LA TEORIA DE LA EXTRAPOLACION ESTADISTICA

Por el Dr. Norbert Wiener [1]

Presentado a la Sociedad Matemática Mexicana, el día 27 de abril de 1945.

Tengo mucho gusto en manifestar mi satisfacción por el honor que se me confiere con la invitación de la Sociedad Matemática Mexicana para presentar ante ella este trabajo.

El asunto que he escogido es la teoría de la extrapolación estadística, o del pronóstico. En una serie de sucesos bien determinados, no existe ningún problema de pronóstico; tanto el pasado como el futuro, tienen una determinación única. Es solamente cuando el pasado puede admitir varios futuros que aparece el problema del pronóstico.

Sin embargo, este problema no se presenta a menos que introduzcamos en la teoría un elemento nuevo. La multiplicidad de pronósticos posibles no ocasiona ningún futuro preferido si no tenemos alguna distribución de estas posibilidades, alguna medida de posibilidades. Una medida de posibilidades se llama un sistema de probabilidad.

Los sistemas de probabilidad son muy diversos; a pesar de ello algunos son particularmente naturales. Esos se llaman *estables,* y se distinguen de los otros en el hecho de que la medida de probabilidad no cambia cuando todos los valores del tiempo se aumentan igualmente.

En este caso, tenemos una familia S de series P de números P_v, que tienen índices v que van de menos infinito a más infinito. En la familia S se establece un sistema de medida de Lebesgue. Sea m (S_1) la medida o

---

(1) Del Massachusetts Institute of Technology, Cambridge, Mass., E. U. A.

probabilidad del conjunto S_1, y supongamos que m (S) sea 1. Conocemos que la probabilidad combinada de todos los sucesos debe ser *1*. La transformación de P_v en P_{v+1} engendra entonces una transformación de P en TP, y es claro que siempre será

$$m(S_1) = m(TS_1).$$

Estas condiciones son justamente aquellas a las cuales se aplica el teorema ergódico de Birkhoff. Este teorema dice que si f (P) es una función integrable sobre S, entonces existe con excepción de un conjunto de valores de P, de medida 0, el siguiente límite:

$$\lim_{n\to\infty} \frac{1}{n+1} [f(P) + f(TP) + \ldots + f(T^nP)] = f^*(P);$$

y este límite es una función integrable. Resulta también que:

$$\lim_{n\to\infty} \frac{1}{n+1} [f(P) + f(T^{-1}P) + \ldots + f(T^{-n}P)] = f^*(P),$$

con excepción de un conjunto semejante al anterior, de medida 0.

En el caso general, f(P) depende de P, a pesar de que

$$f^*(P) = f^*(TP).$$

Hay sin embargo, un caso en el cual el valor de $f^*(P)$ no depende en ninguna forma del valor de P. Este caso se llama *el caso ergódico*. En el caso ergódico, tenemos la fórmula:

$$f^*(P) = \int_S f(P)\, dV_p.$$

El caso ergódico es el más fundamental de todos. En los demás, según un teorema de von Neumann, es posible reemplazar de una manera esencialmente unívoca la medida original por una medida obtenida si se integra un sistema de medidas ergódicas. En consecuencia no es necesario considerar el caso no ergódico.

Sea $\{g_n(P)\}$ un arreglo de funciones L^2: es decir, de funciones medibles y de cuadrados integrables. Es bien conocido el procedimiento para ortogonalizar este arreglo. Así obtenemos un nuevo arreglo $\varphi_n(P)$ donde:

$$\int_S \varphi_m(P)\, \overline{\varphi_n(P)}\, dV_p = \begin{cases} 0 \text{ si } m \neq n; \\ 1 \text{ si } m = n; \end{cases}$$

$$\varphi_n(P) = \sum_{m=1}^{n} c_{mn}\, g_m(P). \qquad (c_{mn} \neq 0).$$

En este procedimiento, hay muchas integraciones de productos de funciones $g_n(P)$. Ahora bien, en vista del teorema ergódico, es casi siempre posible reemplazar cada integración

$$\int_S g_m(P)\, \overline{g_n(P)}\, dV_p$$

por el valor medio

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} g_m(T^{-k}P)\, \overline{g_n(T^{-k}P)}.$$

Cuando m y k tienen valores dados, el conjunto de valores de P para los cuales no es válida esta equivalencia es de medida cero. No hay más combinaciones de m y k que números naturales; y todos los puntos de tal conjunto de conjuntos de medida cero forman un conjunto igualmente de medida cero.

Una clase de funciones de P que tiene mucho interés, es la de funciones L^2 del pasado y de la actualidad. Estas funciones son: 1. Las funciones $F(P_0, P_{-1}, \ldots, P_{-\mu})$ donde $\mu < \infty$, y F es una función medible de P y de cuadrado integrable; 2. Todas las funciones de L^2 que son "límites en la media" de series de funciones 1. Según un teorema muy general, esta clase posee una base de la potencia de los números naturales en la cual se desarrollan todas las funciones de la clase. Si todas las funciones mencionadas pertenecen a L^2, las funciones monomias en $P_0, P_{-1}, \ldots$ forman tal clase.

Ortogonalizemos esta clase. Obtenemos las funciones $\varphi_n(P)$. Sea $f(P)$ una función de L^2; entonces:

$$f(P) = f_1(P) + f_2(P);$$

donde

$$f_1(P) \sim \sum_1^{\infty} \varphi_n(P) \int_S f(Q)\, \overline{\varphi_n(Q)}\, dV_Q,$$

y

$$\int_S f_2(P)\, \overline{\varphi_n(P)}\, dV_P = 0 \qquad (n = 1, 2, \ldots)$$

La función $f_1(P)$ se puede escribir:

$$f_1(P) \sim \sum_1^{\infty} \varphi_n(P) \lim_{\nu \to \infty} \frac{1}{\nu + 1} \sum_{k=0}^{\nu} f(T^{-k}Q)\, \overline{\varphi_n(T^{-k}Q)}.$$

Esta es siempre una función de lo pasado y de la actualidad. Se obtiene como límite de algunos polinomios en P_0, P_{-1}, ... Es decir, las operaciones con las cuales se obtiene de P_0 son: (1), la generación de una demora, que cambie P_0 en P_{-k}; (2), la multiplicación de estas cantidades retardadas; (3), la adición de las cantidades (2); (4), el cálculo del límite. Con excepción de (4), todos estos procedimientos se pueden realizar con aparatos eléctricos. Se conocen muy bien los circuitos de demora, así como la combinación aditiva de voltajes o corrientes. La combinación multiplicativa es también posible, si poseemos un aparato que transforme un voltaje o una corriente en su cuadrado, porque

$$xy = 1/4\, [(x + y)^2 - (x - y)^2].$$

Existen tales aparatos; lo que falta ahora es la construcción de aparatos con propiedades exactas y reproducibles. El problema no es más difícil que muchos otros que han resuelto los ingenieros. Es decir que podemos obtener $f_1(P)$ del pasado y de la actualidad de P, cuando se presenta como una serie de voltajes en función del tiempo, si empleamos un aparato eléctrico que podemos idear en una forma bien determinada y precisa. Partimos de la estadística de la serie P_ν, donde ν debe ser un parámetro que represente el tiempo, y de la suposición de que el procedimiento tiene un equilibrio estadístico.

No es difícil demostrar que entre todas las funciones del pasado y la actualidad, f_1 es aquella para la cual

$$\int_S | f_1(P) - f_2(P) |^2\, dV_P$$

es un mínimo, y que el valor medio de $f(P)$ cuando conocemos el pasado y la actualidad de P es $f_1(P)$. Como consecuencia de esto, podemos determinar el valor medio de $e^{i\lambda f(P)}$ cuando son determinados el pasado y la actualidad. Sea este valor medio $F(\lambda)$ en función de λ. Entonces es $F(\lambda)$ la función

$$\int_{-\infty}^{\infty} e^{i\lambda u}\, d\varphi(u),$$

donde $\varphi(u_2 - 0) - \varphi(u_1 - 0)$ es la probabilidad que $f(P)$ esté entre u_1 y u_2. Esta función $\varphi(u)$ de distribución es todo lo que podemos saber de $f(P)$, cuando conocemos el pasado y la actualidad de P. Repito, lo sabemos sólo con probabilidad uno, no con certeza completa.

Si $f(P)$ es P_ν $(\nu \geqq 1)$, hemos resuelto el problema del pronóstico estadístico, por una computación en el caso en el cual se presentan los datos lentamente, y por una máquina apropiada en el caso en el cual no podemos seguir los datos con lápiz y papel. En ambos casos es necesario un conocimiento de todo el pasado. Es claro que puede cambiarse este método por un método estadístico práctico cuando tan sólo conocemos el pasado después de una época dada. Sin embargo, esta teoría no ha sido desarrollada todavía.

Hasta ahora hemos considerado la teoría de series sencillas de sucesos: es decir, de series en las cuales en un tiempo dado no se determina más que un número. Si reemplazamos cada número por un par de números, o por un arreglo ordenado de números no se cambia nada de importancia en la teoría. En un caso particular, supongamos que se determinan en cada instante dos cantidades, y que la primera representa la intensidad de un mensaje (por ejemplo, un mensaje telefónico), la segunda la intensidad de un ruido. Entonces tenemos una distribución estadística simultánea de dos series: P_ν, el mensaje, y Q_ν, el ruido. Sea R_ν una combinación de los pasados y las actualidades de P_ν y Q_ν que transforme R_ν en $R_{\nu+\lambda}$, la misma operación que transforma P_ν en $P_{\nu+\lambda}$ y Q_ν en $Q_{\nu+\lambda}$. Entonces el problema de determinar el valor óptimo de $P_{\nu\mp\mu}$ cuando conocemos todos los valores de R_σ para $\sigma \leqq \nu$, es un problema muy importante en la electrotécnica. Es la extensión en una región no lineal, del problema conocido de filtración eléctrica. Puede ser muy importante para la modulación de frecuencia y para otras cuestiones técnicas.

Los métodos que se desarrollan aquí son puramente operacionales si empleamos esta palabra con el mismo significado que han introducido los físicos en la teoría de los quanta. Los problemas de diseño de un aparato telefónico, no tienen una solución perfecta; sólo tienen una solución óptima cuando conocemos la naturaleza estadística de los mensajes y de los ruidos. En este caso el problema de diseño no es más que una aplicación del cálculo de variaciones. Obtenemos por una vía única las características matemáticas del aparato. En seguida empleamos métodos ya bien conocidos para realizar el aparato "en el metal", como decimos en Estados Unidos. Este es un método de inventar, radicalmente distinto de los que suponen las leyes de patentes.