REGRESION LINEAL APLICADA A PROBLEMAS DE DIFUSION

E.Santos, E.Marengo y F.Dyment

Departamento de Metalurgia
Comisión Nacional de Energía Atómica

R E S U M E N

Se hace una revisión general de los métodos de cálculo de la "mejor recta" por cuadrados mínimos con error en ambas coordenadas y se sugiere una expresión para atribuir peso a datos provenientes de experiencias destructivas. Una subrutina FORTRAN que calcula la pendiente y la ordenada al origen con sus errores es aplicada en cuatro programas para obtener los parámetros que interesan en procesos de difusión. Se agrega, además, una subrutina que grafica con una impresora de línea, la mejor recta calculada y los datos experimentales.

Este trabajo está disponible como publicación interna de CNEA; El presente es sólo una breve información sobre su contenido.

INTRODUCCION

En muchas ramas de la ciencia el investigador debe enfrentar frecuentemente el problema de trazar la "mejor recta" a través de un conjunto de datos experimentales.

Cuando se trata de aplicar el método de cuadrados mínimos surgen dos dificultades: 1°) Si los pares de puntos(x_i, y_i) de valores observados no están afectados por errores similares, debe asignarse un peso estadístico a cada punto en el análisis de regresión; 2°) en experiencias destructivas o difíciles de repetir bajo condiciones constantes, el peso no puede calcularse y depende de criterios subjetivos del investigador: su experiencia, conocimiento del sistema e instrumentos, y, esencialmente, su "buen sentido".

Error estadístico en ambas coordenadas

El ajuste de una recta considerando errores en ambas variables en experiencias no destructivas ya ha sido estudiado[1-3]. Se obtiene un estimador de la varianza en cada punto.

Si las variables x e y verifican la relación funcional F(X,Y,Bk), donde Bk son constantes para determinar el peso de cada punto (X_i,Y_i) está dado por la expresión:

$$\frac{1}{w_i} = \frac{1}{w_{y_i}}\left(\frac{\partial F}{\partial Y}\right)_i^2 + \frac{1}{w_{x_i}}\left(\frac{\partial F}{\partial X}\right)_i^2$$

-2

donde

$$W_{x_i} = \frac{\sigma_1}{\sigma_{x_i}^2} \quad ; \quad W_{y_i} = \frac{\sigma_2}{\sigma_{y_i}^2}$$

σ_1, σ_2 son constantes de proporcionalidad.

La regresión lineal se calcula de acuerdo a[3] iterando sobre una ecuación "lineal"

Este método se aplica en la subrutina CUAMI, la cual itera hasta que 2 valores consecutivos de la pendiente difieren entre sí en menos de la tolerancia admitida, consistente con las condiciones del problema (para nosotros ese valor se fijó en 0.01). Además de su rápida convergencia, el método presenta otra ventaja importante sobre los casos en que se itera sobre una ecuación "cúbica": es innecesario estimar un valor inicial de la pendiente ya que suponiéndola nula en el comienzo, los valores obtenidos en la primera iteración corresponden a la pendiente y ordenada al origen cuando la variable x ha sido medida con error despreciable. El número de iteraciones en la subrutina es, por consiguiente, un índice de cuán significativos son los errores en la coordenada x comparados con los de la coordenada y.

Asignación de peso cuando no hay una distribución de puntos para cada valor de la variable

Nuestro caso es el de experiencias destructivas donde cada medida tiene un error que no puede ser estimado estadísticamente. En tal situación la experiencia personal del investigador, su conocimiento del sistema en estudio, del instrumental empleado y fundamentalmente, su juicio personal, determinan los criterios de asignación de pesos.

Para eliminar, tanto como fuese posible, la subjetividad de estas apreciaciones, se ha usado un criterio que se considera razonable si bien no riguroso desde el punto de vista matemático.

Cuando se cuenta con un conjunto de mediciones para cada valor de las variables $[(x_i^k, y_i^k) ; k = 1, \cdots, p_i]$ $\forall i$, se está en condiciones de calcular la varianza de la distribución para cada valor $\underline{i}$. Su inversa es el "peso relativo de ese i o valor con respecto al total de valores medidos en la regresión lineal. (Los valores con mayor varianza, o sea aquellos en que la dispersión es mayor "pesan" menos en el cómputo de la recta).

Por analogía, si se puede calcular de algún modo el error absoluto en la medición experimental de cada valor (error apreciación del instrumental, propagación de errores, por ejemplo), se adopta como aproximación sensata al "peso" la inversa del cuadrado de este error.

Códigos de Computación

Lo expuesto se ha aplicado al cálculo de coeficientes de difusión y parámetros de su proceso de difusión: energía de activación y factor de frecuencia.

Se han desarrollado 4 programas que hacen uso de la subrutina CUAMI.

RESIDUAL y DIRECTO calculan el coeficiente de difusión que resulta de la resolución de la ecuación:

$$\frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial x^2}$$

con la condición inicial:

$$c(x,0) = C_o\, \delta(x)$$

y las condiciones de contorno:

$$\left.\frac{\partial c}{\partial x}\right|_{x=0} = 0 \qquad ; \qquad \lim_{x \to \infty} c(x,t) = C$$

RESIDUAL se aplica cuando se usa el método de Gruzin y DIRECTO, cuando se emplea el seccionamiento directo.

LIMITE calcula los coeficientes de difusión en borde de grano por los métodos de Fisher y Suzuoka y ENERGIA calcula la energía de activación y el factor de frecuencia de un proceso de difusión.

Las subrutinas CUAMI y CURVA (subrutina de graficación) y los 4 códigos de difusión se han escrito en FORTRAN IV.

Todos ellos están disponibles para quien los requiera.

REFERENCIAS

1. Deming W.E. Sample Design in Business Research. J.Wiley, 1960.

2. York D., Can. J. Phys. 44, 1079 (1966).

3. Williamson J.H., Can. J. Phys. 46, 1845 (1968)