

ANÁLISIS DE REGRESIÓN NO LINEAL: 1-TÉCNICAS PARA DETERMINAR PARÁMETROS

Javier Armijo C.

Facultad de Química e Ingeniería Química, Universidad Nacional Mayor de San Marcos

RESUMEN

En esta primera parte se revisan las técnicas comúnmente usadas en la determinación de parámetros de modelos matemáticos no lineales.

Palabras claves: Análisis, regresión, no lineal, modelos, parámetros.

ABSTRACT

In this first part we shall describe several techniques which have been effectively employed to estimate nonlinear model parameter.

Key words: Analysis, regression, nonlinear, models, parameters.

INTRODUCCIÓN

El análisis de regresión es la aplicación de métodos matemáticos-estadísticos al análisis de datos experimentales y el ajuste de modelos matemáticos a dichos datos mediante la estimación de parámetros desconocidos del modelo.

Desde el punto de vista del análisis de regresión, los modelos se clasifican en modelos lineales y modelos no lineales. Un modelo es lineal si tiene una combinación lineal de los parámetros como por ejemplo

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 \ln x_1$$

Donde β_j son los parámetros, x_i son las variables independientes o controladas y η es la variable dependiente o la variable respuesta. Un modelo no lineal en los parámetros y en la variable independiente es por ejemplo:

$$\eta = \sqrt{\beta_0 + \beta_1 x_1 + \beta_2 x_2}$$

Así, un modelo es lineal con respecto a los parámetros si la primera derivada parcial de la variable dependiente respecto a cada uno de los parámetros no es una función de los parámetros.

La determinación de parámetros óptimos de un modelo, por ajuste con datos experimentales, consiste en minimizar la función suma de los cuadrados, es decir, buscamos

$$\text{Minimizar } \phi = \sum_{i=1}^n w_i [Y_i - \eta_i(x_k, \beta_j)]^2 \quad (1)$$

Donde Y_i son los datos experimentales y η representa al modelo no lineal:

$$\eta = \eta(x_1, x_2, \dots, x_q, \beta_1, \beta_2, \dots, \beta_m)$$

En la ecuación 1, w_i representa la calidad o peso de los datos experimentales (varía entre 0 y 1).

El análisis de regresión supone:

- 1) El valor observado de Y_i es aleatorio alrededor de η_i . Se denomina ecuación de regresión:

$$Y_i = \eta_i + \varepsilon_i$$

- 2) La variable independiente, x_k está libre de error. Ambas variables x , Y son continuas.
- 3) Los errores ε_i son variables aleatorias con una media esperada alrededor de cero, de varianza constante y estadísticamente son independientes.
- 4) La varianza de ε_i es la misma que la varianza de Y_i .
- 5) Los errores ε_i tienen una distribución normal. Esta suposición convierte a cualquier técnica de ajuste en el método de los mínimos cuadrados.

En general se consideran dos tipos de errores. Uno es el error en la *medición* de la variable dependiente; el otro es el error en la *forma* del modelo. Si los errores de medición y del modelo están presentes en un experimento, ε debe representar la combinación de ambos efectos.

TÉCNICAS DE OPTIMIZACIÓN

La solución de la ecuación (1) es un problema de optimización de funciones donde los parámetros β_j son las variables. Las técnicas de optimización pueden dividirse en dos grupos, los que no usan derivadas y las que usan derivadas. Entre las que no utilizan derivadas tenemos el método *Simplex*. Entre los que utilizan derivadas tenemos el método de *Gauss-Siedel* (o *Newton-Raphson*), el método de la *gradiente* y el método de *Marquardt*.

Dichas técnicas utilizan procedimientos iterativos cuya estrategia general es la siguiente:

- 1) Supone valores iniciales para el vector de parámetros β , llamado también punto base. Donde β son los estimados de los verdaderos parámetros β .
- 2) Se calcula la dirección de exploración B , que debe seguirse desde el punto base, en busca del mínimo de la función objetivo ϕ .
- 3) Selecciona la distancia τ , que se debe recorrer en la dirección calculada en el

paso anterior. Este se le considera como un factor de aceleración y es un escalar cuyo valor oscila entre 0 y 1.

- 4) El nuevo vector de parámetros de la siguiente iteración ($k+1$) se calcula como sigue:

$$b^{k+1} = b^k + t B^k \quad (2)$$

El método Simplex de Nelder y Mead

Este método está basado en la comparación de los valores de la función objetivo en ($m+1$) vértices de un simplex general y moviendo el simplex hacia el punto óptimo. Este movimiento es conseguido a través de tres operaciones básicas: reflexión, contracción y expansión.

Sea el vector columna de los estimados iniciales de los parámetros $(b_1, b_2, \dots, b_m)^T$. Luego, los ($m+1$) vértices del poliedro son:

$$\begin{aligned} v_1 &= (b_1, b_2, \dots, b_m)^T \\ v_2 &= (\pi_1 + b_1, \pi + b_2, \dots, \pi + b_m)^T \\ v_3 &= (\pi + b_1, \pi_1 + b_2, \pi + b_3, \dots, \pi + b_m)^T \\ &\dots \\ v_{m+1} &= (\pi + b_1, \pi + b_2, \pi + b_3, \dots, \pi_1 + b_m)^T \end{aligned}$$

Donde

$$\pi_1 = \left[\frac{\sqrt{m+1} - 1 + m}{m\sqrt{2}} \right] S, \pi = \left[\frac{\sqrt{m+1} - 1}{m\sqrt{2}} \right] S$$

y S es un factor de escala.

La función objetivo ϕ de la ecuación (1) se calcula en cada uno de los vértices y de este conjunto determinamos ϕ_{mayor} , $\phi_{\text{segundo mayor}}$, ϕ_{menor} , así como los correspondientes parámetros b_j .

Las tres operaciones básicas del método son:

reflexión(r)

$$b_j^r = (1 + \alpha) b_j^{\text{centroide}} - \alpha b_j^{\text{mayor}}, \quad \alpha > 0 \quad (3)$$

expansión(e)

$$b_j^e = \gamma b_j^r + (1-\gamma) b_j^{\text{centroide}}, \quad \gamma > 1 \quad (4)$$

contracción(c)

$$b_j^c = \beta b_j^{\text{mayor}} + (1-\beta) b_j^{\text{centroide}}, \quad 0 < \beta < 1 \quad (5)$$

Donde:

$$b_j^{\text{centroide}} = \frac{1}{m} \sum_1^{m+1} b_j \quad (6)$$

La sumatoria no incluye al vector de parámetros que corresponde al ϕ_{mayor} . Los coeficientes toman los siguientes valores $\alpha = 1$, $\gamma = 2$, $\beta = 0.5$. El criterio para terminar la búsqueda es:

$$\sqrt{\frac{\sum (\phi_i - \bar{\phi})^2}{m}} \leq \tau \quad (7)$$

$\bar{\phi}$: es el promedio de todos los ϕ

donde τ es un número pequeño seleccionado.

Los pasos del método son:

- 1) Se calculan los vértices $\mathbf{b}^{\text{mayor}}$, $\mathbf{b}^{\text{segundo mayor}}$, $\mathbf{b}^{\text{menor}}$ y el centroide $\mathbf{b}^{\text{centroide}}$. La prueba de convergencia se verifica según la ecuación (7).
- 2) El vector $\mathbf{b}^{\text{mayor}}$ es reflejado y se calcula el valor de $\phi^r(\mathbf{b}^r)$.
- 3) Si $\phi^{\text{segundo mayor}} \geq \phi^r \geq \phi^{\text{menor}}$, luego $\mathbf{b}^{\text{mayor}}$ es reemplazado por \mathbf{b}^r y el proceso se repite desde el paso 1.
- 4) Si $\phi^r(\mathbf{b}^r) < \phi^{\text{menor}}$, se expande el simplex en la dirección $\mathbf{b}^r - \mathbf{b}^{\text{centroide}}$. La expansión continua si es que $\phi^e < \phi^{\text{menor}}$ y en este caso $\mathbf{b}^{\text{mayor}}$ es reemplazado por \mathbf{b}^e . Si no fuera el caso se reemplaza $\mathbf{b}^{\text{mayor}}$ por \mathbf{b}^r .
- 5) Si la reflexión del paso 2 produce un vector \mathbf{b}^r tal que $\phi^{\text{mayor}} > \phi^r > \phi^{\text{segundo mayor}}$, se reemplaza $\mathbf{b}^{\text{mayor}}$ por \mathbf{b}^r y procedemos a la contracción. La contracción también se aplica si $\phi^r \geq \phi^{\text{mayor}}$.
- 6) Si $\phi^{\text{mayor}} > \phi^c$, luego $\mathbf{b}^{\text{mayor}}$ es reemplazado por \mathbf{b}^c y el proceso se reinicia desde el

paso 1. Si por otro lado, $\phi^{\text{mayor}} \leq \phi^c$ el simplex actual se quiebra alrededor del punto \mathbf{b}^i que se calcula como sigue:

$$\mathbf{b}^i = 0.5(\mathbf{b}^i + \mathbf{b}^{\text{menor}}), \quad i=1,2,\dots,m+1$$

y comenzamos en el paso 1.

El método de Gauss-Siedel

Este es uno de los métodos que requiere calcular derivadas de forma analítica o numérica. El método, denominado también de Newton-Rahpson, consiste en linealizar el modelo a través de la serie de Taylor trunca da para luego resolver un sistema de ecuaciones lineales. A través de un proceso iterativo, el método estima nuevos parámetros que minimizan la función de la ecuación (1).

Sea el modelo:

$$\eta = \eta(x_1, x_2, \dots, x_q, \beta_1, \beta_2, \dots, \beta_m)$$

Se expande el modelo al truncar el modelo a solo los dos primeros términos:

$$\eta \cong \eta_0 + \sum_{j=1}^m \left(\frac{\partial \eta}{\partial \beta_j} \right) (\beta_j - b_j^0)$$

Reemplazando esta ecuación en la ecuación (1) obtenemos:

$$\phi \cong \sum_{i=1}^n w_i \left[Y_i - (\eta_i)_0 - \sum_{j=1}^m \left(\frac{\partial \eta}{\partial \beta_j} \right) (\beta_j - b_j^0) \right] \quad (8)$$

$$\eta_0 = \eta(x_1, x_2, \dots, x_q, b_1^0, b_2^0, \dots, b_m^0)$$

$$\left(\frac{\partial \eta}{\partial \beta_j} \right)_0 = \frac{\partial \eta(x_1, x_2, \dots, x_q, b_1^0, b_2^0, \dots, b_m^0)}{\partial \beta_j}$$

Donde $b_1^0, b_2^0, \dots, b_m^0$ son los estimados iniciales de los parámetros.

Se deriva la ecuación 8 respecto a cada parámetro e igualamos a cero. Se obtiene un sistema de ecuaciones que después de acomodar y empleando notación matricial resulta en la forma siguiente:

$$(\mathbf{X}^T \mathbf{w} \mathbf{X})^0 \mathbf{B}^0 = (\mathbf{X}^T \mathbf{w} \mathbf{E})^0 \quad (9)$$

Despejando **B**:

$$\mathbf{B}^0 = [(\mathbf{X}^T \mathbf{w} \mathbf{X})^0]^{-1} (\mathbf{X}^T \mathbf{w} \mathbf{E})^0 \quad (10)$$

Donde:

$$X = \begin{bmatrix} \frac{\partial \eta_1}{\partial \beta_1} & \frac{\partial \eta_1}{\partial \beta_2} & \dots & \frac{\partial \eta_1}{\partial \beta_m} \\ \frac{\partial \eta_2}{\partial \beta_1} & \frac{\partial \eta_2}{\partial \beta_2} & \dots & \frac{\partial \eta_2}{\partial \beta_m} \\ \dots & \dots & \dots & \dots \\ \frac{\partial \eta_n}{\partial \beta_1} & \frac{\partial \eta_n}{\partial \beta_2} & \dots & \frac{\partial \eta_n}{\partial \beta_m} \end{bmatrix}_{n \times m} \quad B = \begin{bmatrix} \beta_1 - b_1 \\ \beta_2 - b_2 \\ \dots \\ \beta_m - b_m \end{bmatrix}_{m \times 1}$$

$$w = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & w_n \end{bmatrix}_{n \times n} \quad E = \begin{bmatrix} Y_1 - \eta_1 \\ Y_2 - \eta_2 \\ \dots \\ Y_n - \eta_n \end{bmatrix}_{n \times 1}$$

En la matriz X, las derivadas se evalúan para cada valor de las variables independientes correspondientes a un dato experimental $i = 1, 2, \dots, n$. \mathbf{X}^T es la matriz transpuesta de **X**.

Un problema de este método es que la matriz $(\mathbf{X}^T \mathbf{w} \mathbf{X})^0$ puede ser singular (vectores columna linealmente independientes).

Después de determinar el vector de exploración **B** de la ecuación (10), los siguientes valores del vector de parámetros **b** se calculan de la ecuación (2) y se continúa con el esquema iterativo hasta alcanzar cierta tolerancia prefijada para el valor de la función ϕ calculada de la ecuación 1.

El método de la gradiente

A diferencia del método de Gauss-Siedel, este método linealiza la ecuación (1)

$$\phi \cong \phi_0 + \sum_{j=1}^m \left(\frac{\partial \phi}{\partial \beta_j} \right)_0 (\beta_j - b_j^0)$$

El vector gradiente de ϕ se define como sigue:

$$\nabla \phi = \frac{\partial \phi}{\partial \beta_1} \delta_1 + \frac{\partial \phi}{\partial \beta_2} \delta_2 + \dots + \frac{\partial \phi}{\partial \beta_m} \delta_m$$

donde $\delta_1, \delta_2, \dots, \delta_m$ son los vectores unitarios en las direcciones de los parámetros $\beta_1, \beta_2, \dots, \beta_m$ respectivamente. El vector gradiente evaluado en el vector \mathbf{b}^0 es equivalente al término de derivada que aparece en la ecuación de expansión.

El negativo del vector gradiente direcciona la búsqueda hacia el valor mínimo de la función objetivo, es por eso que a este método también se le conoce como el de **descenso brusco**.

El método recomienda construir el vector unitario como sigue:

$$\frac{\nabla \phi}{\|\nabla \phi\|} = \frac{\frac{\partial \phi}{\partial \beta_1} \delta_1 + \frac{\partial \phi}{\partial \beta_2} \delta_2 + \dots + \frac{\partial \phi}{\partial \beta_m} \delta_m}{\sqrt{\left(\frac{\partial \phi}{\partial \beta_1}\right)^2 + \left(\frac{\partial \phi}{\partial \beta_2}\right)^2 + \dots + \left(\frac{\partial \phi}{\partial \beta_m}\right)^2}} \quad (11)$$

Los componentes negativos del vector unitario evaluado en \mathbf{b}^0 determinan al vector de exploración **B**, en dirección del mínimo, así por ejemplo:

$$B_j = \beta_j - b_j^0 = - \frac{\left(\frac{\partial \phi}{\partial \beta_j}\right)_0}{\sqrt{\sum \left(\frac{\partial \phi}{\partial \beta_j}\right)^2}}$$

Se calcula todos los B_j y se determina el vector **B**, el mismo que se reemplaza en la ecuación 2 y se continúa la iteración.

Al inicio se avanza rápidamente hacia el mínimo, pero luego la velocidad de convergencia se hace lenta. El negativo de la gradiente de ϕ apunta en la dirección que minimiza ϕ solamente en una región local y no en dirección de un mínimo global de ϕ .

Método de Marquardt

El método de Marquardt es una combinación del método de Gauss-Siedel y del método de la gradiente. En la ecuación $(\mathbf{X}^T \mathbf{w} \mathbf{X})^0 \mathbf{B}^0 = (\mathbf{X}^T \mathbf{w} \mathbf{E})^0$ se demuestra que el término del lado derecho es igual a:

$$X^T wE = -\frac{1}{2} \begin{bmatrix} \frac{\partial \phi}{\partial \beta_1} \\ \frac{\partial \phi}{\partial \beta_2} \\ \vdots \\ \frac{\partial \phi}{\partial \beta_m} \end{bmatrix}_{m \times 1} = -\frac{1}{2} \nabla \phi$$

Luego tenemos $2(X^T wX)B = -\nabla \phi^T$.

Marquardt introduce el factor t (tamaño del paso) en la ecuación anterior y resulta:

$$(2X^T wX + tI)B = -\nabla \phi^T$$

Donde I es la matriz identidad. Cuando $t = 0$ tenemos el método de Gauss - Siedel, y si t tiende al infinito tenemos al método de la gradiente.

El método de Marquardt busca la solución de la ecuación 9 transformándola a la forma:

$$(A + tI)B = C \quad (12)$$

donde $A = X^T wX$ y $C = X^T wE$.

Para efectos de cálculo se recomienda multiplicar la ecuación 12 por la matriz diagonal D , cuyos elementos están dados por:

$$d_{ii} = (a_{ii})^{-1/2} \quad (13)$$

La ecuación 12 queda ahora:

$$\begin{aligned} D(A + tI)DD^{-1}B &= DC \\ (DAD + tD^2)D^{-1}B &= DC \\ (A^* + tD^2)B^* &= C^* \end{aligned} \quad (14)$$

Donde los elementos de las matrices A^* , B^* y C^* son respectivamente:

$$a_{ij}^* = d_{ii} d_{jj} (a_{ij}) \quad (15)$$

$$c_j^* = d_{jj} c_j \quad (16)$$

$$b_j^* = b_j / d_{jj} \quad (17)$$

El valor de t se selecciona de la siguiente manera:

Sea $\nu > 1$ y t^{k-1} el valor de t en la iteración anterior (valor inicial $t^0 = 0.01$). Se calcula $\phi(t^{k-1})$ y $\phi(t^{k-1}/\nu)$, luego:

1. Si $\phi(t^{k-1}/\nu) \leq \phi^k$, luego se hace $t^k = t^{k-1}/\nu$.
2. Si $\phi(t^{k-1}/\nu) > \phi^k$ y $\phi(t^{k-1}) \leq \phi^k$, luego se hace $t^k = t^{k-1}$.
3. Si $\phi(t^{k-1}/\nu) > \phi^k$ y $\phi(t^{k-1}) > \phi^k$ se incrementa t por multiplicaciones sucesivas con ν hasta que $\phi(t^{k-1}\nu) \leq \phi^k$. Luego se hace $t^k = \nu t^{k-1}$.

La ecuación 12 o la ecuación 14 deben resolverse para determinar el vector de exploración B o B^* respectivamente. Con la ecuación 2 se calcula los nuevos parámetros para la siguiente iteración.

ANÁLISIS ESTADÍSTICO

Además de estimar los parámetros del modelo, debe obtenerse alguna medida de la dispersión de los parámetros estimados y de la dispersión del valor de la variable dependiente Y .

Intervalo de confianza de los parámetros

El intervalo de confianza se estima de la siguiente ecuación:

$$b_k - t_s s_Y \sqrt{a_{ii}} \leq \beta_k \leq b_k + t_s s_Y \sqrt{a_{ii}} \quad (18)$$

Donde:

a_{ii} son los elementos de la diagonal principal de la matriz A ,

b_k es el k -ésimo parámetro estimado del parámetro verdadero β_k ,

t_s es la variable aleatoria t -Student y se determina de tablas para $(n - m)$ grados de libertad y 97.5% de confianza,

s_Y representa la desviación estándar entre el valor experimental repetido y el valor del modelo, de la variable dependiente. Para estimar s_Y se necesita conocer la desviación estándar de los errores experimentales, si este no se conoce se reemplaza s_Y por la desviación estándar de los residuales s_r :

$$s_r^2 = \frac{E^T wE}{n - m} = \frac{\phi}{n - m} \quad (19)$$

Donde s_r es una medida de la falta de ajuste del modelo a los datos experimentales.

La matriz de los coeficientes de correlación

Los intervalos de confianza estimados con la ecuación 18 son válidos para cada parámetro de forma independiente siempre y cuando éstos no estén correlacionados. Usualmente los parámetros están correlacionados, siendo necesario cuantificar dicha correlación a través de la matriz de correlación \mathbf{R} cuyos elementos son:

$$r_{ij} = \frac{a_{ij}}{\sqrt{a_{ii}a_{jj}}} \quad (20)$$

Donde a_{ij} son los elementos de la matriz \mathbf{A}^{-1} :

$$\mathbf{A}^{-1} = (\mathbf{X}^T \mathbf{w} \mathbf{X})^{-1}$$

La matriz \mathbf{R} es simétrica y sus elementos r_{ij} tienen valores en el rango de -1 a 1, y representa la correlación existente entre los parámetros b_i y b_j . Cuanto más próximo a ± 1 es el valor de r_{ij} , mayor es la correlación entre los dos parámetros, lo que implica que es muy difícil de obtener estimados independientes de los parámetros con los datos experimentales disponibles. Una correlación negativa entre dos parámetros indica que los errores que originan que el estimado de uno de ellos sea elevado, también originan que el estimado del otro sea bajo.

La región conjunta de confianza de los parámetros

Los intervalos de confianza calculados de la ecuación 18 representan rangos individuales e independientes. Normalmente los parámetros están, en alguna medida, correlacionados. Es decir no es posible determinar el valor de un parámetro sin que afecte al otro. La correlación entre los parámetros se observa de la construcción de la región de confianza conjunta. El lugar geométrico que encierra esta región se estima de la ecuación 21:

$$[(\beta_1 - b_1) \dots (\beta_m - b_m)] \mathbf{X}^T \mathbf{w} \mathbf{X} \begin{bmatrix} \beta_1 - b_1 \\ \dots \\ \beta_m - b_m \end{bmatrix} = s_Y^2 m F_{1-\alpha} \quad (21)$$

Donde s_Y se reemplaza por s_r cuando no se conoce la desviación estándar de los errores experimentales.

La función estadística $F(m, n-m)$ se determina de tablas con un nivel de significancia $1-\alpha$ y con grados de libertad: $m, n-m$.

CONCLUSIONES

Las técnicas matemáticas que se usan para determinar parámetros por un análisis de regresión requieren el desarrollo de algoritmos para ejecución en computadora. Existe en la literatura especializada muchos algoritmos de técnicas que son modificaciones mejoradas de las que hemos presentado aquí.

En lo que sigue, presentaremos ejemplos de cálculo y los algoritmos correspondientes en un programa de computadora de fácil uso que permita al lector asimilar las técnicas matemáticas empleadas en el análisis de regresión no lineal.

BIBLIOGRAFÍA

- [1] Draper N., H. Smith; «Applied Regression Analysis», John Wiley & Sons Inc, 1981.
- [2] Gupta B.C., A.M. Mathai; «Regression and Analysis of Variance techniques», Instituto de Matemática de U. Federal de Río de Janeiro, 1980.
- [3] Himmelblau David M. «Process Analysis by Statistical Methods», John Wiley & Sons Inc, 1970.
- [4] Jacoby S.L., J.S. Kowalik, J.T. Pizzo; «Iterative methods for nonlinear optimization problems», Prentice Hall Inc, 1972.
- [5] Murray W. «Numerical methods for unconstrained optimization», Academic Press, 1972.