

ESTRATEGIA DE CÓMPUTO PARA LA RESOLUCIÓN DE MODELOS BIOLÓGICOS NO LINEALES

María V. Mederos*, Mayra Tejuca**, Emir Salas***³, María de los A. Chávez***

* Dpto. de Matemática Aplicada, Facultad de Matemática y Computación, UH.

** Dpto. de Bioquímica, Facultad de Biología, UH.

***Centro de Estudio de Proteínas, Facultad de Biología, UH.

ABSTRACT

A computational strategy is presented for the least squares curve fit of nonlinear models which allows obtaining a good initial approximation that guarantees convergence of the iterative process required by its solution. This strategy is illustrated with biological models.

KEY WORDS: nonlinear models, biological models

MSC: 65H05

RESUMEN

Se presenta una estrategia de cómputo para el ajuste de modelos no lineales, lo cual permite obtener una buena aproximación inicial que garantice la convergencia del proceso iterativo necesario para su resolución. Dicha estrategia se ilustra con modelos biológicos.

1. INTRODUCCIÓN

En este trabajo se describe la elaboración y validación de una herramienta de cómputo, empleando el asistente matemático Matlab, que permite el cálculo con precisión de modelos de regresión no lineales que aparecen en aplicaciones de la Biología. El procedimiento se realiza mediante una estrategia que consiste en una linealización previa de los modelos con vista a la determinación de una aproximación inicial que garantice la convergencia del proceso iterativo necesario para la aproximación mínimo cuadrática.

2. REGRESIÓN NO LINEAL

Dada una tabla de valores experimentales X / Y que representa una función $y = f(x; \alpha, \beta)$, donde α, β son parámetros a determinar, el conocido método de los mínimos cuadrados para la regresión estadística consiste en buscar \hat{f} dentro de una clase prefijada F de funciones, tal que se minimice el error cuadrático $E(\alpha, \beta)$:

$$E(\alpha, \beta) = \sum_{i=1}^n [y_i - f(x_i, \alpha, \beta)]^2$$

es decir,

$$\sum_{i=1}^n [y_i - \hat{f}(x_i, \alpha, \beta)]^2 = \min_{\hat{f} \in F} \sum_{i=1}^n [y_i - f(x_i, \alpha, \beta)]^2$$

con lo cual se obtenga la mejor función de aproximación que describe la tendencia del fenómeno representado por la tabla.

Matemáticamente, se trata de un problema de búsqueda de un extremo, que será mínimo absoluto de la función de dos variables $E(\alpha, \beta)$, lo que se logra mediante la resolución del sistema formado por la anulación de las derivadas parciales

$$\frac{\partial E}{\partial \alpha} = 0, \quad \frac{\partial E}{\partial \beta} = 0$$

denominado ecuaciones normales. Este se reduce al sistema de ecuaciones

$$\begin{aligned} \sum_{i=1}^n [y_i - f(x_i, \alpha, \beta)] \frac{\partial f}{\partial \alpha} &= 0 \\ \sum_{i=1}^n [y_i - f(x_i, \alpha, \beta)] \frac{\partial f}{\partial \beta} &= 0 \end{aligned} \quad (1)$$

Cuando la función f depende de α y β linealmente, la resolución de (1) no ofrece dificultades para la determinación del punto de mínimo representado por el vector $(\hat{\alpha}, \hat{\beta})$ que define a \hat{f} . Sin embargo, no sucede así cuando la dependencia es *no lineal*. En este caso la resolución de (1) debe hacerse por aproximaciones sucesivas convergentes a la solución $(\hat{\alpha}, \hat{\beta})$. Pero todo proceso de aproximaciones sucesivas requiere el conocimiento de una aproximación inicial (α_0, β_0) , y ésta debe ser suficientemente buena para garantizar la convergencia. El método más usado para llevar a cabo este proceso iterativo es el de Newton, resultante del truncamiento del desarrollo en serie de Taylor de la función f . Su aplicación no está exenta de dificultades de cálculo, y existen diversas variantes encaminadas a superarlas, pero la clave del éxito está en poder disponer de una buena aproximación inicial. Siendo f no lineal, si mediante alguna transformación matemática se logra su linealización, entonces el sistema (1) será también lineal y su resolución es inmediata. Esta solución del sistema (1) linealizado permite obtener la aproximación inicial requerida a partir de la transformación inversa de la realizada anteriormente, y luego el punto de mínimo por iteración de Newton. Por el contrario, si no es posible linealizar f , la aproximación inicial habrá que definirla teniendo en cuenta el conocimiento del fenómeno en cuestión, o investigando el comportamiento de la función error $E(\alpha, \beta)$ en una región adecuada del espacio paramétrico.

3 APLICACIÓN EN LOS MODELOS BIOLÓGICOS

Existe una diversidad de sistemas de cómputo profesionales que pueden usarse para la resolución de los modelos matemáticos no lineales provenientes del campo de la Biología. Pero todos ellos presuponen conocida esa aproximación inicial (α_0, β_0) adecuada para garantizar la aplicación exitosa de la iteración de Newton, lo cual no siempre es la realidad.

En este trabajo exponemos la estrategia que permite la obtención de una buena aproximación inicial mediante la resolución previa de modelos linealizados, y su aplicación posterior segura a la resolución del verdadero modelo no lineal, lo cual ilustramos con cuatro modelos biológicos. Estos son:

- 1) Hipérbola de la forma: $y = \frac{a x}{x + b}$ (modelo de Michaelis-Menten, que refleja la dependencia de la velocidad inicial de reacciones catalizadas por enzimas que no presentan cooperatividad, en función de la concentración inicial de sustrato presente en el medio biológico donde ocurre la misma).

$y = V_0$, velocidad inicial de reacción

$x = [S_0]$, concentración de sustrato

$a = V_{\max}$ y $b = K_M$, constantes que caracterizan cinética y termodinámicamente el sistema enzima-sustrato estudiado, y que dependen de las condiciones experimentales usadas.

2) Ecuación de Morrison, empleada en la Enzimología para el estudio de inhibidores de unión fuerte. La misma describe la dependencia de la fracción de inhibición con respecto a la potencia del inhibidor y diversas condiciones experimentales:

$$y = 1 + \frac{-([E_0] + [I_0] + Ki) + \sqrt{([E_0] + [I_0] + Ki)^2 - 4[E_0][I_0]}}{2[E_0]},$$

$$0 < y < 1, \quad (2)$$

para dos problemas relacionados, el Problema 1, donde los parámetros desconocidos son $[E_0]$ y Ki , y el Problema 2, donde estos son $[I_0]$ y Ki . La variable $y = V_i / V_o$, representa la razón de las velocidades de reacción en presencia de inhibidor (V_i) y en ausencia del mismo (V_o), y la variable x es $[I_0]$, concentración de inhibidor en el Problema 1, y $[E_0]$, concentración de enzima en el Problema 2. Ki es la constante de inhibición.

3) Función exponencial de la forma: $y = -a \exp(-x/b) + c_0$, que se usa para describir el curso temporal de procesos de primer orden, como por ejemplo, la dimerización de una molécula, donde
 x es la variable independiente, el tiempo
 y es la variable dependiente que en este ejemplo sería indicativa del fenómeno de dimerización
 a/b es la pendiente de la curva en el origen
 b es la constante de tiempo
 c_0 es la asíntota horizontal, es decir el valor máximo al que tiende la variable dependiente, en este ejemplo, la dimerización

4) Sigmoide $y = \left(\frac{x^a}{x^a + b^a} \right)$, que se aplica para describir cinéticamente fenómenos de unión que

impliquen cooperatividad, por ejemplo la saturación de la hemoglobina por el oxígeno o de enzimas por sus sustratos, donde

y representa la fracción de saturación de sitios de unión
 $x = [S]$, es la concentración de oxígeno o de sustrato
 a representa idealmente el número de sitios para la unión del oxígeno o del sustrato

En la práctica este tratamiento se aplica en estudios de unión de enzimas con sus sustratos. La obtención de un valor de $n > 1$ es un indicativo de la existencia de cooperatividad positiva en la unión del sustrato, $n=1$, de la ausencia de cooperatividad, y en el caso de enzimas que presentan cooperatividad negativa en la unión de su sustrato, se obtendrá un valor de $n < 1$.

b es la concentración de oxígeno o de sustrato que da lugar a una fracción de saturación de 0.5, es decir, que satura el 50 % de los sitios de unión

4 LINEALIZACIÓN

1) La linealización de la hipérbola se logra planteando el modelo recíproco, por tanto,

$$\frac{1}{y} = \frac{x+b}{ax} \rightarrow \frac{1}{y} = \frac{1}{a} + \frac{b}{a} \frac{1}{x} \rightarrow y' = c_1 + c_2 x'$$

La matriz de datos tiene la forma $M = \begin{bmatrix} 1 & 1/x_1 \\ \vdots & \vdots \\ 1 & 1/x_n \end{bmatrix}$ y el vector término independiente, $t = \begin{bmatrix} 1/y_1 \\ \vdots \\ 1/y_n \end{bmatrix}$.

Después de resolver el modelo lineal sobredeterminado anterior $Mc=t$, con $c = (c_1, c_2)$ se hace la conversión a los parámetros originales, obteniéndose $a = 1/c_1$ y $b = c_2/c_1$. La aproximación

inicial para la resolución del modelo no lineal estará dada entonces por el vector

$$d = (d_1, d_2) = (1/c_1, c_2/c_1).$$

2) Para la linealización de la ecuación de Morrison en el Problema 1, basta despejar $[I_0]$ en la expresión (2):

$$[I_0] = [E_0](1 - y) + Ki \left(\frac{1 - y}{y} \right),$$

y resolverlo usando regresión bilineal con la base dada por las funciones $g_1(y) = 1 - y$ y

$$h_1(y) = (1 - y)/y. \text{ La matriz de datos tiene la forma } M = \begin{bmatrix} 1 - y_1 & 1 - y_1 / y_1 \\ \vdots & \vdots \\ 1 - y_n & 1 - y_n / y_n \end{bmatrix}, \text{ y el vector}$$

$$\text{término independiente, } t = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}. \text{ Resolviendo el sistema lineal sobredeterminado } Mc = t \text{ se obtiene}$$

la solución

$c = (c_1, c_2)$, que se tomará directamente como aproximación inicial $d^{(0)} = (c_1, c_2)$, para la resolución del

modelo no lineal con la ecuación de Morrison.

Para la linealización en el Problema 2, se despeja $[E_0]$ en la expresión (2):

$$[E_0] = [I_0] \left(\frac{1}{1 - y} \right) + Ki \left(\frac{-1}{y} \right),$$

y se resuelve usando también regresión bilineal, pero con la base dada por las funciones

$g_2(y) = 1/(1 - y)$ y $h_2(y) = -1/y$. La solución $c = (c_1, c_2)$ del sistema $Mc = t$, con matriz de datos

$$M = \begin{bmatrix} 1/(1 - y_1) & -1/y_1 \\ \vdots & \vdots \\ 1/(1 - y_n) & -1/y_n \end{bmatrix} \text{ y término independiente } t = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \text{ se tomará también como}$$

aproximación inicial $d^{(0)} = (c_1, c_2)$ para la resolución de la ecuación de Morrison.

3) La función exponencial planteada es intrínsecamente no lineal, pero el conocimiento de una aproximación de la asíntota horizontal a partir de los datos, permite asignar un valor inicial al parámetro c_0 , y con ello transformar el modelo original en $c_0 - y = a \exp(-x/b)$, que ahora ya es lineizable mediante aplicación de logaritmos, siempre que se tome $c_0 > y_{\max}$. Se obtiene

$$\ln(c_0 - y) = \ln(a) - x/b \rightarrow y' = c_1 + c_2 x.$$

Después de resuelto el modelo linealizado anterior con matriz $M = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$ y término

independiente $t = \begin{bmatrix} \ln(c_0 - y_1) \\ \vdots \\ \ln(c_0 - y_n) \end{bmatrix}$, se obtiene el vector de parámetros $c = (c_1, c_2)$, y se hace la

conversión a los parámetros originales, calculándose $a = \exp(c_1)$ y $b = -1/c_2$. La aproximación inicial para la resolución del modelo no lineal estará dada entonces por el vector $d^{(0)} = (d_1, d_2, d_3) = (\exp(c_1), -1/c_2, c_0)$.

4) La linealización de la sigmoide sólo es posible si se consideran los valores de y en términos de fracciones de saturación \bar{y} , es decir, divididos por su valor máximo. Se obtiene entonces

$$\bar{y}(x^a + b^a) = x^a \rightarrow \bar{y}b^a = x^a(1 - \bar{y}) \rightarrow \frac{\bar{y}}{(1 - \bar{y})} = \frac{x^a}{b^a} \rightarrow \ln\left(\frac{\bar{y}}{1 - \bar{y}}\right) = -a \ln(b) + a \ln x$$

luego el modelo linealizado queda $y' = c_1 + c_2 x'$, y después de resuelto con matriz de datos

$$M = \begin{bmatrix} 1 & \ln(x_1) \\ \vdots & \vdots \\ 1 & \ln(x_n) \end{bmatrix} \text{ y término independiente } t = \begin{bmatrix} \ln\left(\frac{\bar{y}_1}{1 - \bar{y}_1}\right) \\ \vdots \\ \ln\left(\frac{\bar{y}_n}{1 - \bar{y}_n}\right) \end{bmatrix}, \text{ se hace la conversión a}$$

los parámetros originales, obteniéndose $a = c_2$ y $b = \exp\left(-c_1/c_2\right)$. La aproximación inicial para la resolución del modelo original con los valores originales de y será finalmente

$$d^{(0)} = \left(c_2, \exp\left(-c_1/c_2\right)\right).$$

5 RESOLUCIÓN DEL MODELO NO LINEAL

Para la aplicación del método de Newton contamos con el subprograma '*lsqcurvefit*' del asistente matemático Matlab, que es específico para el cálculo científico. Este facilita los cálculos mediante subprogramas como el mencionado, que actúan como caja negra en la resolución de los problemas matemáticos fundamentales, entre los cuales se encuentra el de ajuste de datos por el método de los mínimos cuadrados para modelos no lineales.

Habiendo garantizado en cada uno de los modelos ilustrativos expuestos la obtención de una aproximación inicial adecuada $d^{(0)} = (\alpha_0, \beta_0)$ o $d^{(0)} = (\alpha_0, \beta_0, \gamma_0)$, según el caso, mediante la resolución de modelos linealizados, basta ahora aplicar la función de Matlab '*lsqcurvefit*' para el logro del objetivo final de resolver con precisión el modelo no lineal. Sólo habrá que hacer el llamado de dicha función especificando los datos x, y , el vector $d^{(0)}$ y la expresión matemática del modelo correspondiente.

Por ejemplo, si para el modelo de la hipérbola se obtuvo como solución del modelo linealizado el vector de coeficientes $c = (0.0048, 0.0003)$, la conversión a los parámetros originales da la aproximación inicial

$$d^{(0)} = \left(\frac{1}{c_1}, \frac{c_2}{c_1}\right) = (208.33, 0.0625), \text{ y la aplicación del método de Newton usando la función}$$

de Matlab mencionada converge y proporciona para los coeficientes del modelo no lineal el vector final $d = (210.2, 0.0698)$. Por tanto la solución del modelo será $y = \frac{210.2 x}{x + 0.0698}$.

6 CONCLUSIONES

Los modelos mencionados como ilustración de la estrategia presentada han sido validados con datos reales,

poniéndose en evidencia con los resultados del estadígrafo r^2 cómo mejora la calidad del modelo no lineal, y garantizándose la convergencia del proceso iterativo al usar una buena aproximación inicial. Por último, es importante reiterar que esta estrategia despersonaliza el cómputo, y ahorra tiempo por eliminar la necesidad de barrer el espacio paramétrico para ubicar la aproximación inicial requerida de los parámetros, en ausencia de su conocimiento.

RECEIVED JULY 2009
REVISED NOVEMBER 2009

REFERENCIAS

- [1] CHÁVEZ, M.A., DÍAZ, J., PÉREZ, U., DELFÍN, J. (2006): **Temas de Enzimología**, Vol.1, Editorial Félix Varela, Ciudad de La Habana
- [2] DRAPER, N., SMITH, H. (1998): **Applied Regression Analysis**, 3rd edition, J. Wiley & Sons, New York
- [3] MATLAB (1994): **The Matrix Laboratory**. The Math Works Inc. Version 7.3, Illinois.
- [4] MONTGOMERY, D.C., PECK, E.A., VINING, G.G. (2002): **Introduction To Linear Regression Analysis**, Cía Editorial Continental S.A., México
- [5] MORRISON, J.F. (1969): Kinetics of the reversible inhibition of enzyme catalyzed reactions by tight-binding inhibitors. **Biochimica et Biophysica Acta** 185: 269-286