cualquier rotación de esos polígonos regulares dará también origen a un diseño  $G$ -óptimo.** Igualmente por ser la matriz  $H$  invariante bajo transformaciones lineales, utilizando dichas transformaciones sobre un diseño base que cumpla la condición- $h$  se puede lograr diseños  $G$ -óptimos.

Por otro lado, cuando el valor de  $n$  es factorizable en dos enteros  $a$  y  $k$ , se puede trazar un polígono de  $a$  lados y realizar  $k$  experimentos en cada uno de los vértices.

Así para  $n = 12$  tenemos las siguientes alternativas:

- 1 experimento en cada vértice de un dodecágono regular
- 3 experimentos en cada vértice de un cuadrado o rectángulo
- 4 experimentos en cada vértice de un triángulo isósceles

El modelo puede estar escrito en términos de funciones de las  $x_i$ , como por ejemplo

$$y_i = \beta_0 + \beta_1/x_{1i} + \beta_2/x_{2i} + \varepsilon_i$$

En este caso el polígono tiene que ser dibujado en el plano  $1/x_1, 1/x_2$  y no en el plano  $x_1, x_2$ . En una forma más general podemos decir que si el modelo es de la forma

$$y_i = \beta_0 + \beta_1 f_1(x_{1i}) + \beta_2 f_2(x_{2i}) + \varepsilon_i$$

el polígono elegido para producir un diseño  $G$ -óptimo tiene que ser trazado en el plano  $f_1(x_1), f_2(x_2)$ . Sin embargo hay por lo menos dos casos en que puede dibujarse el polígono en el plano  $x_1, x_2$  y la condición  $h$  también se cumplirá para el modelo escrito en funciones de estas variables. Estos dos casos -  $f_1$  y  $f_2$  no representan transformaciones lineales de las variables  $x_1$  y  $x_2$ , debido a que la matriz  $H$  es invariante bajo transformaciones no lineales siempre y cuando el modelo incluya el término independiente.

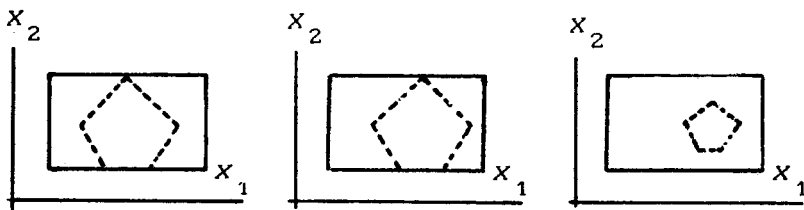


-  $f_1$  y  $f_2$  no representan necesariamente transformaciones lineales pero el polígono elegido es un rectángulo. Esto se debe a que como los puntos elegidos de la región experimental pueden ser escritos de la forma  $(a,b)$   $(a,c)$   $(d,b)$   $(d,c)$  los puntos  $(f(a),f(b))$   $(f(a),f(c))$   $(f(d),f(b))$   $(f(d),f(c))$  también describen un rectángulo.

Si hay tres regresores independientes entre sí, diseños formados por los vértices de poliedros regulares satisfacen la condición- $h$ .

## **ii) Diferencia intrínseca entre $G$ - y $D$ -Optimalidad**

El hecho que al imponer la condición  $h$  produce diseños  $G$ -óptimos hace ver que la  $G$ -optimalidad, a diferencia de la  $D$ -optimalidad, no se preocupa por la cobertura que se logra de la región experimental. Dada una región experimental rectangular y  $n=5$  cualquiera de los siguientes pentágonos produciría diseños  $G$ -óptimos al ubicar los experimentos en los vértices.



Obviamente mientras sea menor el área del pentágono menor será el determinante de la matriz de información y peor el diseño desde el punto de vista de la  $D$ -optimalidad.

Eso sí la  $G$ -optimalidad persigue un balance en la ubicación de los experimentos.

### iii) Diseños que son simultáneamente $D$ - y $G$ - óptimos

En 1960 *Kiefer & Wolfowitz* demostraron, para la teoría aproximada, la equivalencia entre los criterios de  $D$  y  $G$  optimalidad; sin embargo posteriormente en 1985 *Kiefer* demostró mediante un contraejemplo que esta equivalencia no necesariamente se cumple para los diseños exactos de  $n$  puntos. Nosotros podemos apreciar esto claramente a partir de ii).

Si la región experimental es rectangular y deseamos hacer un diseño que sea  $G$ -óptimo y  $D$ -óptimo a la vez, es evidente que para lograr cobertura total de la región experimental por parte de la CVI y balance en la posición de los experimentos necesitamos que  $n$  sea un múltiplo de 4. En ese caso se realizan un número igual de experimentos en cada vértice de la región experimental.

Resulta evidente también de ii) que para lograr diseños que sean  $D$  y  $G$  óptimos para el caso de dos regresores, si la región experimental no es rectangular debe tener una forma tal que sus vértices sean equidistantes del centro y que el número de observaciones debe ser múltiplo del número de vértices. Sólo así pueden satisfacerse cobertura y balance simultáneamente.

#### iv) Sobre el número de experimentos

Es evidente que para el modelo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_r x_r + \varepsilon$$

el mínimo número de experimentos es  $p = r + 1$  si los parámetros se van a estimar en forma puntual. Por ello  $n$  podrá tomar los valores  $p + 1, p + 2, p + 3, \dots$  etc. Si además deseamos imponer los criterios de  $D$  y  $G$  optimalidad simultáneamente habrá que hacer algunas restricciones sobre los valores de  $n$  según se desprende de iii). La comparación entre dos diseños con diferente número de experimentos puede hacerse en función de las varianzas de los estimadores de los parámetros y de la variable respuesta.

Es conveniente notar que como

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \dots + \hat{\beta}_r x_{ri}$$

la varianza de  $\hat{y}_i$  puede expresarse en función de las varianzas de las  $\hat{\beta}_j$  y las covarianzas entre las  $\hat{\beta}_i$  y  $\hat{\beta}_j$ . Si las varianzas y covarianzas de las  $\hat{\beta}_j$  se ven todas reducidas en un mismo porcentaje es de esperar que las varianzas de las  $\hat{y}_i$  se

vean reducidas en el mismo porcentaje. Nosotros trabajaremos primero a nivel de las varianzas de las  $\hat{y}_i$  por cuanto veremos que esto es muy sencillo para los diseños que satisfacen la condición- $h$ .

Si un diseño  $D_1$  satisface la condición  $h$  y comprende  $n_1$  experimentos se cumple:

$$\text{Var}_{D_1}(\hat{y}_i) = \frac{p}{n_1} \sigma^2 \quad \forall i \quad (2)$$

Para otro diseño  $D_2$  que también satisface la condición  $h$  pero que comprende  $n_2$  experimentos, tendremos:

$$\text{Var}_{D_2}(\hat{y}_i) = \frac{p}{n_2} \sigma^2 \quad \forall i \quad (2.a)$$

De las expresiones (2) y (2.a) observamos que si los dos diseños cumplen la condición  $h$  (por lo que además son  $G$ -óptimos) la relación entre sus varianzas es inversamente proporcional a la razón del número de experimentos de cada uno

$$\frac{Var_{D_1}(\hat{y}_i)}{Var_{D_2}(\hat{y}_i)} = \frac{\frac{p}{n_1} \sigma^2}{\frac{p}{n_2} \sigma^2} = \frac{n_2}{n_1} \quad \forall i \quad (3)$$

Es interesante notar que esta expresión depende sólo del número de experimentos de cada diseño y es independiente de los valores de la región experimental bajo la condición de que ambos diseños satisfagan la condición-*h*. Así la ganancia o pérdida relativa en términos de varianza de los estimadores al pasar de un diseño  $D_1$  a uno  $D_2$  podemos definirla como

$$\frac{Var_{D_1}(\hat{y}_i) - Var_{D_2}(\hat{y}_i)}{Var_{D_1}(\hat{y}_i)} \times 100 \quad (4)$$

De las expresiones anteriores vemos que (4) es igual a

$$\frac{n_2 - n_1}{n_2} \times 100 \quad (5)$$

De esta manera, si  $n_2$  es mayor que  $n_1$  la expresión en (5) tomará un valor positivo,

es decir ganaremos algo en términos de reducción de varianza de estimadores por haber aumentado el número de experimentos. En caso contrario, es decir si reducimos el número de experimentos, la expresión (5) tomará un valor negativo que representa la pérdida relativa de precisión.

Lo fácil que ha sido deducir una expresión general, independiente de la región experimental, es una razón más para preferir trabajar con diseños que satisfagan la condición- $h$ .

El cálculo de las expresiones (3) y (5) para diferentes valores de  $n_1$  y  $n_2$  es muy útil para ayudarnos a tomar la decisión acerca del número de experimentos a realizar. Una vez calculadas y graficadas esas expresiones utilizando como  $n_1$  el mínimo número de experimentos determinado de acuerdo a los criterios de la sección anterior, el estadístico y el experimentador pueden decidir cual disminución relativa en la varianza del estimador consideran que vale la pena y obtendrán como resulta-

do un valor adecuado para el número de experimentos a realizar.

Experimentalmente trabajamos analizando también la reducción en la varianza de los estimadores de los parámetros. Para diseños que asignan un número igual de experimentos en cada vértice de un rectángulo, la reducción en las varianzas y covarianzas de los estimadores de los parámetros se da en forma uniforme y coincidente en términos porcentuales con la reducción en  $Var(\hat{y}_i)$  pero es mucho más sencillo obtener una forma general  $Var(\hat{y}_i)$  tal como lo hemos hecho.

Hasta el momento la discusión acerca del número de experimentos ha sido llevada a cabo tomando en cuenta sólo la precisión pero no el costo. El Estadístico, como producto de la información que el experimentador le proporcione al respecto, puede definir una función costo  $c(n)$  que representa el costo (en unidades monetarias) de ejecución de un diseño con  $n$  experimentos.



Para poder contrastar el costo de realizar  $n$  experimentos y lo que ganamos al realizar esos  $n$  experimentos, necesitamos definir una función ganancia (en las mismas unidades monetarias que la función costo). Definamos, por ejemplo, como  $q$  la cantidad de unidades monetarias que ganamos al reducir  $Var(\hat{y}_i)$  y  $Var(\hat{\beta}_j)$  en un 10% con respecto a los valores que estas varianzas toman para el caso de realizar el mínimo número de experimentos, al que llamaremos  $n^*$ . Se puede definir una función ganancia  $g(n)$  que representa el número de unidades monetarias que ganamos al realizar un diseño con  $n$ . Si el diseño satisface la condición- $h$ , gracias a la expresión (5) se puede escribir:

$$g(n) = \frac{n - n^*}{n} \times q \times 10$$

En base a la función ganancia y a la función costo podemos definir otra función a la que podemos llamar función utilidad

$$u(n) = g(n) - c(n)$$

Cuando la función de utilidad toma valores negativos significa que lo que se gana en precisión no justifica el costo de su realización.

**v) La G-optimalidad y los diseños que incluyen observaciones en el centro de la reunión experimental**

La imposición de la condición- $h$  descartaría la inclusión de observaciones en el centro del polígono descrito por los otros datos. Este detalle puede llamar la atención a los usuarios de algunos diseños estándar formados por ejemplo por los vértices de un cuadrado y el centro del mismo. Cabe aclarar que aquí hemos estado trabajando bajo el supuesto que deseamos estimar un modelo de regresión múltiple de tipo predeterminado al cual hemos podido llegar por conocimiento del área de aplicación (conocimiento o suposición de relaciones funcionales entre las variables o por previa experimentación). El caso es diferente cuando la experimentación se realiza con fines exploratorios a fin de

determinar el orden de la superficie a estimar y donde los puntos centrales y otros puntos interiores de la región experimental sí cumplen un cometido: el de ofrecer información acerca de la curvatura de la superficie y por ende de la necesidad de incluir términos cuadráticos y de productos cruzados de las variables.

## 6. Conclusiones

1) El uso de las propiedades de la matriz  $H$  permite una mejor comprensión de los criterios de  $V$ - y  $G$ - optimalidad.

2) La  $V$ -optimalidad nos indica que el número de observaciones debe ser lo más grande posible en relación al número de parámetros pero no aporta nada en la ubicación de los experimentos para el caso del modelo de regresión.

3) La  $G$ -optimalidad busca un balance en la ubicación de los experimentos pero no se preocupa de la cobertura de la región experimental por parte de CVI.

4) Los diseños que asignan una influencia potencial igual a todos los puntos del diseño (lo que hemos definido como condición- $h$ ) gozan de las siguientes propiedades:

i) son  $G$ -óptimos

ii) la varianza de  $y_i$  es  $p \sigma/n$ , independientemente de los valores de las  $x_j$  y por tanto podemos conocer la reducción en la varianza del estimador de la variable respuesta al incrementar el número de observaciones, independientemente de la ubicación de la región experimental.

5) En el caso de los diseños exactos, para que los criterios de  $D$ -optimalidad y  $G$ -optimalidad puedan ser satisfechos simultáneamente hay que imponer una restricción en cuanto a los valores que puede tomar  $n$  (número de experimentos) y a la forma de región experimental.

6) Cuando se van a realizar experimentos con miras a estimar los parámetros de un modelo previamente elegido y se desea que el diseño sea  $G$ -óptimo, queda descar-

tada la inclusión de experimentos en el centro de la CVI.

*Este artículo es consecuencia del trabajo realizado por el autor en la atención de una consulta para lograr un diseño G-óptimo para ciertos experimentos a realizarse dentro del ámbito de la Ingeniería electrónica, durante su permanencia en la Clínica Matemática del Claremont Graduate School de California. El autor agradece a la Clínica Matemática de C.G.S. y al CONCYTEC, que hicieron posible su visita a dicha institución.*

### **Bibliografía.**

- [1] Atkinson, A.C. (1988) Recent Developments in the Methods of Optimun and Related Experimental Designs. *International Statistical Review* Vol 56, #2 pág. 99-115.
- [2] John, P. (1972) *Experimental Design*. John Wiley.

- [3] Hoaglin, D.C. and Welsch, R.E. (1978)  
The Hat Matrix in Regression and ANOVA  
*The American Statistician* Vol 32 #1.
- [4] Kiefer, J. and Wolfowitz, J. (1960)  
The Equivalence of Two Extremum Problems.  
*Canadian Journal of Mathematics*.  
Vol 12.
- [5] Myers, R. (1986) Modern and Classical  
Regression Analysis with applications.  
Duxbury Press.
- [6] Seier, E. & Romberger, A. (1990) Opti-  
mality Criteria and Exact Experimental  
Designs for Regression Models. (*Clinic  
Report*) *The Mathematics Clinic, Clare-  
mont Graduate School, California*.

## **IRRACIONALIDAD Y TRASCENDENCIA DE $e$ Y $\pi$**

**José TOLA\***

*Las demostraciones de la trascendencia de los números  $e$  y  $\pi$  constituyen hechos importantes de la historia de las Matemáticas que se deben a Charles Hermite (1872) y Ferdinand Lindemann (1882) respectivamente. En particular, la demostración de Lindemann dejó definitivamente establecido que es imposible construir un cuadrado cuya área sea exactamente igual a la de un círculo dado. Quedó así resuelto en sentido negativo el clásico problema de los geómetras griegos que se conoce con el nombre de **cuadratura del círculo**. En la creencia de que esas demostraciones serán de interés para los lectores de*

---

\*

Profesor Principal de la PUCP

*Pro-Mathematica se exponen en este artículo las demostraciones de la trascendencia de  $e$  y  $\pi$  en la forma, más sencilla que las originales, que se deben a David Hilbert . Para mayor ilustración se dan, además, las pruebas, considerablemente más simples, de la irracionalidad de los mencionados números.*

## § 1. Introducción

Es un hecho bien conocido que  $e$  y  $\pi$  son números trascendentes, es decir que no son raíces de ninguna ecuación algebraica

$$a_0 + a_1x + \dots + a_nx^n = 0$$

en que los coeficientes  $a_0, a_1, \dots, a_n$  son números enteros. Las demostraciones han sido muy difundidas y no es difícil hallarlas en la literatura matemática, particularmente en otros idiomas. No obstante, considero de interés darla a conocer mejor en nuestro medio.



En parte, la importancia de la trascendencia de  $\pi$  está en el hecho, importante en la historia de las Matemáticas, de que puso punto final, después de muchos siglos, a los innumerables y estériles intentos de llevar a cabo, con sólo la ayuda de la regla y el compás, la construcción de un cuadrado cuya área fuera *exactamente* igual a la de un círculo de radio dado, problema que, con el nombre de *cuadratura del círculo*, se remonta a la antigüedad clásica.

El problema de la cuadratura del círculo se reduce, dado un segmento cuya longitud se toma por unidad, a construir con sólo la ayuda de la regla y el compás otro segmento cuya longitud sea exactamente igual a  $\pi$ . Desde luego, es preciso entender claramente que esta construcción debe concebirse en un sentido ideal, en que los medios empleados sean absolutamente perfectos y en que el resultado no esté afectado por error alguno.

Ahora bien, dado el segmento de longitud unidad, todos los números *construibles*

con la regla y el compás son aquellos y sólo aquellos que pertenecen a una extensión  $\mathbb{F}_n$  del campo  $\mathbb{F}_0$  de los números racionales, tal que existe una sucesión de campos  $\mathbb{F}_0, \mathbb{F}_1, \dots, \mathbb{F}_n$  que tiene la propiedad de que cada campo  $\mathbb{F}_i$  ( $i = 1, \dots, n$ ) resulta del precedente  $\mathbb{F}_{i-1}$  por adjunción a éste de un número  $\sqrt[k]{k}$ , donde  $k$  pertenece al campo  $\mathbb{F}_{i-1}$  pero  $\sqrt[k]{k}$  no; es decir que  $\mathbb{F}_i$  está formado por todos los números de la forma  $a + b \sqrt[k]{k}$  donde  $a$  y  $b$  pertenecen a  $\mathbb{F}_{i-1}$ .

Cada número construible es *algebraico*, es decir que es raíz de una ecuación algebraica con coeficientes enteros. <sup>(1)</sup>

Por tanto, para comprobar que  $\pi$  no es construible con la regla y el compás basta probar que no es un número algebraico, es decir que es *trascendente*.

La historia de la demostración de la trascendencia de  $\pi$  está estrechamente li-

---

(1) Richard Courant y Herbert Robbins, *What is Mathematics?*, Oxford University Press (1941), pág. 133.

gada a la de la trascendencia de  $e$ , la base de los logaritmos neperianos :

*Johann Heinrich Lambert* (1728-1777) demostró en 1764 que  $\pi$  es un número irracional, con lo cual quedó demostrado que el problema de la cuadratura del círculo no puede ser resuelto con sólo el auxilio de la regla.

En 1872 *Charles Hermite* (1822-1905) [2] logró demostrar que  $e$  es trascendente, a cuyo fin se valió de un polinomio obtenido con ayuda de reducidas de fracciones continuas que había utilizado Lambert. Finalmente, en 1882, *Ferdinand Lindemann* (1852-1939) [2] demostró, mediante una modificación de la demostración de Hermite, que  $\pi$  es también trascendente. Posteriormente, varios matemáticos, y entre ellos los más grandes de su tiempo como *Karl Weierstrass* (1815-1897) y *David Hilbert* (1862-1943), dieron variantes y modificaciones de las demostraciones originales.

*Las demostraciones de la trascendencia de  $e$  y  $\pi$  que se dan aquí se deben a David*

Hilbert [3]. Naturalmente demuestran al mismo tiempo que esos números son irracionales. No obstante, para mayor ilustración y por ser muy sencillo, probaremos previamente la irracionalidad de  $e$  y  $\pi$ , aún cuando eso no es requerido para la demostración de su trascendencia.

## I. IRRACIONALIDAD Y TRASCENDENCIA DE $e$ .

### § 2. Irracionalidad de $e$

$e$  es igual a la suma de la serie  $\sum_{n=0}^{\infty} \frac{1}{n!}$ , en donde, en particular,  $0! = 1$  por definición.

Si  $e$  fuera racional, podríamos suponer que  $e = \frac{p}{q}$  donde  $p$  y  $q$  son enteros primos entre sí. Demostraremos que esta suposición conduce a una contradicción. En efecto, de ella se deduce que

$$0 < \frac{p}{q} - \left( 1 + \frac{1}{1!} + \frac{1}{2!} + \dots + \frac{1}{q!} \right) = \sum_{n=q+1}^{\infty} \frac{1}{n!} \quad (1)$$

Ahora bien, podemos escribir

$$\begin{aligned} \sum_{n=q+1}^{\infty} \frac{1}{n!} &= \frac{1}{q!} \left( \frac{1}{q+1} + \frac{1}{(q+1)(q+2)} + \dots \right) \\ &< \frac{1}{q!} \left( \frac{1}{q+1} + \frac{1}{(q+1)^2} + \dots \right) = \\ &= \frac{1}{q!} \frac{\frac{1}{q+1}}{1 - \frac{1}{q+1}} = \frac{1}{q!} \frac{1}{q} . \end{aligned}$$

Luego, de (1) se deduce que

$$0 < q! \left( \frac{p}{q} - 2 - \frac{1}{2!} - \dots - \frac{1}{q!} \right) < \frac{1}{q} .$$

Esta desigualdad es imposible pues en tanto el término de enmedio es un número entero positivo,  $\frac{1}{q}$  es menor que 1. Por consiguiente  $e$  es irracional.

### § 3. Trascendencia de $e$ .

Para demostrar que  $e$  es un número trascendente es necesario probar que no es raíz de una ecuación

$$a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n = 0$$

en que  $a_0 \neq 0$  y  $a_0, a_1, a_1, \dots, a_n$  son números *enteros*. Con tal fin bastará entonces con demostrar que una relación de la forma

$$a_0 + a_1 e + a_2 e^2 + \dots + a_n e^n = 0 \quad (1)$$

conduce a una contradicción.

La idea de la demostración es la siguiente: determinaremos los números *enteros*  $M, M_1, M_2, \dots, M_n$  tales que se cumplan las relaciones

$$e = \frac{M_1 + \varepsilon_1}{M}, \quad e^2 = \frac{M_2 + \varepsilon_2}{M}, \quad \dots, \quad (2)$$

$$e^n = \frac{M_n + \varepsilon_n}{M},$$

donde  $\varepsilon_1/M, \varepsilon_2/M, \dots, \varepsilon_n/M$  son números positivos muy pequeños. Valiéndonos de las expresiones (2) de las potencias de  $e$  la relación (1) que suponemos cumplida toma la forma

$$[a_0 M + a_1 M_1 + a_2 M_2 + \dots + a_n M_n] + \quad (3)$$

$$[a_1 \varepsilon_1 + a_2 \varepsilon_2 + \dots + a_n \varepsilon_n] = 0.$$

La primera expresión entre corchetes es un entero. En el curso de la demostración probaremos que no es nulo. En lo que respecta a la expresión encerrada en el segundo par de corchetes, podremos establecer que puede hacerse menor que la unidad.

Resultará así una contradicción porque es imposible que un entero aumentado en un número menor que 1 resulte igual a cero. De esa manera quedará demostrado que la relación (1) es imposible, y por tanto, que  $e$  es trascendente.

En el curso de la demostración que sigue haremos uso del teorema según el cual si un entero no es divisible por un cierto número no puede ser igual a cero, por cuanto cero es divisible por cualquier número dado. Mostraremos entonces que los números  $M_1, \dots, M_n$  son divisibles por un cierto número primo  $p$ , pero que  $a_0^M$  no es divisible por dicho número. Resultará así que  $a_0^M + a_1^M + \dots + a_n^M$  no es divisible por  $p$  y por tanto es diferente de cero.

La ayuda más importante que tendremos es la de la *integral de Hermite*

$$M = \int_0^{\infty} \frac{z^{p-1} [(z-1)(z-2)\dots(z-n)]^p e^{-z}}{(p-1)!} dz, \quad (4)$$

donde  $n$  es el grado de la ecuación (1) y  $p$  es un número *impar primo* que después vamos a determinar. Si el intervalo de integración lo dividimos en dos partes por el número  $\nu$ , podremos definir los números  $M_\nu$  y  $\epsilon_\nu$ , ( $\nu = 1, \dots, n$ ) mediante las fórmulas

$$M_\nu = e^\nu \int_\nu^{\infty} \frac{z^{p-1} [(z-1)(z-2)\dots(z-n)]^p e^{-z}}{(p-1)!} dz, \quad (5)$$

y

$$\epsilon_\nu = e^\nu \int_0^\nu \frac{z^{p-1} [(z-1)(z-2)\dots(z-n)]^p e^{-z}}{(p-1)!} dz, \quad (6)$$

De las fórmulas (4), (5) y (6) se deduce la ecuación

$$e^\nu = \frac{M_\nu + \epsilon_\nu}{M}$$



que para  $\nu = 1, 2, \dots, n$ , da las relaciones (2).

Podemos entrar ahora a detallar la demostración.

1. Comencemos por recordar la fórmula mediante la cual se define la función gama :

$$\Gamma(\rho) = \int_0^{\infty} z^{\rho-1} e^{-z} dz. \quad (1)$$

Esta fórmula la vamos a necesitar únicamente para valores enteros y positivos de  $\rho$ . Esto supuesto, y que  $\rho > 1$ , se obtiene integrando por partes

$$\begin{aligned} \Gamma(\rho) &= \int_0^{\infty} z^{\rho-1} e^{-z} dz = [-z^{\rho-1} e^{-z}]_0^{\infty} + \\ &+ \int_0^{\infty} (\rho-1) z^{\rho-2} e^{-z} dz = \end{aligned}$$

---

(1) Ver, por ejemplo, Tola, Análisis II, Lima (1971), pág. 235.

$$\begin{aligned}
&= (\rho-1) \int_0^{\infty} z^{\rho-2} e^{-z} dz = (\rho-1)\Gamma(\rho-1) = \\
&= (\rho-1)(\rho-2) \Gamma(\rho-2) = \\
&= (\rho-1)(\rho-2)(\rho-3) \dots 2 \cdot 1 \Gamma(1) .
\end{aligned}$$

y puesto que

$$\Gamma(1) = \int_0^{\infty} e^{-z} dz = 1 .$$

resulta

$$\Gamma(\rho) = \int_0^{\infty} z^{\rho-1} e^{-z} dz = (\rho-1)! \quad (7)$$

Por consiguiente  $\Gamma(\rho)$  aumenta rápidamente cuando el entero  $\rho$  crece.

2. La fórmula que acabamos de obtener nos permitirá calcular con facilidad la integral de Hermite. Con ese fin nos valdremos de la fórmula para elevar a una potencia

un polinomio <sup>(1)</sup>, y tendremos el siguiente polinomio en potencias de  $z$  :

$$\begin{aligned} [(z-1)(z-2)\dots(z-n)]^p &= [z^n + \dots + (-1)^n n!]^p = \\ &= z^{np} + \dots + (-1)^n (n!)^p, \end{aligned} \quad (8)$$

en donde hemos escrito únicamente los términos en la mayor y en la menor potencia de  $z$ , y hemos tenido en cuenta nuestra suposición de que  $p$  es un número impar, y por tanto se tiene que  $(-1)^{np} = [(-1)^p]^n = (-1)^n$ . Se sabe además que todos los coeficientes del polinomio obtenido en (8) son enteros. La integral de Hermite es entonces dada por

(1) La fórmula para elevar a potencia un polinomio es

$$\begin{aligned} &(a_1 + a_2 + \dots + a_p)^m \\ &= \sum \frac{m!}{\alpha_1! \alpha_2! \dots \alpha_p!} a_1^{\alpha_1} a_2^{\alpha_2} \dots a_p^{\alpha_p}. \end{aligned}$$

donde la suma se extiende a todos los conjuntos  $\{\alpha_1, \alpha_2, \dots, \alpha_p\}$  de números enteros positivos o nulos tales que  $\alpha_1 + \alpha_2 + \dots + \alpha_p = m$ , y los coeficientes son números enteros. (Weber, Lehrbuch der Algebra, I, pág. 53).

$$M = \frac{(-1)^n (n!)^p}{(p-1)!} \int_0^\infty z^{p-1} e^{-z} dz +$$

$$\sum_{\rho=p+1}^{np+p} \frac{C_\rho}{(p-1)!} \int_0^\infty z^{\rho-1} e^{-z} dz,$$

donde los  $C_\rho$  son constantes enteras. Las integrales de la ecuación precedente pueden ser calculadas mediante la fórmula (7) y se obtiene

$$M = (-1)^n (n!)^p + \sum_{\rho=p+1}^{np+p} C_\rho \frac{(\rho-1)!}{(p-1)!}.$$

El índice de sumación  $\rho$  es siempre mayor que  $p$ . Por tanto la expresión  $(\rho-1)!/(p-1)!$  es un número entero que tiene a  $p$  como factor, de manera que puede sacarse en factor común a  $p$  en la sumatoria y resulta

$$M = (-1)^n (n!)^p + p[C_{p+1} + C_{p+2}(p+1) +$$

$$C_{p+3}(p+2)\dots].$$

El número  $M$  no será divisible por  $p$  si no lo es el primer término del segundo

miembro para lo cual, dado que  $p$  es primo, basta que sea  $p > n$ , lo que es posible porque los números primos son infinitos. Podemos suponer por tanto que  $(-1)^n (n!)^p$ , y por consiguiente  $M$ , no es divisible por el número primo  $p$ . También podemos escoger a  $p$  de manera que el término  $a_0$  de (1), que es diferente de cero, no sea divisible por  $p$ , para lo cual basta que sea  $p > |a_0|$ , lo que es siempre posible. De la manera dicha queda probado que  $a_0 M$  no es divisible por  $p$ , que es uno de los hechos que nos proponíamos establecer.

3. Vamos a probar ahora que los números  $M_\nu$  definidos por (5) son todos divisibles por  $p$ . Si en dicha ecuación introducimos el factor  $e^\nu$  dentro del signo integral y hacemos el cambio de la variable  $z$  por la variable  $\zeta = z - \nu$ , que varía entre 0 e  $\infty$  cuando varía entre  $\nu$  e  $\infty$ , se obtiene

$$M_\nu = \int_0^\infty \frac{1}{(p-1)} \left[ (\zeta + \nu)^{p-1} [(\zeta + \nu - 1)(\zeta + \nu - 2) \dots \dots (\zeta + \nu - n)]^p e^{-\zeta} \right] d\zeta$$

Si se llevan a cabo las multiplicaciones del numerador se obtiene un polinomio en  $\zeta$  en que el término de menor grado es en  $\zeta^p$  y el de mayor grado es en  $\zeta^{(n+1)p-1}$ , y cuyos coeficientes son números enteros. La integral del numerador será entonces una combinación lineal de las integrales

$$\int_0^{\infty} \zeta^p e^{-\zeta} d\zeta, \int_0^{\infty} \zeta^{p+1} e^{-\zeta} d\zeta, \dots, \int_0^{\infty} \zeta^{(n+1)p-1} e^{-\zeta} d\zeta,$$

con coeficientes enteros. Puesto que dichas integrales, según (7), son iguales a  $p!$ ,  $(p+1)!$ , ...,  $[(n+1)p-1]!$ , la integral del numerador será igual a  $p!$  multiplicado por un número entero  $A_\nu$ , de manera que

$$M_\nu = \frac{p! A_\nu}{(p-1)!} = p A_\nu. \quad (\nu = 1, 2, \dots, n),$$

lo cual prueba, como ya habíamos anunciado que  $M_1, \dots, M_n$  son números enteros divisibles por  $p$ . Este resultado y el hecho ya probado de que  $a_0^M$  no es divisible por  $p$ , permite deducir que  $a_0^M + a_1^M + \dots + a_n^M$  no es divisible por  $p$  y por tanto es diferente de cero.

4. Vamos a tratar ahora acerca de una nueva parte de la demostración que se refiere a la suma  $a_1 \varepsilon_1 + \dots + a_n \varepsilon_n$  donde los  $\varepsilon_\nu$  son dados por las fórmulas (6). Vamos a probar que estos números pueden suponerse arbitrariamente pequeños. Para ese fin haremos uso del hecho de que  $p$  puede ser elegido tan grande como se quiera, pues las condiciones que hasta ahora hemos impuesto a  $p$  son la de ser un número primo mayor que  $n$  y que  $|a_0|$  y que, por lo demás, puede ser arbitrariamente grande.

Designemos por  $G$  y  $g_\nu$  a los máximos de los valores absolutos de  $z(z-1)\dots(z-n)$  y  $(z-1)(z-2)\dots(z-n)e^{-z+\nu}$ , respectivamente, para  $z$  en el intervalo  $(0, n)$ , es decir que

$$\left. \begin{array}{l} |z(z-1)\dots(z-n)| \leq G \\ |(z-1)(z-2)\dots(z-n)e^{-z+\nu}| \leq g_\nu \end{array} \right\} \text{ para } 0 \leq z \leq n .$$

Puesto que  $\nu \leq n$ , se tiene para  $\nu = 1, \dots, n$ ,

$$|\epsilon_\nu| \leq \int_0^\nu \left| \frac{z^{p-1} [(z-1)(z-2)\dots(z-n)]^p e^{-z+\nu}}{(p-1)!} \right| dz$$

$$\leq \int_0^\nu \frac{G^{p-1} g_\nu}{(p-1)!} dz ,$$

o sea

$$|\epsilon_\nu| \leq \frac{G^{p-1} g_\nu \cdot \nu}{(p-1)!} \quad (9)$$

Ahora bien,  $G$ ,  $g_\nu$  y  $\nu$  son números constantes independientes de  $p$ , y es claro que la fracción  $G^{p-1}/(p-1)!$  es tan pequeña como se quiera si  $p$  se elige suficientemente grande.

Luego,

$$|a_1 \epsilon_1 + a_2 \epsilon_2 + \dots + a_n \epsilon_n| \leq |a_1| |\epsilon_1| + |a_2| |\epsilon_2| +$$

$$\dots + |a_n| |\epsilon_n| \leq (|a_1| g_1 + |a_2| 2g_2 +$$

$$\dots + |a_n| n g_n) \frac{G^{p-1}}{(p-1)!} .$$

Puesto que el valor de la expresión entre paréntesis es independiente de  $p$ , re-



sulta de aquí que si  $p$  es suficientemente grande la suma  $|a_1 \varepsilon_1 + \dots + a_n \varepsilon_n|$  es tan pequeña como se quiera y, en particular, menor que la unidad.

Con esto hemos dado término a la demostración, que fue descrita al comienzo, de que una relación de la forma (1) en que  $a_0$  es diferente de cero y  $a_1, \dots, a_n$  son enteros, conduce a una contradicción, y por consiguiente el número  $e$  no es algebraico. Es por tanto trascendente.

## II. IRRACIONALIDAD Y TRASCENDENCIA DE $\pi$

### § 4. Irracionalidad de $\pi$

En la demostración que sigue<sup>(1)</sup> desempeña un papel importante la función

$$f(x) = \frac{x^n(1-x)^n}{n!} \quad (1)$$

donde  $n$  es un número entero.

---

(1) Niven, Bull. of the Am. Math. Soc., 53 (1947) pág.209).

Designemos por  $D^k f$  a la  $k$ -ésima derivada de  $f$ . Para calcularla podemos emplear la conocida fórmula de *Leibnitz* de la derivada de un producto de dos funciones<sup>(1)</sup>:

$$D^k(gh) = \sum_{j=0}^k \binom{k}{j} D^j g \cdot D^{k-j} h.$$

donde  $D^0 g = g$  y  $D^0 h = h$ .

Se tiene por tanto

$$D^k f = \frac{1}{n!} \sum_{j=0}^k \binom{k}{j} D^j g D^{k-j} h$$

donde  $g(x) = x^n$  y  $h(x) = (1-x)^n$ . Por consiguiente, puesto que

$$D^j g(0) = \begin{cases} 0 & \text{si } j < n \\ n! & \text{si } j = n \\ 0 & \text{si } j > n \end{cases} \quad \text{y}$$

$$D^1 h(0) = (-1)^1 \frac{n!}{(n-1)!},$$

---

(1) R. Courant, *Differential and Integral Calculus*, Blackie and Son Limited, vol. I, (1942) pág. 202.

se tiene

$$D^k f(0) = 0 \quad \text{si} \quad k < n$$

y

$$D^k f(0) = \frac{1}{n!} \binom{k}{n} n! D^{k-n} h(0) \quad \text{si} \quad k \geq n. \quad (2)$$

Además, se deduce de (1) que  $f(1-x) = f(x)$ , y por tanto, para  $k = 0, 1, 2, \dots$  es  $D^k f(1-x) = (-1)^k D^k f(x)$ . En particular

$$D^k f(1) = (-1)^k D^k f(0). \quad (3)$$

Puesto que tanto los coeficientes binomiales como  $D^{k-n} h(0)$  son enteros, resulta de (2) y (3) que  $D^k f(0)$  y  $D^k f(1)$ , para  $k = 0, 1, 2, \dots$ , son números enteros.

2. Es claro que si probamos que  $\pi^2$  no es racional tampoco lo será  $\pi$ . Demostraremos pues que si se supone que

$$\pi^2 = \frac{p}{q} \quad (4)$$

donde  $p/q$  es una fracción irreducible se llega a una contradicción. Con este fin nos valdremos de la función  $f$  antes considerada, y supondremos que se cumple (4).

Sea la función

$$F(x) = q^n \{ \pi^{2n} f(x) - \pi^{2n-2} D^2 f(x) + \\ + \pi^{2n-4} D^4 f(x) - \dots + \\ \pi^2 (-1)^{n-1} D^{2n-2} f(x) + (-1)^n D^{2n} f(x) \}. \quad (5)$$

En virtud de lo demostrado en la parte 1 y de (4),

$$F(0) \text{ y } F(1) \text{ son números enteros} \quad (6)$$

Teniendo en cuenta que, por ser  $f(x)$  un polinomio de grado  $2n$ , sus derivadas de órdenes mayores que  $2n$  son nulas, podemos obtener las derivadas primera y segunda de  $F(x)$  en la forma

$$F'(x) = q^n \{ \pi^{2n} Df(x) - \pi^{2n-2} D^3 f(x) + \pi^{2n-4} D^5 f(x) - \\ - \dots + \pi^2 (-1)^{n-1} D^{2n-1} f(x) \}$$

$$F''(x) = q^n \{ \pi^{2n} D^2 f(x) - \pi^{2n-2} D^4 f(x) + \\ + \pi^{2n-4} D^6 f(x) - \dots + \pi^2 (-1)^{n-1} D^{2n} f(x) \};$$

de donde se deduce haciendo uso de (5),

$$F''(x) + \pi^2 F(x) = q^n \pi^{2n+2} f(x) = p^n \pi^2 f(x) \quad (7)$$

Ahora bien, puede observarse que

$$\begin{aligned}
 (F'(x)\operatorname{sen} \pi x - \pi F(x)\operatorname{cos} \pi x)' &= \pi F'(x)\operatorname{cos} \pi x + \\
 F''(x)\operatorname{sen} \pi x + \pi^2 F(x)\operatorname{sen} \pi x - \pi F'(x)\operatorname{cos} \pi x &= \\
 F''(x)\operatorname{sen} \pi x + \pi^2 F(x)\operatorname{sen} \pi x .
 \end{aligned}$$

La ecuación (7) nos permite escribir entonces

$$(F'(x)\operatorname{sen} \pi x - \pi F(x)\operatorname{cos} \pi x)' = p^n \pi^2 f(x)\operatorname{sen} \pi x \quad (8)$$

Se deduce de aquí que

$$\begin{aligned}
 p^n \pi^2 \int_0^1 f(x)\operatorname{sen} \pi x . dx &= (F'(x)\operatorname{sen} \pi x - \\
 - \pi F(x)\operatorname{cos} \pi x) \Big|_0^1 &= \pi F(1) + \pi F(0),
 \end{aligned}$$

o sea

$$\pi p^n \int_0^1 f(x)\operatorname{sen} \pi x . dx = F(1) + F(0). \quad (9)$$

Puesto que, según (6), el segundo miembro es entero, también lo será el primer miembro cualquiera que sea el número  $n$ . Bastará entonces con probar que  $n$  puede ser elegido de modo que el primer miembro de (9) sea un número *positivo* estrictamente

menor que 1, para que (9) sea una relación absurda, y por consiguiente la fórmula (4) implique una contradicción. Lograremos ese propósito si suponemos que  $n$  es tal que  $\frac{2p^n}{n!} < 1$ , lo que es posible porque  $\lim_{n \rightarrow \infty} \frac{p^n}{n!} = 0$

En efecto, debido a que  $x(1-x) = x-x^2$  es positivo y menor que 1 para  $0 < x < 1$ , en este intervalo se cumple que

$$0 < f(x) < \frac{1}{n!}$$

y por consiguiente

$$\begin{aligned} 0 < \pi p^n \int_0^1 f(x) \operatorname{sen} \pi x \cdot dx < \pi \frac{p^n}{n!} \int_0^1 \operatorname{sen} \pi x \cdot dx \\ &= \frac{2p^n}{n!} < 1 ; \end{aligned}$$

es decir que el primer miembro de (9) es positivo y menor que 1 y por tanto, como hemos señalado, esa ecuación es absurda y la suposición (4) implica una contradicción.  $\pi$  es por consiguiente un número irracional.

## § 5. Trascendencia de $\pi$

### 5.1 Definiciones y proposiciones preliminares

La demostración de la trascendencia del número  $\pi$  que vamos a exponer aquí se debe a *Hilbert* como ya hemos señalado, y requiere recordar algunas definiciones y proposiciones algebraicas que damos a continuación.

Se llama *función racional* de las variables  $x_1, x_2, \dots, x_n$  a una función que es definida por un cociente de la forma

$$f(x_1, x_2, \dots, x_n) = \frac{\sum_{\alpha} c_{\alpha} x_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n}}{\sum_{\alpha} c'_{\alpha} x_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n}},$$

donde las sumas se extienden a todos los conjuntos

$$\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$$

de  $n$  números enteros positivos o nulos, y en que los coeficientes  $c_{\alpha}$  y  $c'_{\alpha}$  son núme-

ros complejos, un número finito de los cuales, a lo sumo, es diferente de cero; pero sin que el denominador se reduzca a cero. Si la expresión que define a  $f$  se reduce al numerador

$$f(x_1, \dots, x_n) = \sum_{\alpha} c_{\alpha} x_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n},$$

se llama *función racional entera* de  $x_1, x_2, \dots, x_n$  o *polinomio* en  $x_1, x_2, \dots, x_n$ . De lo contrario se llama *función racional fraccionaria*.

Se dice que la función  $f(x_1, \dots, x_n)$  es *simétrica* si su valor no cambia cuando se realiza una permutación cualquiera de los valores que se atribuya a las variables, es decir si se cumple que

$$f(a_1, \dots, a_n) = f(a_{\sigma(1)}, \dots, a_{\sigma(n)}),$$

cualesquiera que sean los números  $a_1, \dots, a_n$  del dominio en que la función es definida y la permutación  $\sigma$  de los subíndices  $1, 2, \dots, n$ . Las funciones racionales enteras que son simétricas se denominan en



forma abreviada *polinomios simétricos* en las variables  $x_1, \dots, x_n$ .

Atención especial merecen los polinomios evidentemente simétricos

$$s_1 = x_1 + x_2 + \dots + x_n$$

$$s_2 = x_1x_2 + x_1x_3 + \dots + x_1x_n + x_2x_3 + \dots + x_{n-1}x_n$$

$$s_3 = x_1x_2x_3 + x_1x_2x_4 + \dots + x_{n-2}x_{n-1}x_n$$

.....

$$s_n = x_1x_2 \dots x_n, \quad (1)$$

los cuales se llaman *funciones simétricas elementales* de las variables  $x_1, \dots, x_n$ .

Consideremos ahora la ecuación algebraica de grado  $n$

$$a_0 + a_1x + \dots + a_{n-1}x^{n-1} + x^n = 0, \quad (1)$$

cuyos coeficientes son números complejos, y en que el coeficiente de  $x^n$  es 1, razón

---

(1) Observar que cada término  $x_{j_1}x_{j_2} \dots x_{j_p}$  de  $s_p$  es tal que  $j_1 < j_2 < \dots < j_p$

por la cual diremos que la ecuación es *mónica*. Si  $\alpha_1, \alpha_2, \dots, \alpha_n$  son las raíces de la ecuación, es sabido que se cumple la identidad

$$a_0 + a_1 x + \dots + a_{n-1} x^{n-1} + x^n = (x - \alpha_1)(x - \alpha_2) \dots (x - \alpha_n),$$

de donde se deduce que

$$-a_{n-1} = \alpha_1 + \alpha_2 + \dots + \alpha_n$$

$$a_{n-2} = \alpha_1 \alpha_2 + \alpha_1 \alpha_3 + \dots + \alpha_1 \alpha_n + \alpha_2 \alpha_3 + \dots + \alpha_{n-1} \alpha_n$$

$$-a_{n-3} = \alpha_1 \alpha_2 \alpha_3 + \alpha_1 \alpha_2 \alpha_4 + \dots + \alpha_{n-2} \alpha_{n-1} \alpha_n$$

.....

$$(-1)^n a_0 = \alpha_1 \alpha_2 \dots \alpha_n .$$

Puede reconocerse de inmediato que los segundos miembros de las ecuaciones anteriores son la funciones simétricas elementales de las raíces de la ecuación y se cumple por tanto que *las funciones simétricas elementales  $s_1, s_2, \dots, s_n$  de las raíces de una ecuación mónica son iguales a los coeficientes de la ecuación con sus propios signos o con signos contrarios, de acuerdo*

con la fórmula

$$s_p = (-1)^p a_{n-p}, \quad (p = 1, 2, \dots, n).$$

La siguiente proposición, que luego vamos a utilizar, es consecuencia del teorema fundamental de la teoría de las funciones simétricas<sup>(1)</sup>

### Teorema 1.

a) Cada polinomio simétrico  $P$  de las variables  $x_1, x_2, \dots, x_n$  es igual a un polinomio  $Q$  en las funciones simétricas elementales  $s_1, \dots, s_n$  de dichas variables y en los coeficientes de  $P$ ; y los coeficientes de  $Q$  son números enteros.

b) Si todos los coeficientes de  $P$  son números racionales,  $Q$  es un polinomio en las funciones simétricas elementales de las variables y sus coeficientes son racionales.

c) Si todos los coeficientes de  $P$  son números enteros,  $Q$  es un polinomio en las

---

(1) Ver por ejemplo, Van der Warden, Moderne Algebra I, Springer (1937).

funciones simétricas elementales de las variables y sus coeficientes son *enteros*.

Las observaciones hechas precedentemente acerca de la ecuación mónica (1) da lugar a la siguiente consecuencia del teorema anterior.

### **Teorema 2.**

a) Cada polinomio simétrico  $P$  en las raíces  $\alpha_1, \dots, \alpha_n$  de la ecuación mónica (1) es igual a un polinomio  $Q$  en los coeficientes  $a_0, a_1, \dots, a_{n-1}$  de la ecuación y en los coeficientes de  $P$ ; y los coeficientes de  $Q$  son enteros.

b) Si todos los coeficientes de  $P$  y los de la ecuación (1) son números *racionales*,  $P$  es igual a un número *racional*.

c) Si todos los coeficientes de  $P$  y los de la ecuación (1) son números *enteros*,  $P$  es igual a un número *entero*.

Si la ecuación que se considera no es mónica, y si  $a_n \neq 0$  es el coeficiente de  $x^n$ , lo expresado anteriormente se cumple sus-

tituyendo a los coeficientes  $a_0, a_1, \dots, a_n$  por los números  $\frac{a_0}{a_n}, \frac{a_1}{a_n}, \dots, \frac{a_{n-1}}{a_n}$ .

Por último, recordaremos los siguientes conceptos acerca de los *números algebraicos*<sup>(1)</sup>:

Un número complejo se llama *algebraico* si es una raíz de una ecuación algebraica con coeficientes enteros, o bien, equivalentemente, de una ecuación *mónica* con coeficientes racionales.

Un número algebraico se llama *entero algebraico* si es raíz de una ecuación *mónica* con coeficientes enteros racionales. En adelante mantendremos la denominación de *enteros* para los números enteros racionales:

La suma, la diferencia, el producto y el cociente, de números algebraicos (con divisor diferente de cero en el último ca-

---

(1) L.W.Reid, The Elements of the Theory of Algebraic Numbers, The Macmillan Co. (1910).

so) son números algebraicos; y la suma, la diferencia y el producto de números enteros algebraicos son también enteros algebraicos.

En el anexo se ha reunido una información complementaria acerca de los números algebraicos; en particular, se da la demostración de las proposiciones enunciadas anteriormente.

## 5.2 Demostración de la trascendencia de $\pi$

De la ecuación

$$e^z = e^{x+iy} = e^x(\cos y + i \operatorname{sen} x)$$

mediante la cual se define a la función exponencial de la variable compleja  $z$ , resulta la relación

$$1 + e^{i\pi} = 0 \quad (2)$$

La trascendencia de  $e$ , que ya fue demostrada, asegura que no puede existir una ecuación de la forma

$$a_0 + a_1 e + \dots + a_n e^n = 0$$

donde  $a_0 \neq 0$  y  $a_0, a_1, \dots, a_n$  son números enteros racionales.

*Lindemann* demostró en 1882 que tampoco es posible que se cumpla una relación de la forma

$$a_0 + a_1 e^{b_1} + a_2 e^{b_2} + \dots + a_n e^{b_n} = 0 \quad (3)$$

en que los números  $a_j$  y  $b_j$  ( $j = 0, 1, \dots, n$ ) sean *algebraicos*. Ahora bien, la unidad imaginaria  $i$  es un número algebraico por cuanto es raíz de la ecuación  $1 + x^2 = 0$ . Luego, si  $\pi$  fuera algebraico también lo sería  $i\pi$ , y no podría cumplir la ecuación (2) en virtud del teorema de *Lindemann*. Resulta así que  $\pi$  es trascendente.

Aquí no vamos a demostrar el teorema general de *Lindemann* sino un caso particular. La demostración de *Hilbert* que vamos a exponer es una modificación de la que fue dada para demostrar la trascendencia de  $e$ .

Comenzamos pues suponiendo, contrariamente a lo que queremos probar, que  $\pi$  es

algebraico, es decir que es raíz de una ecuación algebraica con coeficientes *enteros*. Según lo dicho  $i\pi$  satisfaría también a una ecuación de la misma clase, cuyas raíces ( $i\pi$  entre ellas) vamos a designar por  $\alpha_1, \alpha_2, \dots, \alpha_n$ . A causa de (2) se cumple entonces la relación

$$(1+e^{\alpha_1})(1+e^{\alpha_2})\dots(1+e^{\alpha_n}) = 0,$$

que puede escribirse en la forma

$$\left. \begin{aligned} &(1+e^{\alpha_1}+e^{\alpha_2}+\dots+e^{\alpha_n})+(e^{\alpha_1+\alpha_2}+e^{\alpha_1+\alpha_3}+\dots \\ &+e^{\alpha_{n-1}+\alpha_n})+\dots+(e^{\alpha_1+\alpha_2+\alpha_3}+e^{\alpha_1+\alpha_2+\alpha_4}+\dots \\ &+e^{\alpha_{n-2}+\alpha_{n-1}+\alpha_n})+\dots+(e^{\alpha_1+\alpha_2+\dots+\alpha_n})=0 \end{aligned} \right\} (4)$$

Puede ocurrir que algunos de los exponentes de esta ecuación sean nulos. Cada vez que eso ocurra uno de los términos se reducirá a 1. Si los sumamos con el 1 que ya aparece en el primer término resultará un número racional entero positivo  $a_0$ . Los demás exponentes, todos ellos diferentes



de cero, los designamos por  $\beta_1, \beta_2, \dots, \beta_N$ , y por tanto (4) puede escribirse en la forma

$$a_0 + e^{\beta_1} + e^{\beta_2} + \dots + e^{\beta_N} = 0, \quad (a_0 \text{ entero} > 0) \quad (5)$$

Podemos probar que los números  $\beta_1, \beta_2, \dots, \beta_N$  son las raíces de una ecuación algebraica con coeficientes enteros racionales. En efecto, a partir de la ecuación cuyas raíces son  $\alpha_1, \alpha_2, \dots, \alpha_n$  podemos construir otra ecuación con coeficientes enteros cuyas raíces son las sumas de dos términos  $\alpha_1 + \alpha_2, \alpha_1 + \alpha_3, \dots, \alpha_{n-1} + \alpha_n$  <sup>(1)</sup>.

(1) Los coeficientes de la ecuación mónica de grado  $n(n-1)/2$   
 $(x - (\alpha_1 + \alpha_2))(x - (\alpha_1 + \alpha_3)) \dots (x - (\alpha_{n-1} + \alpha_n))$ ,  
 cuyas raíces son los números  $\alpha_i + \alpha_j$ ,  
 donde  $i, j = 1, 2, \dots, n, e i < j$ , son funciones simétricas de los  $\alpha_k$  y por tanto son funciones racionales enteras de los coeficientes del polinomio mónico cuyas raíces son  $\alpha_1, \alpha_2, \dots, \alpha_n$ , y por consiguiente son números racionales. Luego, los números  $\alpha_i + \alpha_j$  son raíces de una ecuación algebraica con coeficientes enteros.

Semejantemente, a partir de la ecuación cuyas raíces son estos últimos números podemos construir otra, también con coeficientes enteros, cuyas raíces son las sumas de tres términos  $\alpha_1+\alpha_2+\alpha_3, \alpha_1+\alpha_2+\alpha_4, \dots, \alpha_{n-2}+\alpha_{n-1}+\alpha_n$  ; y así sucesivamente. Por último,  $\alpha_1+\alpha_2+\dots+\alpha_n$  que es el coeficiente de mayor grado de la ecuación mónica, cuyas raíces son  $\alpha_1, \alpha_2, \dots, \alpha_n$ , es un número racional y satisface por tanto a una ecuación *lineal* de coeficientes enteros. Multiplicando todas las ecuaciones así obtenidas se obtendrá una ecuación de coeficientes enteros algunas de cuyas raíces pueden ser nulas, y cuyas restantes raíces son  $\beta_1, \beta_2, \dots, \beta_N$ . Si la dividimos por la potencia de la incógnita cuyo exponente es igual al número de raíces nulas, resultará una ecuación de grado  $N$  cuyas raíces son los  $N$  números  $\beta_1, \beta_2, \dots, \beta_N$  :

$$b_0 + b_1 z + b_2 z^2 + \dots + b_N z^N = 0, \quad (b_0, b_N \neq 0 ;$$

$$b_i \text{ entero para } i = 0, 1, \dots, N). \quad (6)$$

Resumiendo: si se supone que  $\pi$  es algebraico tendrá que cumplirse la ecuación (5) en donde los exponentes  $\beta_1, \beta_2, \dots, \beta_N$  son raíces de una ecuación (6). Si demostramos que de (5) y (6) resulta una contradicción quedará probado que  $\pi$  es un número trascendente. Podemos decir en consecuencia que para demostrar que  $\pi$  es trascendente basta probar la siguiente proposición que es un caso particular del teorema de Lindemann : Si  $\beta_1, \beta_2, \dots, \beta_N$  son las raíces de una ecuación (6), no es posible que se cumpla la relación (5).

La demostración se hará siguiendo pasos semejantes a los de la demostración de la trascendencia de  $e$ , de la que difiere en aspectos muy importantes. Tal como entonces pudimos obtener expresiones apropiadas de las potencias enteras de  $e$ , ahora las obtendremos para las potencias de  $e$  que figuran en la ecuación (5). Ellas serán

$$e^{\beta_1} = \frac{M_1 + \epsilon_1}{M}, \quad e^{\beta_2} = \frac{M_2 + \epsilon_2}{M}, \quad \dots,$$

$$e^{\beta_N} = \frac{M_N + \epsilon_N}{M}, \quad (7)$$

en donde, como probaremos luego,  $M$  será un entero y los números  $\varepsilon_1, \dots, \varepsilon_N$ , que ahora podrán ser complejos, serán tan pequeños como se quiera en valor absoluto; pero en que, a diferencia de lo que tuvo lugar en el caso de  $e$ , la suma  $\sum_{\nu} M_{\nu}$ , y no cada uno de los números  $M_{\nu}$ , será un entero divisible por un cierto número primo  $p^{(1)}$ .

Sustituyendo en (5) los valores (7) se obtiene

$$[a_0 M + M_1 + M_2 + \dots + M_N] + [\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_N] = 0 \quad (8)$$

En el curso de la demostración probaremos que  $a_0 M$ , y por tanto la primera expresión entre corchetes, es un entero no divisible por un número primo  $p$ , y es por consiguiente un entero diferente de cero. Probaremos también que basta elegir al número primo  $p$  suficientemente grande, para que la suma  $\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_N$  se haga tan pequeña

---

(1) Como observaremos mas tarde, los números  $M_{\nu}$  serán ahora enteros algebraicos, no necesariamente enteros (racionales).

como se quiera. Resultará así que (9) es una relación imposible, lo que probará que  $\pi$  es trascendente.

Procederemos ahora a exponer los detalles de la demostración que acabamos de describir.

1. Para definir al número  $M$  utilizaremos una adecuada modificación de la integral de *Hermite* que fue utilizada para demostrar la trascendencia de  $e$ .  $M$  será dado por

$$M = \int_0^{\infty} \frac{1}{(p-1)!} z^{p-1} \left[ (z-\beta_1)(z-\beta_2)\dots \dots (z-\beta_N) \right]^p e^{-z} \cdot b_N^{(N-1)p} dz,$$

donde  $b_N$  es el coeficiente de  $z^n$  en (6), y  $p$  es un cierto número primo que luego determinaremos. Puesto que  $\beta_1, \dots, \beta_N$  son las raíces de la ecuación (6), puede escribirse

$$(z-\beta_1)(z-\beta_2)\dots(z-\beta_N) = \frac{1}{b_N} (b_0 + b_1 z + \dots + b_N z^N) \quad (9)$$

y por consiguiente

$$M = \int_0^{\infty} \frac{e^{-z} z^{p-1} dz}{(p-1)!} [b_0 + b_1 z + \dots + b_N z^N]^p \cdot b_N^{(N-1)p-1} \quad (10)$$

2. Desarrollemos ahora el integrando según las potencias de  $z$ . La menor potencia es  $z^{p-1}$  y da origen al término

$$\int_0^{\infty} \frac{e^{-z} z^{p-1} dz}{(p-1)!} b_0^p b_N^{(N-1)p-1} \quad (11)$$

de la expresión de  $M$ .

La fórmula

$$\Gamma(\rho) = \int_0^{\infty} z^{\rho-1} e^{-z} dz = (\rho-1)! \quad (12)$$

que da el valor de la función *gama* para  $\rho$  entero positivo permite escribir el término (11) en la forma

$$b_0^p b_N^{(N-1)p-1}.$$

Los demás términos del integrando de la expresión de  $M$  contienen a  $z$  elevado a la potencia  $p$  o mayores que  $p$ , de manera que sus integrales son de la forma

$$\frac{K_q}{(p-1)!} \int_0^\infty e^{-z} z^{q-1} dz, \quad (q-1 \geq p),$$

donde  $k_q$  es un entero. En virtud de (12) esos términos tienen valores

$$K_q \frac{(q-1)!}{(p-1)!} = K_q \cdot p \dots (q-1), \quad (q-1 \geq p),$$

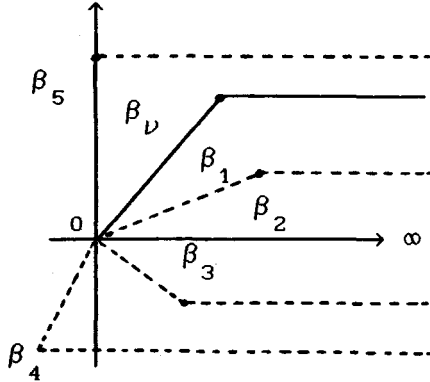
de manera que todos ellos son números enteros divisibles por  $p$ . Resulta así que  $M$  es la suma de  $b_0 b_N^{(N-1)p-1}$  y de un múltiplo de  $p$ . Basta con que elijamos al número primo  $p$  de modo que sea  $p > |b_0|$  y  $p > |b_N|$  para que  $b_0 b_N^{(N-1)p-1}$  no sea divisible por  $p$ , y por tanto, para que  $M$  tampoco sea divisible por  $p$ . Por último, puesto que  $a_0$  es un entero positivo,  $a_0 M$  no será divisible por  $p$  si elegimos a  $p$  con la condición adicional  $p > a_0$ .

3. Ahora vamos a determinar a los números  $M_\nu$  y  $\epsilon_\nu$  de las fórmulas (7). El procedimiento que fue empleado en el caso de  $e$  requiere ahora una importante modificación porque para definir esos números vamos a emplear los números  $\beta_\nu$  como uno de los límites de integración, que pueden ser complejos (de hecho: uno de ellos es  $i\pi$ ). Por consiguiente las integrales mediante las cuales definiremos a  $M_\nu$  y  $\epsilon_\nu$  deberán considerarse sobre el plano complejo, a cuyo fin podemos observar que los integrandos son funciones analíticas uniformes de la variable compleja  $z$ , regulares salvo en el punto  $z = \infty$  en que hay una singularidad esencial. Debemos por tanto fijar convenientemente el camino de integración.

Para cada valor de  $\nu$  la integral que expresa el valor de  $M$  en la fórmula (10) puede ser elegido, como se ve en la figura de la pag. 109, formado por un segmento de recta que parte de  $0$  y llega a  $\beta_\nu$ , y por una semirecta paralela al eje real, que parte de  $\beta_\nu$ , y va a  $\infty$ . Como es sa-



bido, el valor de la integral resulta ser el mismo para todos del número  $\nu$ .



La integral a lo largo del segmento  $0\beta_\nu$  definirá el valor de  $\epsilon_\nu$  que, como luego veremos podrá hacerse arbitrariamente pequeño si se elige al número primo  $p$  suficientemente grande; y la integral a lo largo de la semirrecta  $\beta_\nu^\infty$  proporcionará el valor de  $M_\nu$ . Se tiene por tanto

$$\epsilon_\nu = e^{\beta_\nu} \int_0^{\beta_\nu} \frac{e^{-z} z^{p-1} dz}{(p-1)!} [b_0 + b_1 z + \dots + b_N z^N]^p .$$

$$\cdot b_N^{(N-1)p-1} , \tag{13}$$

$$M = \varepsilon \beta_\nu \int_{\beta_\nu}^{\infty} \frac{e^{-z} z^{p-1} dz}{(p-1)!} [b_0 + b_1 z + \dots + b_N z^N]^p .$$

$$. b_N^{(N-1)p-1} , \quad (14)$$

valores que es fácil observar que satisfacen a las relaciones (7).

4. Trataremos ahora acerca de los  $\varepsilon_\nu$  .

Como se sabe, el valor de la integral de una función de variable compleja es no mayor que el producto del máximo del valor absoluto del integrando sobre el camino de integración por la longitud de éste que en este caso es  $|\beta_\nu|$ .

El integrando puede escribirse en la forma

$$\frac{[z(b_0 + b_1 z + \dots + b_N z^N) b_N^{N-1}]^{p-1}}{(p-1)!} (e^{-z} (b_0 + b_1 z + \dots + b_N z^N) b_N^{N-2}) .$$

Si  $G$  denota al máximo de  $|z(b_0 + b_1 z + \dots + b_N z^N) b_N^{N-1}|$  en una región del plano comple-

jo que contiene a todos los segmentos que unen a  $o$  con los  $\beta_\nu$ , resultará que  $|\varepsilon_\nu|$  será no mayor que el producto de  $G^{p-1}/(p-1)!$  por factores que son independientes de  $p$ . Resultará así que si se aumenta convenientemente el valor de  $p$ , lo que es siempre posible, los números  $\varepsilon_\nu$  y por tanto la suma  $\varepsilon_1 + \dots + \varepsilon_\nu$  puede hacerse tan pequeña como se quiera; en particular, menor que la unidad.

5. Nos ocuparemos ahora de los números  $M_\nu$ .

Si reemplazamos en el integrando de la ecuación (14) la expresión entre corchetes por su valor deducido de (9) y sustituimos luego la variable  $z$  por la variable  $\zeta_\nu = z - \beta_\nu$ , que deberá variar entre  $0$  e  $\infty$ , se obtiene

$$M_\nu = \int_0^\infty \frac{e^{-\zeta} d\zeta}{(p-1)!} (\zeta + \beta_\nu)^{p-1} (\zeta + \beta_\nu - \beta_1)^p \dots \zeta^p \dots$$

$$\dots (\zeta + \beta_\nu - \beta_N)^p b_N^{Np-1} = \int_0^\infty \frac{e^{-\zeta} d\zeta}{(p-1)!} \zeta^p \Phi_\nu(\zeta), \quad (15)$$

en donde

$$\begin{aligned}
 \Phi_{\nu}(\zeta) &= b_N^{Np-1} (\zeta + \beta_{\nu})^{p-1} (\zeta + \beta_{\nu} - \beta_1)^p \dots \\
 &\dots (\zeta + \beta_{\nu} - \beta_{\nu-1})^p (\zeta + \beta_{\nu} - \beta_{\nu+1})^p \dots \\
 &\dots (\zeta + \beta_{\nu} - \beta_N)^p.
 \end{aligned}
 \tag{16}$$

Se tiene entonces que

$$\sum_{\nu=1}^N M_{\nu} = \sum_{\nu=1}^N \int_0^{\infty} \frac{e^{-\zeta} d\zeta}{(p-1)!} \zeta^p \Phi(\zeta),
 \tag{17}$$

en donde

$$\begin{aligned}
 \Phi(\zeta) &= \sum_{\nu=1}^N b_N^{Np-1} (\zeta + \beta_{\nu})^{p-1} (\zeta + \beta_{\nu} - \beta_1)^p \dots \\
 &\dots (\zeta + \beta_{\nu} - \beta_{\nu-1})^p (\zeta + \beta_{\nu} - \beta_{\nu+1})^p \dots \\
 &\dots (\zeta + \beta_{\nu} - \beta_N)^p
 \end{aligned}$$

es un polinomio en  $\zeta$ , en cada uno de cuyos sumandos una de las  $N$  cantidades  $\beta_1, \beta_2, \dots, \beta_N$  juega un papel particular; pero en el conjunto de ellos ninguna de esas cantidades tiene un rol diferente del de las demás. Por esa razón, si se desarrollan las potencias y se llevan a cabo los productos

de la expresión de  $\Phi(\zeta)$ , cada uno de los coeficientes de las potencias de  $\zeta$  es una función racional entera y simétrica de  $\beta_1, \dots, \beta_N$  con coeficientes enteros. Ahora bien, los números  $\beta_1, \dots, \beta_N$  son todas raíces de la ecuación (6) cuyos coeficientes son enteros, pero que no es mónica. Por tanto el teorema 2 de la sección 5.1 nos permite establecer que los mencionados coeficientes de las potencias de  $\zeta$  del polinomio  $\Phi(\zeta)$  son racionales. Pero nosotros deseamos probar que esos coeficientes son *enteros*. Esto podemos lograrlo valiéndonos del factor entero  $b_N^{Np-1}$  que aparece en todos los términos de  $\Phi(\zeta)$ . Puede escribirse en efecto

$$\begin{aligned} \Phi(\zeta) = & \sum_{\nu=1}^N (b_N \zeta + b_N \beta_\nu)^{p-1} (b_N \zeta + b_N \beta_\nu - b_N \beta_1)^p \dots \\ & \dots (b_N \zeta + b_N \beta_\nu - b_N \beta_{\nu-1})^p (b_N \zeta - b_N \beta_\nu - b_N \beta_{\nu+1})^p \\ & \dots (b_N \zeta + b_N \beta_\nu - b_N \beta_N)^p. \end{aligned}$$

Cuando se lleva a cabo el cálculo de los coeficientes de las potencias de  $\zeta$  median-

te la fórmula de la potencia de un polinomio se obtienen funciones enteras racionales simétricas de los productos  $b_N \beta_1, b_N \beta_2, \dots, b_N \beta_N$ , con coeficientes enteros. Ahora bien, estos últimos  $N$  productos son todas las raíces de la ecuación que resulta de la ecuación (4) cuando se reemplaza la incógnita  $z$  por la incógnita  $w$  tal que  $w = b_N z$ , o sea la ecuación

$$b_0 + b_1 \left(\frac{w}{b_N}\right) + \dots + b_{N-1} \left(\frac{w}{b_N}\right)^{N-1} + b_N \left(\frac{w}{b_N}\right)^N = 0,$$

y también, por tanto, de la que resulta de ésta al multiplicarla por  $b_N^{N-1}$ , o sea

$$b_0 b_N^{N-1} + b_1 b_N^{N-2} w + \dots + b_{N-2} b_N w^{N-2} + b_{N-1} w^{N-1} + w^N = 0, \quad (18)$$

que es una ecuación mónica con coeficientes enteros. Resulta así, según el teorema 2 de la sección 5.1, que los coeficientes del polinomio  $\Phi(\zeta)$  son números enteros que denotaremos por  $A_0, A_1, \dots, A_{N-1}$ . Se tiene por consiguiente

$$\sum_{\nu=1}^N M_{\nu} = \int_0^{\infty} \frac{e^{-\zeta} \zeta^p d\zeta}{(p-1)!} (A_0 + A_1 \zeta + A_2 \zeta^2 + \dots + A_{Np-1} \zeta^{Np-1}),$$

y en virtud de (12) resulta

$$\sum_{\nu=1}^N M_{\nu} = \frac{1}{(p-1)!} (A_0 p! + A_1 (p+1)! + \dots + A_{Np-1} (Np)!), \quad (19)$$

que prueba que  $\sum_{\nu=1}^N M_{\nu}$  es un número entero divisible por  $p$ .<sup>(1)</sup>

---

(1) Cabe aquí observar que, tal como indicamos en una nota de pie de página precedente, los números  $M_{\nu}$  son enteros algebraicos. En efecto, puesto que  $b_N \beta_1, \dots, b_N \beta_N$  son todas las raíces de la ecuación mónica (18) con coeficientes enteros, dichos números son, por definición, enteros algebraicos. De (15) puede deducirse entonces, mediante la fórmula (12), que  $M_{\nu}$  es una expresión racional entera de números algebraicos. De lo dicho al final de la sección 5.1 resulta que  $M_{\nu}$  es un número entero algebraico.

En la parte 2 de la demostración hemos probado que  $a_0 M$  es un entero que no es divisible por  $p$ .

Por consiguiente la suma

$$a_0 M + \sum_{\nu=1}^N M_\nu$$

es necesariamente un número racional entero que no es divisible por  $p$ , y por tanto es diferente de cero. Resulta así que la ecuación

$$\left[ a_0 M + \sum_{\nu=1}^N M_\nu \right] + \left[ \sum_{\nu=1}^N \varepsilon_\nu \right] = 0,$$

que es la ecuación (8) no puede cumplirse, pues no es posible que la suma de un entero diferente de cero y de un número que es menor que la unidad pueda ser igual a cero. Esto prueba, según fue explicado al comenzar la demostración, que si se supone que  $\pi$  es un número algebraico, la ecuación (2), que es cierta sin duda, no podría cumplirse. Luego  $\pi$  es un número trascen-



dente, y como consecuencia resulta la imposibilidad de resolver el problema clásico de la cuadratura del círculo.

## § 6. ANEXO

### Los Números Algebraicos

Para las definiciones de los números algebraicos y de los números enteros algebraicos remitimos al lector a la sección 5.1 .

Sea  $\mathbb{Q}$  el campo de los números racionales y  $\mathbb{Q}[x]$  el anillo de los polinomios en  $x$ , con coeficientes racionales.

Un polinomio  $f(x) \in \mathbb{Q}[x]$  se llama *irreducible* si no es posible expresarlo como un producto de dos polinomios ninguno de los cuales sea de grado 0, es decir elementos pertenecientes a  $\mathbb{Q}$ .

#### **Teorema 3.**

La suma, la diferencia, el producto y el cociente (en este último caso con divisor

diferente de cero) de dos números algebraicos, son números algebraicos.

**Demostración.**

Sean  $\alpha$  y  $\beta$  dos números algebraicos. Por definición serán raíces de las respectivas ecuaciones mónicas

$$x^m + a_1 x^{m-1} + \dots + a_m = 0 \quad (1)$$

$$x^n + b_1 x^{n-1} + \dots + b_n = 0 \quad (2)$$

cuyos coeficientes son números racionales.

Sean  $\alpha_0 = \alpha, \alpha_1, \dots, \alpha_{m-1}$  las raíces de (1) y  $\beta_0 = \beta, \beta_1, \dots, \beta_{n-1}$  las raíces de (2).

Consideremos la ecuación mónica

$$\prod_{i,j} (x - (\alpha_i + \beta_j)) = x^{mn} + c_1 x^{mn-1} + \dots + c_{mn} = 0, \quad (3)$$

donde el producto del primer miembro se extiende a todos los pares  $(i, j)$  tales que  $i = 0, 1, \dots, m-1, j = 0, 1, \dots, n-1$ .

Los coeficientes  $c_1, c_2, \dots, c_{mn}$  son funciones enteras simétricas de las raíces de

las ecuaciones (1) y (2); son por tanto funciones racionales enteras de los coeficientes de dichas ecuaciones, es decir que son números racionales. Puesto que  $\alpha + \beta = \alpha_0 + \beta_0$  es una raíz de (3), será un número algebraico. De manera semejante se prueba que  $\alpha - \beta$  y  $\alpha\beta$  son números algebraicos. Para probar que  $\alpha/\beta$ , con  $\beta \neq 0$ , es algebraico, puede observarse que  $1/\beta$  es algebraico porque es raíz de la ecuación de coeficientes racionales

$$b_n y^n + b_{n-1} y^{n-1} + \dots + b_1 y + 1 = 0$$

Resulta entonces que el producto  $\alpha \cdot \frac{1}{\beta} = \frac{\alpha}{\beta}$  es algebraico. ■

### **Corolario 1.**

Toda función racional de números algebraicos con coeficientes racionales, en que ningún divisor es nulo, es un número algebraico.

### **Demostración.**

Resulta del Teorema porque todo número racional es algebraico. ■

## Corolario 2.

La suma, la diferencia y el producto de dos números enteros algebraicos son enteros algebraicos.

### Demostración.

En este caso, los coeficientes  $c$  de la ecuación (3), en el caso de la suma  $\alpha + \beta$  de dos enteros algebraicos, son funciones racionales enteras de los coeficientes  $a$  y  $b$ ; y puesto que estos son ahora enteros por hipótesis, los  $c$  serán también enteros. Luego  $\alpha + \beta$  será un número entero algebraico. Semejantemente para  $\alpha - \beta$  y  $\alpha\beta$ . ■

La siguiente proposición no es utilizada en los párrafos precedentes. Se incluye aquí únicamente a título de información complementaria de interés en la teoría de los números algebraicos.

## Teorema 4.

Si  $\omega$  es una raíz de la ecuación

$$F(x) = x^n + \alpha_1 x^{n-1} + \dots + \alpha_n = 0,$$

donde los coeficientes son números alge-

braicos, es también raíz de una ecuación con coeficientes racionales y por tanto es un número algebraico.

### Demostración.

Cada uno de los números  $\alpha_i (i=1, \dots, n)$ , por cuanto es algebraico, será raíz de una ecuación mónica  $\varphi_i(x) = 0$  de grado  $m_i$  con coeficientes racionales, cuyas raíces vamos a designar por

$$\alpha_i^0 = \alpha_i, \alpha_i', \alpha_i'', \dots, \alpha_i^{(m_i-1)}.$$

Sea  $m = m_1 m_2 \dots m_n$ , y definamos los  $m$  polinomios

$$F_j(x) = x^n + \alpha_1^{j_1} x^{n-1} + \alpha_2^{j_2} x^{n-2} + \dots + \alpha_{n-1}^{j_{n-1}} x + \alpha_n^{j_n}$$

donde  $j = \{j_1, \dots, j_n\}$ ,  $j_k \in \{0, 1, 2, \dots, m_k - 1\}$  para  $k = 1, 2, \dots, n$ ; de manera que

$$F_{(0, \dots, 0)} = F(x).$$

Consideremos el polinomio

$$f(x) = \prod_j F_j(x)$$

donde el producto se extiende a todos los

conjuntos  $j$ . El número  $\omega$  es raíz de la ecuación

$$f(x) = 0.$$

Ahora bien, los coeficientes de  $f(x)$  son evidentemente funciones racionales enteras simétricas de todas las raíces de la ecuación mónica  $\varphi_1(x)\varphi_2(x)\dots\varphi_n(x) = 0$  de coeficientes racionales; por consiguiente, según el teorema 2, son funciones racionales de estos últimos coeficientes, y por consiguiente ellos mismos son números racionales. Puesto que  $\omega$  es raíz de la ecuación  $f(x) = 0$ , será un número algebraico. ■

### Ejemplo.

Sea  $\omega$  una raíz de la ecuación

$$F(x) = x^2 + \sqrt{2}x + \sqrt{3} = 0.$$

Tenemos ahora:  $\alpha_1 = \sqrt{2}$  y  $\alpha_2 = \sqrt{3}$ , de modo que  $\varphi_1(x) = x^2 - 2$  y  $\varphi_2(x) = x^2 - 3$ . Por tanto

$$\alpha_1^{\circ} = \sqrt{2}, \alpha_1' = -\sqrt{2}, \alpha_2^{\circ} = \sqrt{3}, \alpha_2' = -\sqrt{3},$$

y

$$F_{(0,0)} = x^2 + \sqrt{2}x + \sqrt{3}$$

$$F_{(1,0)} = x^2 - \sqrt{2}x + \sqrt{3}$$

$$F_{(0,1)} = x^2 + \sqrt{2}x - \sqrt{3}$$

$$F_{(1,1)} = x^2 - \sqrt{2}x - \sqrt{3}.$$

Se tiene entonces

$$f(x) = F_{(0,0)} \cdot F_{(1,0)} \cdot F_{(0,1)} \cdot F_{(1,1)} = \\ x^8 - 4x^6 - 2x^4 - 12x^2 + 9 = 0.$$

Puesto que  $\omega$  es una raíz de esta última ecuación, es un número algebraico. Puede observarse que es un entero algebraico; lo cual es consecuencia de que los coeficientes del polinomio  $F(x)$  son enteros algebraicos, según resulta del siguiente corolario.

### Corolario.

Si  $\omega$  es raíz de la ecuación

$$F(x) = x^n + \alpha_1 x^{n-1} + \dots + \alpha_n = 0,$$

donde  $\alpha_1, \alpha_2, \dots, \alpha_n$  son enteros algebraicos, entonces él mismo es un entero algebraico.

### **Demostración.**

Usando las notaciones de la demostración del teorema anterior, los coeficientes de  $f(x)$  son funciones racionales enteras de los coeficientes de las ecuaciones  $\varphi_i(x) = 0$ , los cuales son ahora enteros racionales. Luego los coeficientes de  $f(x)$  son racionales enteros y  $\omega$  es un entero algebraico. ■

### **Bibliografía.**

- [1] *Courant and Robbins*, What is Mathematics? Oxford University Press, (1941).
- [2] *Hermite*, Comptes Rendus, vol.77(1873); págs. 18-24, 74-79, 226-233, 285-293.
- [3] *Hilbert*, Mathematische Annalen, 43(1893)
- [4] *Lindemann*, Sitzungsberichte der Berliner Akademie. (1882), pág.679.  
Mathematische Annalen, 20 (1882), pág. 213.
- [5] *Niven*, Bull. of the Am. Math. Soc., 53 (1947).



- [6] *Reid L.W.* The Elements of the Theory of Algebraic Numbers, The Mac Millan Co. (1910).
- [7] *Van der Waerden,* Moderne Algebra I, (1937).
- [8] *Weber,* Lehrbuch der Algebra I, Braunschweig, Vierweg und Sohn, (1912)

## **FUNCIONES REALES $\alpha$ -HOMOGENEAS DE GRADO UNO**

**Fernando E. TORRES ORIHUELA\***

---

\* Estudiante de Doctorado del Instituto de Matemática Pura y Aplicada, Río de Janeiro, Brasil.

Una función  $f : \mathbb{R}^g \rightarrow \mathbb{R}$  será llamada  $\alpha$  - homogénea de grado uno si

$$\alpha \in \mathbb{R}^+ \setminus \{1\} \text{ y } \forall x \in \mathbb{R}^g: f(\alpha x) = \alpha x.$$

Tales funciones - por ejemplo - aparecen como funciones de producción en Economía (Ver [1]) y el objetivo de este artículo es responder a ¿qué podemos decir acerca de una tal función?. Es fácil ver que una condición necesaria para que una función  $f$  sea  $\alpha$  - homogénea es que  $f(0) = 0$  y que toda función  $\mathbb{R}$  - lineal definida en  $\mathbb{R}^g$  es  $\alpha$  - homogénea. Sin embargo, como se verá a continuación, una función no tiene por qué ser continua en el 0, y, aun permitiendo discontinuidad en un conjunto de medida de Lebesgue cero, no se obtendrá la linealidad.

Sea  $\mathfrak{F}_\alpha := \{ f : \mathbb{R}^g \rightarrow \mathbb{R} / f \text{ es } \alpha \text{ - homogénea de grado uno} \}$ . A continuación sólo consideraremos el caso  $\alpha > 1$  debido que el otro puede ser tratado similarmente con  $\alpha^{-1}$ . Nótese también que  $\mathbb{R}^g$  puede ser reemplazado por un Espacio de Banach.

Sea  $\mathbb{C}_\alpha := \{x \in \mathbb{R}^g : 1 \leq \|x\| < \alpha\}$  (especie de "hiper-casquete" sólido) donde  $\| \cdot \|$  denota una norma de  $\mathbb{R}^g$ , e.g. la euclidiana.

Sea  $\mathfrak{F}_{\alpha,1} := \mathbb{R}^{\mathbb{C}_\alpha}$ .

**Lema.**  $\forall x \in \mathbb{R}^g \setminus \{0\} \exists! n \in \mathbb{Z} / \alpha^n \leq \|x\| < \alpha^{n+1}$ .

**Prueba.** La unicidad es clara. Para la existencia: Si  $\|x\| \geq 1 \Rightarrow n := \max \{m \in \mathbb{Z} : \alpha^m \leq \|x\|\}$ . Si  $\|x\| < 1 \Rightarrow n := \min \{m \in \mathbb{Z}_0^+ : \|x\| \leq \alpha^{-m}\}$ . ■

Tomemos  $\sigma \in \mathfrak{F}_{\alpha,1}$  y sea  $f_\sigma : \mathbb{R}^g \rightarrow \mathbb{R}$  definida por :

$$x \mapsto \begin{cases} 0 & \text{si } x = 0 \\ \alpha^n \sigma(\alpha^{-n}x) & \text{si } x \neq 0 \end{cases}$$

y  $n$  es como el lema.

**Proposición 1.**

$$\mathfrak{F}_\alpha = \{f_\sigma : \sigma \in \mathfrak{F}_{\alpha,1}\}$$

### Prueba.-

" $\subseteq$ " : Es suficiente restringir  $f \in \mathfrak{F}_\alpha$  a  $\mathbb{C}_\alpha$  y usar su  $\alpha$ -homogeneidad.

" $\supseteq$ " : Sean  $\sigma \in \mathfrak{F}_{\alpha,1}$ ,  $x \in \mathbb{R}^g \setminus \{0\}$  y  $n$  como en el lema. Entonces  $\alpha^{n+1} \leq \| \alpha x \| < \alpha^{n+2}$ . Luego,  $f_\sigma(\alpha x) = \alpha^{n+1} \sigma(\alpha^{-n-1} \alpha x) = \alpha f_\sigma(x)$ . ■

¿Qué condiciones aseguran un comportamiento "aceptable" para  $f \in \mathfrak{F}_\alpha$ ? Debido a la existencia de funciones continuas sin derivada en ningún punto [2, Thm. 1, pag.37] y  $f(0) = 0$  observamos que 0 tendrá un papel decisivo en el comportamiento global de  $f$ . Es claro que la continuidad en el 0 no es suficiente para garantizar diferenciabilidad en una vecindad de tal.

### Proposición 2.

Si  $f \in \mathfrak{F}_\alpha$  es diferenciable en el 0, entonces  $f$  es diferenciable en  $\mathbb{R}^g$ . Más aún,  $f$  es  $\mathbb{R}$ -lineal.

**Prueba.** - La existencia de la derivada en el 0 ( $Df(0)$ ), implica la existencia de derivadas direccionales, i.e.,  $\forall v \in \mathbb{R}^g$  :

$$Df(0).v = \lim_{t \rightarrow 0} \frac{f(tv)}{t} \quad (*)$$

Consideran la sucesión  $\langle \alpha^{-n} \rangle \rightarrow 0$  de (\*) obtenemos:

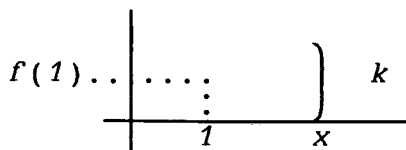
$$Df(0).v = \lim_{n \rightarrow \infty} f(\alpha^{-n}v)\alpha^n \quad (**)$$

La  $\alpha$  - homogeneidad de  $f$  implica :  $\forall n \in \mathbb{Z}_0^+$  :  $f(\alpha^{-n}v) = \alpha^{-n}f(v)$ . De (\*\*) sigue que  $f(v) = Df(0).v$ . ■

A seguir una justificación geométrica de la prop.2 para  $n = 1$ . Para este caso se tiene :

**Escolio:** Si  $f$  es  $\alpha$  - homogénea de  $\mathbb{R}$  en  $\mathbb{R}$  y es diferenciable por la derecha (izquierda) del 0, entonces  $\forall x \geq 0$   $f(x) = \alpha x$  ( $\forall x \leq 0$ ,  $f(x) = \alpha x$ ). ■

En el gráfico :  
y de la existencia  
de la derivada por  
la derecha en el



0, la  $\alpha$ -homogeneidad de  $f$  y una sucesión

$$\langle x_j \rangle \rightarrow 0 \Rightarrow \frac{f(x_j)}{x_j} \rightarrow \frac{k}{x} .$$

Luego si  $k/x \neq f(1)$ ,  $f$  no será diferenciable.

**Agradecimiento.** Al Prof. R. García-Cobián y a A. Lugón por compartir sus ideas al respecto.

### Referencias.

- [1] *Chiappori y Guesnerie*, Strongly Neutral Forecast Funtions, a aparecer en el Jounal of Math. Economics.
- [2] *Singer I.M. and Thorpe J.A.*, Lectures Notes on Elementary Topology and Geometry. Scott, Foresman and Company, 1967.

## **PREDICTORES PARA PRUEBAS DE TIEMPOS DE FALLA**

**Dr. Holger CAPA SANTOS\***

*En este trabajo consideramos una prueba de vida, de tipo II con censura, (terminología anglosajona) de  $N$  elementos cuyos tiempos de vida siguen una ley exponencial y construimos un predictor bayesiano empírico del tiempo en que ocurrirá la  $r$ -ésima falla, habiendo observado la  $k$ -ésima, en la misma prueba. Además comparamos este predictor, con respecto a un predictor clásico y a un predictor bayesiano, a través de la simulación.*

---

\*

Profesor Principal T.C., Esc. Politécnica Nacional, Ecuador.

Profesor Principal T.M., Universidad Central

Consultor del Centro de Investigación Matemáticas Aplicadas a la Ciencia y Tecnología (CIMACYT).



## 1. Introducción.

Consideramos una prueba de vida, de tipo II con censura, de  $N$  elementos cuyos tiempos de vida siguen una ley exponencial y construimos un predictor bayesiano empírico del tiempo en que ocurrirá la  $r$ -ésima falla, habiendo observado la  $k$ -ésima, en la experiencia actual, a partir de observaciones en experiencias precedentes.

Finalmente, comparamos a través de la simulación a este predictor con un predictor obtenido por métodos clásicos por Lawless (1971) y a un predictor bayesiano obtenido por Dunsmore (1974).

## 2. La Predicción Estadística.

La predicción estadística consiste en utilizar las observaciones de un experimento aleatorio (experimento informativo)  $E$  para predecir resultados de otro experimento aleatorio (experimento futuro)  $F$ . Consideramos que  $E$  y  $F$  son independientes.

Supongamos que la información obtenida de la realización de  $E$  puede expresarse a

través de una variable aleatoria (v.a.)  $X$  o de una estadística suficiente de una muestra de la v.a.  $X$ , que tiene por densidad  $f(x : \theta)$ , donde  $\theta$  pertenece a un conjunto de parámetros  $\Omega$ . Buscamos entonces, predecir las observaciones de una v.a.  $Y$  que tiene por densidad  $f(y : \theta)$  y que describe ciertos resultados de  $F$ .

Para situarnos en el contexto bayesiano, supondremos que  $\theta$  sigue una ley a priori  $G(\theta)$ , ( $\theta \in \Omega$ ). Si  $G(\theta)$  admite una densidad  $f(\theta)$ , entonces la densidad predictiva de  $Y$ , dada  $X$ , se expresa por:

$$f(y : x) = \int_{\Omega} f(y : \theta) f(\theta : x) d\theta \quad (1)$$

donde  $f(\theta : x)$  representa la densidad a posteriori de  $X$ .

Para el caso discreto cambiamos, simplemente, el signo de integral por el de sumatoria.

Se puede demostrar que bajo ciertas condiciones de existencia, poniendo  $E(Y:\theta) = u(\theta)$ , donde  $u$  es una función dada (Capa

Santos 1986, por ejemplo), entonces el predictor bayesiano se puede expresar por:

$$\rho(x) = E[u(\theta):x]$$

es decir, utilizando la expresión explícita de la densidad a posteriori  $f(\theta:x)$  en la fórmula (1), tenemos:

$$\rho(x) = \int_{\Omega} u(\theta) f(x:\theta) f(\theta) d\theta / \int_{\Omega} f(x:\theta) f(\theta) d\theta \quad (2)$$

El análisis bayesiano paramétrico considera conocidas  $f(x:\theta)$  y  $f(\theta)$  y de esta forma se puede calcular  $\rho(x)$  utilizando la expresión (2). El análisis bayesiano empírico paramétrico, al contrario, supone conocida  $f(x:\theta)$  y asume solamente la existencia de la ley  $G$  que sigue  $\theta$ .

Para encontrar una regla de decisión en el contexto bayesiano empírico, suponemos que el experimento  $E$  se ha realizado varias veces en el pasado, de manera independiente entre ellas y que  $\theta$  sigue en todos los casos la misma ley a priori, desconocida (esto puede ser de mucha utilidad en los procesos en los que no exis-

ten cambios demasiado bruscos en ciertos períodos de tiempo, permitiéndose utilizar la información precedente). Es decir, en el momento de tomar la decisión actual, disponemos de una serie de elementos aleatorios independientes:

$$(x_1, \theta_1), (x_2, \theta_2), \dots, (x_n, \theta_n) = (x, \theta)$$

Los elementos  $\theta_1, \dots, \theta_n = \theta$  son, evidentemente, desconocidos y no observables. Supondremos también que  $G$  tiene una interpretación frecuencial y utilizaremos los valores  $x_1, \dots, x_n = x$  para obtener la información sobre  $G$  que nos permita deducir una regla de decisión  $\delta_n(x) = \delta_n(x, x_1, \dots, x_{n-1})$  que aproxime, en un cierto sentido, al predictor bayesiano  $\rho$ .  $\delta_n$  se llamará un *predictor bayesiano empírico (p.b.e.)*

*Lemon y Krutchkoff* (1969) proponen la estimación de la función de distribución  $G$  (o ley a priori) por una función en escalera, con saltos de tamaño  $1/n$  en los puntos  $\hat{\theta}_i (i = 1, \dots, n)$ , donde  $\hat{\theta}_i$  es una estimación clásica de  $\theta$  en la  $i$ -ésima expe-

riencia, en el pasado, basada en estadística suficientes.

### 3. Construcción de un Predictor Bayesiano Empírico.

Supongamos que deseamos medir el tiempo de vida de  $M$  elementos idénticos e independientes. Probaremos a la vez estos elementos y detendremos la experiencia en el momento que ocurra la  $p$ -ésima falla ( $1 \leq p \leq M$ ). Los elementos que fallan durante el experimento no se cambiarán. Definiremos a esta prueba como una prueba de tipo  $(M, B, p)$ , lo que corresponde en la literatura anglosajona a las pruebas censuradas de tipo II.

La v.a.  $T$  representará el tiempo de vida de un elemento y  $(T_1, \dots, T_p)$ , a la muestra ordenada de tiempos de falla. Supondremos que  $T$  sigue una ley exponencial estándar de parámetro  $\theta (\theta > 0)$ , cuya densidad está dada por la fórmula :

$$f(t) = \theta \exp(-\theta t) \quad (t > 0)$$

Si deseamos predecir en la misma prueba el  $r$ -ésimo tiempo de falla, habiendo observado la  $k$ -ésima falla, tenemos el problema de que las experiencias informativa y futura no son independientes. Sin embargo, utilizaremos el hecho de que si el tiempo de vida  $T$  sigue una ley exponencial estándar de parámetro  $\theta$ , entonces (Tanis, 1964) las v.a.:

$$S_k = \sum_{i=1}^k T_i + (N-k)T_k \quad \text{y} \quad Y = T_r - T_k$$

son independientes. Además  $S_k$  sigue una ley  $\Gamma(k, \theta)$  y  $Y$  tiene por densidad:

$$f(y; \theta) = \theta \exp(-(N-r+1)\theta y) (1 - \exp(-\theta y))^{r-k-1} (B(N-r+1, r-k))^{-1}$$

donde,

$$B(N-r+1, r-k) = (N-k)! [(N-r)! (r-k-1)!]^{-1}$$

Tenemos entonces:

$$E(Y; \theta) = (N-k)! [\theta(N-r)!]^{-1} \sum_{i=0}^{r-k-1} (-1)^i [i! (r-k-1-i)! (N-r+i+1)^2]^{-1}$$

Es decir,

$$E(Y:\theta) = \theta^{-1}u(N,r,k)$$

donde  $u$  es una función de las tres variables indicadas e independiente de  $\theta$ . Esto nos permite obtener el predictor bayesiano de  $Y$  (sabiendo que  $S_k = s$ ):

$$\rho(s) = u(N,r,k)E(\theta^{-1}:s)$$

Situemos en el contexto *b.e.* y supongamos aún que los datos  $(k_1, s_{k_1}), \dots, (k_{n-1}, s_{k_{n-1}}) = (k, s_k)$ , donde  $s_k$  es el tiempo de prueba observa hasta la  $k$ -ésima falla en la prueba actual, sean disponibles.

Si  $S_k = s$  y  $\delta_n(s)$  es un estimador *b.e.* de  $1/\theta$  basado en los datos precedentes (el cálculo se hará siguiendo los procedimientos mencionados en (2), entonces:

$$Y_n = u(N,r,k) \delta_n(s)$$

es un predictor *b.e.* de  $Y$ . Esto nos permite dar un predictor *b.e.*  $T_{r,n}$  de  $T_r$  :

$$T_{r,n} = Y_n + T_k$$

Es necesario señalar que el valor de  $T_k$  debe conocerse en la prueba actual.

Tomando como estimador clásico de  $\theta_i$  al estimador de máxima verosimilitud:

$$\hat{\theta}_i = k_i / s_{ki} \quad (i=1, \dots, n)$$

utilizando el procedimiento de Lemon y Krutchkoff y puesto que  $S_k$  sigue una ley  $\Gamma(k, \theta)$ , obtenemos el *p.b.e.* de  $T_r$  siguiente :

$$pbe = u(N, r, k) \left[ \sum_{i=1}^n \hat{\theta}_i^{k-1} \exp(-\hat{\theta}_i s_k) \right] \left[ \sum_{i=1}^n \hat{\theta}_i^k \exp(-\hat{\theta}_i s_k) \right]^{-1} + t_k \quad (3)$$

#### 4. Comparación de los predictores.

Con la finalidad de tener una idea sobre el comportamiento de los predictores *b.e.* de  $T_r$ , consideremos el predictor definido por la fórmula (3) y otros dos predictores obtenidos por métodos diferentes. El primero es un predictor bayesiano (de-



notado  $pb$ ), encontrado por Dunsmore (1974) utilizando una ley gama  $\Gamma(g,h)$  como ley a priori para  $\theta$ , pues estas leyes definen una familia de conjugadas naturales. Para el caso  $r = k + 1$ ,  $pb$  está definido por:

$$pb = H/(N-k) (G-1) \quad (G>1)$$

donde,

$$H = h+S_k \quad \text{y} \quad G = g+k$$

El segundo predictor, es el calculado por Lawless (1971) a través de métodos clásicos, y que corresponde al predictor bayesiano con respecto a una ley no informativa para  $\theta$ ; es decir, cuando tomamos  $h=g=0$ . El predictor de Lawless se expresa entonces como sigue :

$$pc = S_k/(N-k)(k-1) \quad (k>1)$$

La metodología elegida para comparar los predictores puede describirse de la siguiente manera :

Hemos elegido como ley a priori para  $\theta$  una ley gamma  $\Gamma(g,h)$ , pues esta nos permi-

te construir un predictor bayesiano utilizando el criterio de las familias conjugadas naturales. Varias combinaciones de los parámetros han hecho posible la comparación de los predictores con respecto a diferentes leyes a priori  $\Gamma(g, h)$ .

Cuando se ha fijado una combinación de parámetros  $(g, h)$ , se generan  $n$  observaciones  $\theta_1, \dots, \theta_n$  de una ley  $\Gamma(g, h)$ . Con cada una de las primeras  $(n-1)$  observaciones construimos  $(n-1)$  muestras completas de una ley exponencial estándar, todas del mismo tamaño (tomamos  $k_1 = 6$ , y entonces  $k_{i-1} = 5$ ,  $(i = 1, \dots, n-1)$ ). Estas muestras deben ser independientes, condicionalmente a los  $\theta_i$ , y ordenadas para calcular los tiempos de prueba  $S_i$ .

Con la última observación  $\theta_n$  generamos una muestra completa, independiente de las precedentes, de tamaño  $N = 10$ , de una ley exponencial estándar de parámetro  $\theta_n$ , con la cual efectuaremos nuestras previsiones. Luego de ordenarla, calculamos  $S_n$  y  $(k_n = 6)$ .

Finalmente calculamos el estimador de máxima verosimilitud de  $\theta_i$ , con la información de la  $i$ -ésima muestra, de acuerdo a lo indicado en la sección precedente.

Ahora procedemos a calcular los diversos predictores de  $T_7$  (es decir, hemos tomado  $r = k + 1 = 7$ ) en la prueba actual (la  $n$ -ésima). Hay que recordar que  $s$  y  $k$  corresponden, respectivamente, a  $s_n$  y  $k_n = 6$ .

Los valores de  $n$  considerados varían entre 2 y 21. Nos limitaremos a presentar ciertos resultados hasta  $n = 11$ , pues luego de este valor no constatamos diferencias importantes. Puesto que, con observaciones simples era difícil obtener conclusiones definitivas, para cada valor de  $n$  repetimos la experiencia 100 veces, lo que nos permitió aproximar los riesgos bayesianos de los predictores por sus riesgos empíricos, con la fórmula :

$$\left[ \sum_{j=1}^{100} (\text{valor de la } j\text{-ésima previsión} - t_r)^2 \right] / 100$$

Para ciertos análisis complementarios, también calculamos las varianzas empíricas de estas aproximaciones de los riesgos bayesianos.

Hemos calculado además, la media teórica (media de las observaciones marginales):

$$m_t = h(g-1)^{-1} \sum_{i=1}^r (N-i+1)^{-1}$$

al igual que las medias empíricas de  $T_r$ .

En las tablas 1 y 2 presentamos algunos resultados de los ensayos numéricos, con los que podemos constatar la concordancia con los análisis teóricos. Por ejemplo, los riesgos del predictor bayesiano son casi siempre más pequeños que aquellos de los otros predictores. Es necesario observar que los riesgos del *pb* y los del *pbe* están bastante cerca y los del *pbe* son más pequeños que los del *pc*; esto se puede también constatar en la fig. 1, donde hemos representado los comportamientos de *pb*, *pc* y *pbe*, para una ley a priori

$\Gamma(4,6)$ . Notemos, que en el caso en que los riesgos empíricos tengan valores muy cercanos, las varianzas empíricas de los riesgos de *pc* son, en general, más grandes que las otras.

## 5. Conclusiones.

La información que presentamos contiene resultados de tipo bayesiano; esto nos ha permitido controlar la calidad de las simulaciones. Recordemos, sin embargo, que en la mayoría de problemas la ley a priori de  $\theta$  es desconocida; sería entonces, más interesante comparar los comportamientos del predictor *b.e.* y el predictor clásico.

El predictor *pbe*, tiene un comportamiento similar al *pb* y además se lo puede calcular rápidamente con una calculadora de bolsillo, lo que nos haría preferirlo en la práctica.

Aunque no hemos presentado aquí los resultados concernientes a los intervalos predictivos, señalemos las principales conclusiones a las que llegamos por medio

de estos ensayos numéricos: Se puede elegir el intervalo predictivo *b.e.* si se espera que la observación por predecir sea grande. Por el contrario, si ésta es pequeña, sería preferible utilizar el intervalo predictivo clásico.

### **Bibliografía.**

- [1] *Capa Santos H. (1986)*, Méthodes bayésiennes empiriques. Applications en fiabilité, Tesis Doctoral, Universidad de París (París VI), París, Francia.
- [2] *Dunsmore I. (1974)*, The Bayesian Predictive Distribution in Life Testing Models, *Technometrics*. V. 16, #3, pp. 455-460.
- [3] *Lawless J. (1971)*, A prediction problem concerning samples from the exponential distributions, with application in life testing, *Technometrics*, V. 13, pp. 725-730.

- [4] *Lemmon G. y Krutchkoff R. (1969)*, An empirical Bayes smoothing technique, *Biometrika*, V. 56, pp. 361-365.
- [5] *Tanis E. A. (1964)*, Linear forms in the statistics from an exponential distribution, *Ann. Math. Statist.*, V. 35, pp. 270-276.

**TABLA 1**

	<b>pb</b>	<b>pbe</b>	<b>pc</b>	<b>obs.</b>
<i>n</i> = 2				
medias empíricas:	2,16	2,16	2,25	2,13
riesgos empíricos:	0,17	0,20	0,27	
var. empl. riesgos:	0,09	0,11	0,25	
<i>n</i> = 3				
medias empíricas:	2,29	2,30	2,40	2,19
riesgos empíricos:	0,26	0,30	0,40	
var. empl. riesgos:	1,26	1,17	1,27	
<i>n</i> = 4				
medias empíricas:	2,12	2,12	2,22	1,86
riesgos empíricos:	0,50	0,53	0,56	
var. empl. riesgos:	2,02	1,83	1,67	
<i>n</i> = 5				
medias empíricas:	1,95	1,93	1,99	1,86
riesgos empíricos:	0,15	0,17	0,21	
var. empl. riesgos:	0,09	0,09	0,14	
<i>n</i> = 6				
medias empíricas:	2,41	2,43	2,54	2,35
riesgos empíricos:	0,19	0,24	0,35	
var. empl. riesgos:	0,10	0,19	0,54	
<i>n</i> = 7				
medias empíricas:	1,97	1,97	2,04	1,90
riesgos empíricos:	0,16	0,17	0,22	
var. empl. riesgos:	0,14	0,15	0,20	
<i>n</i> = 8				
medias empíricas:	2,26	2,28	2,36	2,19
riesgos empíricos:	0,26	0,23	0,30	
var. empl. riesgos:	0,44	0,20	0,23	



$n = 9$

medias empíricas:	2,07	2,09	2,16	2,15
riesgos empíricos:	0,35	0,38	0,36	
var. empl riesgos:	0,50	0,70	0,59	

$n = 10$

medias empíricas:	2,20	2,22	2,30	2,28
riesgos empíricos:	0,52	0,50	0,53	
var. empl riesgos:	3,16	2,35	2,69	2,76

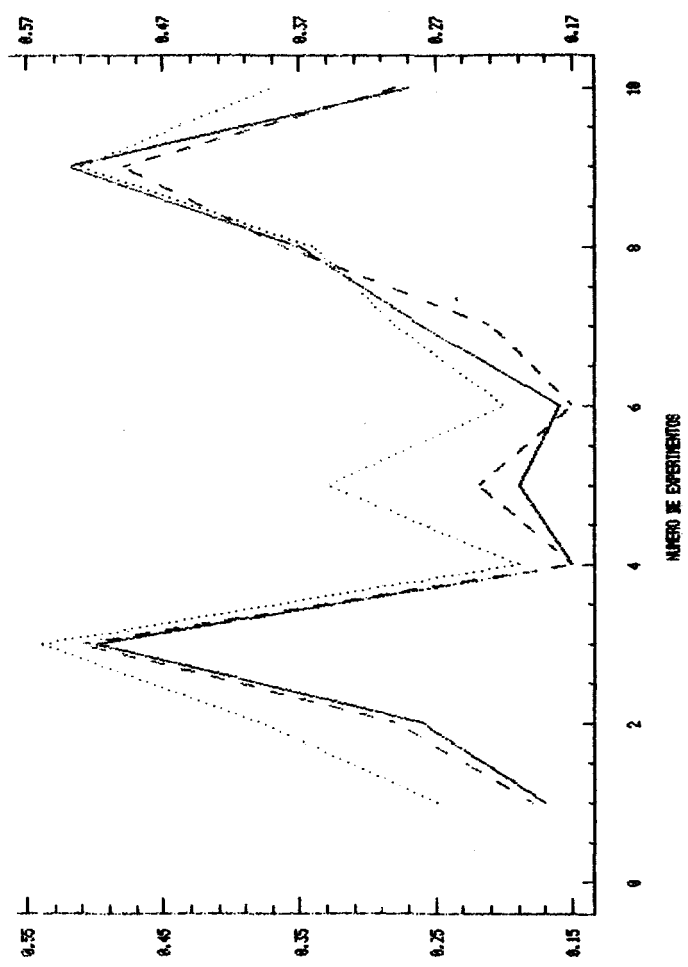
$n = 11$

medias empíricas:	2,22	2,23	2,32	2,28
riesgos empíricos:	0,27	0,30	0,39	
var. empl riesgos:	0,59	0,64	0,78	

Resultados con una ley a priori  $\Gamma(4,6)$ .  
Media teórica  $m_t = 2,19$

RESERVOES EMPÍRICOS DE LOS PREDICTORES

- P<sub>B</sub>
- - P<sub>Ma</sub>
- ..... P<sub>C</sub>



## **CASOS PARTICULARES DE LA DESIGUALDAD DE JOHN-NIRENBERG PARA ESPACIOS $BMO_\varphi$**

**ALEJANDRO ORTIZ FERNANDEZ\***

*En esta nota se dan los detalles de casos particulares de una extensión de la desigualdad de John-Nirenberg,*

$$|\{x \in Q / |f(x) - f_Q| > \lambda\}| \leq c \Psi_{|Q|}^{-1} \left( \frac{\lambda}{\|f\|_{*,\varphi}} \right),$$

donde  $f \in BMO_\varphi$ ,

$$\Psi_\alpha(t) = \int_t^{2^n \alpha} \frac{\varphi(y)}{y} dy, \quad 0 < \alpha < 1,$$

siendo  $\varphi$  una función continua, no-decreciente, de valor real, con  $\varphi(0)=0$ .

Tal extensión nos fue comunicada por el Profesor A. Torchinsky, [6], a quien agradecemos.

---

\* Profesor Asociado de la PUCP.

## 1. Introducción.

El espacio de las funciones de oscilación media acotada,  $BMO$ , fue introducido por F. John - L. Nirenberg en 1961, [3], donde

$$BMO = \{f \in L^1(Q_0) / \|f\|_* = \sup_{Q \subset Q_0} \frac{1}{|Q|} \int_Q |f(x) - f_Q| dx < \infty\}$$

siendo  $Q_0$  un cubo fijo en  $\mathbb{R}^n$ , de medida de Lebesgue  $|Q_0|$  finita, cuyos lados (así como los de los cubos  $Q$ ) son paralelos a los ejes coordenados;  $f_Q$  es el promedio

$$\frac{1}{|Q|} \int_Q f(x) dx. \text{ Con la norma } \|f\|_* +$$

$\|f\|_{L^1(Q_0)}$ ,  $BMO$  es un espacio de Banach.

Obsérvese que  $L^\infty \subset BMO$  siendo la inclusión propia ya que  $\log |x| \in BMO$ . El espacio  $BMO$  está relacionado a otras ramas del análisis (funciones analíticas, ecuaciones en derivadas parciales, martingalas, análisis armónico, ...). La desigualdad de John-Nirenberg es como sigue.

Sea  $\lambda > 0$  un número real y sea el conjunto  $E_\lambda = \{x \in Q / |f(x) - f_Q| > \lambda\}$ . Si  $\omega(\lambda) \equiv |E_\lambda| \leq A e^{-b\lambda} |Q|$ , con  $A$  y  $b$  apropiadas constantes, entonces  $f \in BMO$ .

$$[\text{en efecto, } \int_Q |f(x) - f_Q| dx = \int_0^\infty \omega(\lambda) d\lambda \leq$$

$$A \int_0^\infty e^{-b\lambda} d\lambda \cdot |Q| = \frac{A}{b} |Q|].$$

El recíproco es la parte crucial: si  $f \in BMO$ , entonces existen constantes  $A$  y  $b$  tal que

$$\omega(\lambda) \leq A e^{-b\lambda \|f\|_*^{-1}} \cdot |Q| \quad [+]$$

La prueba original de John-Nirenberg es un tanto técnica; ella fue mejorada por A.P. Calderón, según aparece en Neri [4], en donde se hace uso de la descomposición de Calderón-Zygmund.

Como en la prueba de la extensión de la desigualdad de John-Nirenberg se usará la idea de tal descomposición, veamos algunos detalles de la misma.

### Descomposición de Calderón-Zygmund.

Sea  $f$  una función integrable, definida en un cubo  $Q_0$  y sea  $\lambda > 0$  un real tal que

$$\frac{1}{|Q_0|} \int_{Q_0} |f(x)| dx \leq \lambda .$$

Entonces existe una familia enumerable  $\{Q_k\}$ , de cubos abiertos, disjuntos en  $Q_0$ , tal que

(a)  $|f(x)| \leq \lambda$  a.e. si  $x \in Q_0 - \bigcup_k Q_k$  ;

(b)  $\lambda < \frac{1}{|Q_k|} \int_{Q_k} |f(x)| dx \leq 2^n \lambda$  ;

(c)  $\sum_k |Q_k| < \frac{1}{\lambda} \int_{Q_0} |f(x)| dx .$

### Prueba.

Dividamos  $Q_0$  en  $2^n$  cubos abiertos congruentes (dividiendo por 2 sus lados). Con estos subcubos se presentan dos casos.

(I)  $\frac{1}{|Q'_i|} \int_{Q'_i} |f(x)| dx > \lambda$  ;

$$(II) \quad \frac{1}{|Q''_i|} \int_{Q''_i} |f(x)| dx \leq \lambda .$$

Conservemos los cubos  $Q'_i$  que satisfacen (I) (que es parte de la conclusión (b)), mientras que los cubos  $Q''_i$ , que satisfacen (II), son sometidos al anterior proceso, esto es, son divididos en  $2^n$  nuevos subcubos congruentes, dando origen a una nueva generación de cubos en donde nuevamente tenemos

$$(I) \quad \frac{1}{|Q'_{ii}|} \int_{Q'_{ii}} |f(x)| dx > \lambda ;$$

$$(II) \quad \frac{1}{|Q''_{ii}|} \int_{Q''_{ii}} |f(x)| dx \leq \lambda .$$

Retenemos los cubos  $Q'_{ii}$ , mientras los  $Q''_{ii}$  son sometidos a tal proceso. Y así sucesivamente ... Renumerando obtenemos familias de cubos abiertos, disjuntos, de distintas generaciones  $Q_1, Q_2, \dots, Q_k \dots$  tal que

$$\lambda |Q_k| < \int_{Q_k} |f(x)| dx \leq 2^n \lambda |Q_k| ,$$

que implica (b).

Por otro lado,  $|Q_k| < \frac{1}{\lambda} \int_{Q_k} |f(x)| dx$ ,

de donde  $\sum_k |Q_k| < \frac{1}{\lambda} \int_{Q_0} |f(x)| dx$ , que es (c).

Finalmente, sea  $x \in Q_0 - \bigcup_k Q_k$ , esto es,  $x$  pertenece a algún cubo del tipo  $Q_k''$ , donde (por construcción)  $|Q_k''| \rightarrow 0$  cuando

$k \rightarrow \infty$ . Como  $\frac{1}{|Q_k''|} \int_{Q_k''} |f(y)| dy \leq \lambda$ , por

el teorema de diferenciación de Lebesgue tenemos

$$|f(x)| = \lim_{k \rightarrow \infty} \frac{1}{|Q_k''|} \int_{Q_k''} |f(y)| dy \leq \lambda \quad \text{a.e.}$$

■

## 2. Espacios $BMO_\varphi$ .

El espacio  $BMO$  fue extendido por S. Spanne [5], 1965, vía el espacio  $BMO_\varphi$ , donde  $\varphi(t)$  es una función positiva, no - decreciente, definida sobre  $(0, \infty)$ . Así,



$$BMO_{\varphi} = \left\{ f \in L^1(Q_0) / \|f\|_{*,\varphi} = \sup_{Q \subset Q_0} \frac{1}{\varphi(r)} \frac{1}{|Q|} \int_Q |f(x) - f_Q| dx < \infty \right\}$$

donde  $r$  es la longitud del lado del cubo  $Q$ . Con la respectiva norma,  $BMO_{\varphi}$  es un espacio Banach. Formas particulares de  $\varphi$  hacen coincidir (isomórficamente)  $BMO_{\varphi}$  con algunos clásicos espacios de funciones. Así,

- Si  $\varphi(t) = 1$ , entonces  $BMO_{\varphi} = BMO$  ;
- Si  $\varphi(t) = t^{\alpha}$ ,  $0 < \alpha < 1$ , entonces

$BMO_{\varphi} = \Lambda_{\alpha}$ , donde (espacio de Lipschitz)

$$\Lambda_{\alpha} = \left\{ f \in L^{\infty} / \|f\|_{\Lambda_{\alpha}} = \sup_{x,y} \frac{|f(x) - f(y)|}{|x - y|^{\alpha}} < \infty \right\}$$

Si  $-1 < \alpha < 0$ , entonces se tiene  $BMO_{\varphi} = L^{p,\lambda}$ , con  $\alpha = -\frac{\lambda}{p}$ , siendo (espacio de Morrey)

$$L^{p,\lambda} = \left\{ f \in L^1(Q_0) / \|f\|_{L^{p,\lambda}} = \right. \\ \left. = \sup_Q \left\{ \frac{1}{r^\lambda} \int_Q |f(x)|^p dx \right\}^{1/p} < \infty \right\}.$$

Algunos resultados de Spanne son:

- (i) Si  $\frac{\varphi(t)}{t}$  es casi-decreciente (esto es, si  $t' \leq t$ , existe una constante  $A$  tal que  $\frac{\varphi(t)}{t} \leq A \frac{\varphi(t')}{t'}$ ), entonces

$$f(x) = \int_{|x_1|}^r \frac{\varphi(t)}{t} dt \in BMO_\varphi,$$

$$x = (x_1, \dots, x_n).$$

- (ii) Si  $\frac{\varphi_1(t)}{t}$  es no-creciente, entonces  $BMO_{\varphi_1} \subset BMO_{\varphi_2}$  si y solo si existen constantes  $c, \delta$  tal que  $\varphi_1(r) \leq c \varphi_2(r)$ ,  $0 < r < \delta$ . La inclusión es continua.

- (iii) Si  $\int_0^\delta \frac{\varphi(t)}{t} dt < \infty$ , para algún  $\delta > 0$ , entonces toda  $f \in BMO_\varphi$  es una función continua, y su módulo de continuidad

$\omega(f, r) = \sup_{|x-y| \leq r} |f(x) - f(y)|$  satisfice

$$\omega(f, r) \leq c \int_0^r \frac{\varphi(t)}{t} dt \cdot \|f\|_{*, \varphi}.$$

(iv) Si  $\frac{\varphi(t)}{t}$  es casi decreciente y

$$\int_0^\delta \frac{\varphi(t)}{t} dt = +\infty, \text{ entonces en } BMO_\varphi$$

existen funciones ni acotadas, ni continuas.

En particular, de lo anterior se deduce que

- Si  $\varphi(t) = 1$ , entonces  $\frac{1}{t}$  es casi-decre

ciente y  $\int_0^\delta \frac{dt}{t} = +\infty$ . Luego  $BMO_\varphi = BMO$

contiene a la función no acotada  $f(x) = \log|x|$ .

- Si  $\varphi(t) = t^\alpha$ ,  $0 < \alpha < 1$ , entonces

$$\int_0^r \frac{\varphi(t)}{t} dt = \frac{r^\alpha}{\alpha} < \infty. \text{ Así, } \omega(f, r) \leq$$

$c \frac{r^\alpha}{\alpha} \|f\|_{*, \varphi} \leq c r \|f\|_{*, \varphi}$ . Luego, si  $r$

es pequeño,  $\omega(f,r)$  es pequeño y  $f$  es una función continua. Mas concretamente  $BMO_{t^\alpha} = \Lambda_\alpha$ , lo que constituye el teorema de Meyers-Campanato. Mas generalmente, consideremos el espacio de Lipschitz

$$\Lambda_\varphi = \left\{ f / \|f\|_{\Lambda_\varphi} = \sup_{x,y \in \mathbb{R}^n} \frac{|f(x) - f(y)|}{\varphi(|x-y|)} < \infty \right\}.$$

Se tiene  $\Lambda_\varphi \subset BMO_\varphi$ , con inclusión continua

$$\begin{aligned} & \left[ \text{si } f \in \Lambda_\alpha, \frac{1}{\varphi(r)} \frac{1}{|Q|} \int_Q |f(x) - f_Q| dx \leq \right. \\ & \leq \frac{1}{\varphi(r)} \frac{1}{|Q|} \int_Q \frac{1}{|Q|} \int_Q |f(x) - f(y)| dy dx \leq \\ & \left. \leq \sup_{x,y \in Q} \frac{|f(x) - f(y)|}{\varphi(|x-y|)} < \infty \right]. \end{aligned}$$

En 1971, Ch. Fefferman [1] establece que el espacio  $BMO$  se identifica con el espacio dual del espacio de Hardy  $H^1$ , donde  $H^1 = \{f \in L^1 / R_j f \in L^1, j = 1, \dots, n\}$  siendo  $R_j f$  la transformada de Riesz de  $f$ , defini-

da por  $[R_j f]^\wedge(x) = \frac{x_j}{|x|} \hat{f}(x)$ , con  $\hat{f}$  la transformada de Fourier de  $f$ . Fefferman prueba que tal identificación es equivalente a la caracterización:  $f \in BMO$  si y sólo si  $f = f_0 + \sum_{j=1}^n R_j f_j$ , con  $f_j \in L^\infty$ ,  $j = 0, 1, \dots, n$ . En esta dirección Janson [2], 1976, obtiene una caracterización semejante para  $BMO_\varphi$  imponiendo a  $\varphi$  la (extra) condición de crecimiento

$$\int_r^\infty \frac{\varphi(t)}{t} \frac{dt}{t} \leq c \frac{\varphi(r)}{r} \quad (*) .$$

Se observa que se tiene  $\int_r^\infty \frac{\varphi(t)}{t} \frac{dt}{t} \geq \frac{\varphi(r)}{r}$  y

que la condición (\*) implica que la función (positiva, no-decreciente y continua)

$\Theta(r) = r \int_r^\infty \frac{\varphi(t)}{t} \frac{dt}{t}$  define (por la anterior

caracterización (ii)) al mismo espacio  $BMO_\varphi$ . Esto nos permite asumir que, bajo la condición (\*),  $f \in BMO_\varphi$  es una función continua. Bajo estas consideraciones, Janson prueba que

$f \in BMO_\varphi$  si y sólo si  $f = f_0 + \sum_{j=1}^n R_j f_j$ ,  
 con  $f_j \in \Lambda_\varphi$ ,  $j = 0, 1, \dots, n$ .

### Espacios $L^{p, \Phi}$ .

Sea  $\Phi$  una función positiva, no-decreciente  
 mos (por conveniencia en las aplicaciones)  
 la condición  $\Phi(2t) \leq c \Phi(t)$ .

Sea  $1 \leq p \leq \infty$  y  $r$  la longitud del lado del  
 cubo  $Q$ ,

$$L^{p, \Phi} = \left\{ f \in L^1_{loc}(\mathbb{R}^n) / \|f\|_{*, \Phi} = \right. \\
 \left. \left( \sup_Q \frac{1}{\Phi(r)} \int_Q |f(x) - f_Q|^p dx \right)^{1/p} < \infty \right\} .$$

Identificando las funciones que difieren  
 en una constante, se obtiene la respectiva  
 norma, con la cual  $L^{p, \Phi}$  es un espacio de  
 Banach. Observemos que si  $\Phi(t) = r^n \varphi(t)$ ,  
 donde  $\varphi$  es una función positiva, no-decre-  
 ciente sobre  $(0, \infty)$ , entonces  $L^{1, \Phi} = BMO_\varphi$ .

### 3. Desigualdad de John-Nirenberg para $BMO_\varphi$ (Torchinsky [6]).

En esta ocasión consideramos  $\varphi(t)$ , una función continua, no-decreciente, de valor real tal que  $\varphi(0) = 0$ . Entonces,  $f \in BMO_\varphi$  si  $f \in L^1(Q_0)$  y

$$\|f\|_{*,\varphi} = \sup_{Q \subset Q_0} \frac{1}{\varphi(|Q|)} \frac{1}{|Q|} \int_Q |f(x) - f_Q| dx < \infty.$$

El objetivo es extender a  $BMO_\varphi$  la desigualdad de John-Nirenberg [+] de la sección 1. Bien, sea la función

$$\Psi(t) = \int_t^{2^n} \frac{\varphi(y)}{y} dy. \quad \text{Se tiene la}$$

**Proposición 1.** Sea  $f \in BMO_\varphi$ . Si  $\lim_{t \rightarrow 0^+} \Psi(t) < \infty$ , entonces

$$\sup_{x \in Q} \text{ess.} |f(x) - f_Q| \leq c \left( \int_0^{2^n |Q|} \frac{\varphi(y)}{y} dy \right) \|f\|_{*,\varphi}.$$

### Prueba.

Fijemos  $x \in Q \subseteq Q_0$ . Asumamos que  $x$  sea un punto de Lebesgue de  $f$ , esto es, tal que

$$\lim_{Q \rightarrow x} \frac{1}{|Q|} \int_Q |f(y) - f(x)| dy = 0.$$

Como  $f \in L^1_{loc}$  y casi todo punto de  $\mathbb{R}^n$  es un punto de Lebesgue de  $f$ , es suficiente asumir tal condición para  $x$ . Llamando  $Q_1 = Q$ ,  $Q_2$  es uno de los  $2^n$  subcubos congruentes en que es dividido  $Q_1$ . Este proceso es continuado sucesivamente, y de esta manera se obtiene una sucesión de subcubos (diádicos)  $\{Q_j\}$ , que converge a  $x$ . Además, por el teorema de diferenciación de Lebesgue,

$$\lim_{j \rightarrow \infty} f_{Q_j} = f(x) \text{ a.e.}$$

Pero,

$$\begin{aligned} |f_{Q_{j-1}} - f_{Q_j}| &\leq \frac{1}{|Q_j|} \int_{Q_j} |f(x) - f_{Q_{j-1}}| dx \\ &\leq (\text{desde que } |Q_{j-1}| = 2^n |Q_j|) \end{aligned}$$



$$\leq 2^n \varphi(|Q_{j-1}|) \|f\|_{*,\varphi} .$$

Entonces,

$$|f(x) - f_Q| \leq |f_{Q_2} - f_{Q_1}| + |f_{Q_3} - f_{Q_2}| + \dots + \\ + \lim_{j \rightarrow \infty} |f(x) - f_{Q_j}| =$$

$$= \sum_{j=1}^{\infty} |f_{Q_{j+1}} - f_{Q_j}| \leq 2^n \left( \sum_{j=1}^{\infty} \varphi(|Q_j|) \right) \|f\|_{*,\varphi}$$

(considerando que  $|Q| = 2^{(j-1)n} |Q_j|$ ) =

$$= 2^n \left( \sum_{j=1}^{\infty} \varphi \left( \frac{|Q|}{2^{(j-1)n}} \right) \right) \|f\|_{*,\varphi} .$$

Ahora probemos la desigualdad

$$\varphi \left( \frac{|Q|}{2^{kn}} \right) \leq \frac{1}{n \log 2} \int_{\frac{|Q|}{2^{kn}}}^{\frac{|Q|}{2^{(k-1)n}}} \frac{\varphi(y)}{y} dy , \quad k \geq 1. \quad [*]$$

En efecto,

$$\int_{\frac{|Q|}{2^{kn}}}^{\frac{|Q|}{2^{(k-1)n}}} \frac{\varphi(y)}{y} dy \geq \varphi\left(\frac{|Q|}{2^{kn}}\right) \int_{\frac{|Q|}{2^{kn}}}^{\frac{|Q|}{2^{(k-1)n}}} \frac{dy}{y} = \varphi\left(\frac{|Q|}{2^{kn}}\right) \log\left(\frac{1}{2^{-n}}\right)$$

$$= n \log 2 \varphi\left(\frac{|Q|}{2^{kn}}\right),$$

que implica [\*].

Luego,

$$|f(x) - f_Q| \leq 2^n \sum_{j=1}^{\infty} \left( \frac{1}{n \log 2} \int_{\frac{|Q|}{2^{kn}}}^{\frac{|Q|}{2^{(k-1)n}}} \frac{\varphi(y)}{y} dy \right).$$

•  $\|f\|_{*\varphi} \leq$  (desde que la suma es por bloques)

$$\leq 2^n \left( \frac{1}{n \log 2} \int_0^{2^n |Q|} \frac{\varphi(y)}{y} dy \right) \|f\|_{*\varphi},$$

donde observamos que  $\frac{1}{2^{(j-2)n}} \leq 2^n$  es equivalente a  $2^{2n} \leq 2^{n(j+1)}$ . Así se tiene la tesis con  $c = \frac{2^n}{n \log 2}$ . ■

**Caso.**  $\lim_{t \rightarrow 0^+} \Psi(t) = +\infty$ .

Si esto es el caso, se sabe que

$$\Psi(|x|) = \int_{|x|}^{2^n} \frac{\varphi(y)}{y} dy \in BMO_\varphi$$

(ver sección 2). Sea ahora la función

$$\Psi_\alpha(t) = \int_t^{2^{n_\alpha}} \frac{\varphi(y)}{y} dy, \quad 0 < \alpha < 1.$$

Entonces se tiene el

**Teorema.** Si  $\lim_{t \rightarrow 0^+} \Psi(t) = +\infty$ , entonces existe una constante  $c$  tal que para  $f \in BMO_\varphi$  y todo  $Q \subseteq Q_0$  se tiene

$$\omega(\lambda) \equiv |\{x \in Q / |f(x) - f_Q| > \lambda\}| \leq c \Psi_{|Q|}^{-1} \left( \frac{\lambda}{\|f\|_{*,\varphi}} \right).$$

**Prueba.**

La idea de usar la descomposición de Calderón-Zygmund. Asumamos  $f_Q = 0$  y  $\|f\|_{*,\varphi} = 1$ . (pues hacemos el cambio  $f$  por

$\frac{f-f_Q}{\|f\|_{*,\varphi}}$ , de ser necesario). Así la tesis es  $|\{x \in Q / |f(x)| > \lambda\}| \leq c_1 \Psi_{|Q|}^{-1}(c_2 \lambda)$ .

Al usar tal descomposición obtenemos sucesivas generaciones de sub cubos de  $Q$  en la forma siguiente. La primera generación de cubos es obtenida al nivel

$$\frac{1}{|Q|} \int_Q |f(x)| dx \leq \varphi(|Q|) \leq 2^n \varphi(|Q|),$$

obteniéndose cubos abiertos disjuntos  $Q_j^{(1)}$  tales que

(a)  $|f(x)| \leq 2^n \varphi(|Q|)$  a.e. sobre  $Q - \cup_j Q_j^{(1)}$ ;

(b)  $2^n \varphi(|Q|) \leq \frac{1}{|Q_j^{(1)}|} \int_{Q_j^{(1)}} |f(x)| dx \leq 4^n \varphi(|Q|)$

(c)  $\sum_j |Q_j^{(1)}| \leq \frac{1}{2^n \varphi(|Q|)} \sum_j \int_{Q_j^{(1)}} |f(x)| dx \leq$

$$\leq \frac{|Q|}{2^n \varphi(|Q|) |Q|} \int_Q |f(x)| dx$$

$$\leq \frac{1}{2^n} |Q| .$$

Ahora, en la segunda etapa, aplicamos la descomposición de Calderón-Zygmund a cada función  $(f(x) - f_{Q_j^{(1)}}) \chi_{Q_j^{(1)}}(x)$ , y obtenemos una segunda generación de subcubos. Así, fijemos  $Q_j^{(1)}$  y pongamos  $Q^{(1)} = \cup_j Q_j^{(1)}$ . Entonces, desde que

$$\frac{1}{|Q^{(1)}|} \int_{Q^{(1)}} |f(x) - f_{Q^{(1)}}| \leq \varphi(|Q^{(1)}|)$$

$$< 2^n \varphi(|Q^{(1)}|),$$

obtenemos los subcubos  $Q_j^{(2)}$  de  $Q^{(1)}$ , abiertos y disjuntos, tales que

(a)  $|f(x) - f_{Q^{(1)}}| \leq 2^n \varphi(|Q^{(1)}|)$  a.e. sobre

$$Q^{(1)} - \cup_j Q_j^{(2)}$$

$$\begin{aligned}
 \text{(b)'} \quad 2^n \varphi(|Q^{(1)}|) &\leq \frac{1}{|Q_j^{(2)}|} \int_{Q_j^{(2)}} |f(x) - f_{Q^{(1)}}| dx \\
 &\leq 4^n \varphi(|Q^{(1)}|)
 \end{aligned}$$

$$\begin{aligned}
 \text{(c)'} \quad \sum_j |Q_j^{(2)}| &\leq \frac{1}{2^n \varphi(|Q^{(1)}|)} \sum_j \int_{Q_j^{(2)}} |f(x) - f_{Q^{(1)}}| dx \\
 &\leq \frac{1}{2^n \varphi(|Q^{(1)}|)} \int_{Q^{(1)}} |f(x) - f_{Q^{(1)}}| dx \\
 &\leq \frac{|Q^{(1)}|}{2^n} .
 \end{aligned}$$

Desde que  $|Q^{(1)}| \leq \frac{|Q|}{2^n}$ , de (a)' y (b) obtenemos

$$\begin{aligned}
 |f(x)| &\leq |f(x) - f_{Q^{(1)}}| + \frac{1}{|Q^{(1)}|} \int_{Q^{(1)}} |f(x)| dx \\
 &\leq 4^n \varphi\left(\frac{|Q|}{2^n}\right) + 4^n \varphi(|Q|) ,
 \end{aligned}$$

esto es,

$$|f(x)| \leq 4^n \left( \varphi\left(\frac{|Q|}{2^n}\right) + \varphi(|Q|) \right) . \quad [+]'$$

Ahora, de (b)' y (c), sumando sobre todos los cubos de la primera generación  $Q^{(1)}$ , obtenemos

$$\sum_j |Q_j^{(2)}| \leq \frac{1}{2^n} \sum_j |Q_j^{(1)}| \leq \left(\frac{1}{2^n}\right)^2 |Q|. \quad [++]'$$

Y así podemos continuar... Asumiendo que tenemos seleccionado una  $(k-1)$ -geración de cubos  $Q^{(k-1)}$ , por el método usado antes, se selecciona una  $(k)$ -generación de Calderón-Zygmund cubos, tal que para cada cubo  $Q^{(k-1)}$  se tiene (inductivamente)

$$[+] \quad |f(x)| \leq 4^n \sum_{j=0}^{k-1} \varphi\left(\frac{|Q|}{2^{nj}}\right) \text{ a.e. sobre}$$

$$Q^{(k-1)} - \bigcup_j Q_j^{(k)}$$

$$[++] \quad \sum_j |Q_j^{(k)}| \leq \left(\frac{1}{2^n}\right)^k |Q| .$$

Ahora hagamos el siguiente argumento.

**Caso**  $\lambda > 2^n \varphi(|Q|)$  . Sea  $k$  el más grande entero tal que

$$4^n \sum_{j=0}^{k-1} \varphi\left(\frac{|Q|}{2^{nj}}\right) < \lambda . \text{ Entonces se tiene}$$

$$\begin{aligned} \{x \in Q / |f(x)| > \lambda\} &\subseteq \{x \in Q / 4^n \sum_{j=0}^{k-1} \varphi\left(\frac{|Q|}{2^{nj}}\right) \\ &< |f(x)|\} \subseteq \cup_j Q_j^{(k)} , \text{ por } [+]. \end{aligned}$$

Luego,

$$\begin{aligned} |\{x \in Q / |f(x)| > \lambda\}| &\leq \sum_j |Q_j(x)| \\ &\leq \left(\frac{1}{2^n}\right)^k |Q|. \quad [**] \end{aligned}$$

Ahora, por construcción de  $k$ ,

$$\lambda \leq 4^n \sum_{j=0}^k \varphi\left(\frac{|Q|}{2^{jn}}\right) \leq (\text{por } [*]) \leq \frac{4^n}{n \log 2} \sum_{j=0}^k$$



$$\int_{\frac{|Q|}{2^{jn}}}^{\frac{|Q|}{2^{(j-1)n}}} \frac{\varphi(y)}{y} dy \leq \frac{4^n}{n \log 2} \int_{\frac{|Q|}{2^{kn}}}^{\frac{|Q|}{2^n}} \frac{\varphi(y)}{y} dy =$$

$$C \Psi_{|Q|} \left( \frac{|Q|}{2^{kn}} \right).$$

Luego, usando {\*\*},

$$|\{x \in Q / |f(x)| > \lambda\}| \leq \frac{|Q|}{2^{kn}} \leq \Psi_{|Q|}^{-1}(c_2 \lambda),$$

que es la tesis para este caso.

**Caso**  $\lambda \leq 2^n \varphi(|Q|)$ . Por hipótesis, si  $\lambda \rightarrow 0^+$  entonces  $\Psi_{|Q|}^{-1}(\lambda) \rightarrow \infty$ , luego existe una constante  $c_1$  tal que  $|Q| \leq c_1 \Psi_{|Q|}^{-1}(\lambda)$ ,  $\lambda \leq 2^n \varphi(|Q|)$ . Por lo tanto,  $|\{x \in Q / |f(x)| > \lambda\}| \leq |Q| \leq c_1 \Psi_{|Q|}^{-1}(\lambda)$ . ■

#### 4. Casos particulares.

PROPOSICION 2. Sea  $\varphi(t) = \eta(\log(\frac{1}{t}))$  y  $\Phi'(t) = \eta(t)$ , donde  $\Phi$  es una función continua

(derivable), no-decreciente, con  $\Phi(0) = 0$ .

Si 
$$\Psi_{|Q|}(t) = \int_t^{2^n |Q|} \eta(\log(\frac{1}{y})) \frac{dy}{y},$$

entonces tenemos

$$\Psi_{|Q|}^{-1}(t) \leq C_1 e^{(-1/2)\Phi^{-1}(t)} |Q|.$$

**Prueba.** Tenemos

$$\begin{aligned} \Psi_{|Q|}(t) &= \int_{\frac{1}{t}}^{\frac{1}{2^n |Q|}} \eta(\log s) \left(-\frac{ds}{s}\right) = \int_{\frac{1}{t}}^{\frac{1}{2^n |Q|}} \eta(\log y) \frac{dy}{y} \\ &= \int_{\frac{1}{2^n |Q|}}^{\frac{1}{t}} \eta(\log y) d(\log y) = \Phi(\log(\frac{1}{t})) - \\ &\quad - \Phi(\log(\frac{1}{2^n |Q|})); \end{aligned}$$

esto es,

$$\Psi_{|Q|}(t) + \Phi(\log(\frac{1}{2^n |Q|})) = \Phi(\log(\frac{1}{t})).$$

Llamemos  $\Psi_{|Q|}(t) = t^*$  ; así,  $t = \Psi_{|Q|}^{-1}(t^*)$ .

$$\text{Luego } t^* + \Phi\left(\log\left(\frac{1}{2^n |Q|}\right)\right) = \Phi\left(\log\left(\frac{1}{\Psi_{|Q|}^{-1}(t^*)}\right)\right),$$

$$\begin{aligned} \text{o aun, reescribiendo, } \quad & \Phi\left(\log\left(\frac{1}{\Psi_{|Q|}^{-1}(t)}\right)\right) = \\ & = t + \Phi\left(\log\left(\frac{1}{2^n |Q|}\right)\right), \text{ de donde} \end{aligned}$$

$$\log\left(\frac{1}{\Psi_{|Q|}^{-1}(t)}\right) = \Phi^{-1}\left(t + \Phi\left(\log\left(\frac{1}{2^n |Q|}\right)\right)\right),$$

$$\text{o aún } \Psi_{|Q|}^{-1}(t) = e^{(t + \Phi(\log(\frac{1}{2^n |Q|})))}. \text{ Pero,}$$

desde que  $\Phi^{-1}$  es creciente, en general, si

$$a, b > 0 \text{ se tiene } \Phi^{-1}(a+b) \geq \frac{\Phi^{-1}(a) + \Phi^{-1}(b)}{2}.$$

Luego,

$$\Phi_{|Q|}^{-1}(t) \leq e^{\frac{\Phi^{-1}(t) + \log(\frac{1}{2^n |Q|})}{2}} = e^{\frac{1}{2} \Phi^{-1}(t)}.$$

$$e^{\frac{1}{2} \log(\frac{1}{2^n |Q|})}$$

Considerando que

$$e^{-\frac{1}{2} \log\left(\frac{1}{2^n |Q|}\right)} = \frac{1}{\left(\frac{1}{2^n |Q|}\right)^{1/2}} = 2^{n/2} |Q|^{1/2} \leq c_1 |Q|,$$

tenemos la tesis. ■

**Corolario 1.** Sea  $\eta(y) = 1$  y  $\varphi(t) = 1$ . Como  $\Phi$  es tal que  $\Phi'(t) = 1$ ,  $\Phi(t) = t$ , esto es,  $\Phi^{-1}(t) = t$ . Luego,

$$\Psi_{|Q|}^{-1}(t) \leq c_1 e^{-(1/2)\Phi^{-1}(t)} |Q| = c_1 e^{-(1/2)t} |Q|,$$

y por teorema tenemos

$$|\{x \in Q / |f(x) - f_Q| > \lambda\}| \leq c_1 \Psi_{|Q|}^{-1}\left(\frac{\lambda}{\|f\|_{*,\varphi}}\right) \leq c_1 e^{-(1/2)\lambda \|f\|_{*,\varphi}^{-1}} |Q|,$$

que es la desigualdad de John-Nirenberg para  $BMO$ .

**Corolario 2.** Asumamos  $\eta(y) = y^{-\varepsilon}$ ,  $0 < \varepsilon < 1$  ;  $\varphi(t) = (\log(1/t))^{-\varepsilon} = \frac{1}{(-\log t)^\varepsilon}$  .

Entonces  $\Phi'(t) = t^{-\varepsilon}$  ,  $\Phi(t) = \frac{t^{1-\varepsilon}}{1-\varepsilon}$  ,

$$\Phi^{-1}\left(\frac{t^{1-\varepsilon}}{1-\varepsilon}\right) = t . \quad \text{Si } \frac{t^{1-\varepsilon}}{1-\varepsilon} = s ,$$

$$t = (s(1-\varepsilon))^{1/1-\varepsilon} .$$

Entonces,

$$\Phi^{-1}(s) = (s(1-\varepsilon))^{1/1-\varepsilon} = c s^{1/1-\varepsilon} .$$

Luego,

$$\Psi_{|Q|}^{-1}(t) \leq c_2 e^{-(1/2)t^{1/1-\varepsilon}} \cdot |Q| ,$$

y por tanto se tiene la respectiva desigualdad de John-Nirenberg,

$$\omega(\lambda) \leq c e^{-(1/2)(\lambda \|f\|_{*,\varphi}^{-1})^{1/1-\varepsilon}} |Q| .$$

**Corolario 3.** (Stegenga). Sea  $\eta(y) = \frac{1}{y}$  y  $\varphi(t) = \frac{1}{\log(\frac{1}{t})} = -(\log t)^{-1}$ . Así  $\Phi'(t) = -t^{-1}$  ó  $\Phi(t) = \log(c \frac{1}{t})$ . Entonces  $\Phi^{-1}(t) = c e^{-t}$  y se tiene la correspondiente desigualdad de John - Nirenberg.

### Referencias.

- [1] Fefferman, Ch.-Stein, E.M.: "H<sup>p</sup> spaces of several variables". Acta Math. (1972), 137-193.
- [2] Janson, S.: "On functions with conditions on the mean oscillation". Arkiv. for Math (1976), 189-196.
- [3] John, F.-Nirenberg, L.: "On functions of bounded mean oscillation". Comm.P.Appl.Math.(1961), 415-426.
- [4] Neri, U. : "Some properties of functions with bounded mean oscillation". Studia Math.(1976), 63-75.

- [5] *Spanne, S.*: "Some function spaces defined using the mean oscillation over cubes". *Ann.Sup.Pisa.*(1965), 593-608.
- [6] *Torchinsky, A.* : *Comunicación personal*. Bloomington. (1984).







**PONTIFICIA UNIVERSIDAD CATOLICA DEL PERU**  
**FONDO EDITORIAL**

**REVISTAS**  
(Hasta ediciones 1990)

**EDICIONES ANUALES**

*Anthropologica*

Suscripción	US\$	8.00
Número suelto	US\$	10.00

Envío aéreo certificado:

América Latina	US\$	14.42
Canadá y U.S.A.	US\$	17.50
Europa	US\$	19.80
Asia y Africa	US\$	22.88

Envío vía superficie a cualquier destino:	US\$	5.00
---	------	------

*Boletín del Instituto Riva-Agüero*  
*Debates en Sociología*  
*Derecho*  
*Espacio y Desarrollo*

Suscripción	US\$	6.40
Número suelto	US\$	8.00

Envío aéreo certificado:

América Latina	US\$	7.78
Canadá y U.S.A.	US\$	9.32
Europa	US\$	10.47
Asia y Africa	US\$	12.00

Envío vía superficie a cualquier destino:	US\$	4.00
---	------	------

**EDICIONES SEMESTRALES**

*Areté*  
*Economía*  
*Histórica*  
*Lexis*  
*Pro Mathematica*  
*Revista de la Universidad Católica / Nueva serie*  
*Revista de Psicología*  
*Revista de Química*

Suscripción	US\$	12.80
Número suelto	US\$	8.00

Envío aéreo certificado:

América Latina	US\$	9.00	US\$	4.50
Canadá y U.S.A.	US\$	10.00	US\$	5.00
Europa	US\$	11.00	US\$	5.50
Asia y Africa	US\$	13.00	US\$	6.50

Envío vía superficie a cualquier destino:	US\$	4.00	US\$	2.00
---	------	------	------	------





**PONTIFICIA UNIVERSIDAD CATOLICA DEL PERU  
FONDO EDITORIAL**

**REVISTAS  
(A partir de ediciones 1991)**

**EDICIONES ANUALES**

*Anthropologica*

Suscripción	US\$	12.00
Número suelto	US\$	15.00

**Envío aéreo certificado:**

América Latina	US\$	14.42
Canadá y U.S.A.	US\$	17.50
Europa	US\$	19.80
Asia y Africa	US\$	22.88

<b>Envío vía superficie a cualquier destino:</b>	US\$	5.00
--	------	------

*Boletín del Instituto Riva-Agüero  
Debates en Sociología  
Derecho  
Espacio y Desarrollo*

Suscripción	US\$	9.60
Número suelto	US\$	12.00

**Envío aéreo certificado:**

América Latina	US\$	7.78
Canadá y U.S.A.	US\$	9.32
Europa	US\$	10.47
Asia y Africa	US\$	12.00

<b>Envío vía superficie a cualquier destino:</b>	US\$	4.00
--	------	------

**EDICIONES SEMESTRALES**

*Areté  
Economía  
Histórica  
Lexis  
Pro Mathematica  
Revista de la Universidad Católica / Nueva serie  
Revista de Psicología  
Revista de Química*

Suscripción	US\$	19.20
Número suelto	US\$	12.00

**Envío aéreo certificado:**

Suscripción	Número suelto
-------------	---------------

América Latina	US\$	9.00	US\$	4.50
Canadá y U.S.A.	US\$	10.00	US\$	5.00
Europa	US\$	11.00	US\$	5.50
Asia y Africa	US\$	13.00	US\$	6.50

<b>Envío vía superficie a cualquier destino:</b>	US\$	4.00	US\$	2.00
--	------	------	------	------



