

La stima dei parametri nelle funzioni omografiche tramite una linearizzazione empirica

E. Paolinelli, Istituto di Matematica e Statistica Università di Ancona
L. Piantanelli, Centro di Biochimica INRCA, Ancona

Nella teoria dell'adattamento di funzioni a distribuzioni empiriche è di particolare rilievo la stima dei parametri.

La stima è agevole per i modelli lineari rispetto ai parametri. Lo è meno per i modelli non lineari.

Gli AA. si riferiscono alle funzioni omografiche. Realizzano opportune trasformazioni che consentono l'applicazione delle tecniche lineari e in due esempi fanno notare che le trasformazioni proposte convengono e danno risultati migliori di una procedura iterativa sul modello non lineare.

1. Introduzione

Nelle funzioni algebriche vi sono funzioni razionali intere e funzioni razionali fratte. Nelle funzioni razionali fratte la variabile indipendente è al denominatore del rapporto, i cui termini sono funzioni intere.

Si dicono funzioni omografiche le funzioni lineari fratte del tipo:

$$y = \frac{a_1x + b_1}{a_2x + b_2}$$

dove:

$$a_2 \neq 0 \quad |a_1 b_2 - b_1 a_2| \neq 0$$

Assumendo per asintoti gli assi cartesiani l'equazione dell'iperbole equilatera è la seguente:

$$xy = \frac{a}{2}$$

Più in generale, se gli asintoti sono paralleli agli assi cartesiani:

$$y = \frac{a_1x + b_1}{a_2x + b_2}$$

che è appunto l'espressione di una funzione omografica, funzione omografica bilineare.

2. Il problema

Le relazioni tra i fenomeni e quindi le relazioni tra le variabili sono frequentemente esprimibili solo in forma non lineare. Le funzioni omografiche ne sono un esempio particolarmente noto nella biochimica degli ormoni e dei farmaci (Rodbard *et al.*; 1975, pag. 3). Vogliamo stimare i parametri delle seguenti funzioni omografiche:

$$R = \frac{k_1 q_1}{1 + k_1 F} \quad (1)$$

$$R = \frac{k_1 q_1}{1 + k_1 F} + \frac{k_2 q_2}{1 + k_2 F} \quad (2)$$

$$R = \frac{k_1 q_1}{1 + k_1 F} + k_3 \quad (3)$$

dove F è la variabile indipendente e k_1 , k_2 , k_3 , q_1 e q_2 sono parametri positivi.

Per stimare i parametri effettuiamo le trasformazioni che consentono di linearizzare i modelli ed applichiamo i minimi quadrati ordinari (OLS) ai dati sperimentali trasformati (F.A. Graybill; 1961). Nei casi in cui per le trasformazioni introdotte risultassero rimosse le assunzioni proprie dei minimi quadrati ordinari, procederemo con i minimi quadrati generalizzati (GLS) (A.R. Gallant, 1975; J. Johnstone, 1978).

Gli esempi serviranno a chiarire come le trasformazioni introdotte diano buoni risultati e siano spesso più convenienti di altri metodi (M.A. Sparz, 1962; H.O. Hartley *et al.*, 1965; C.G. Broyden, 1969; Y. Bard, 1970; J. M. Ortega *et al.*, 1970).

3. Le trasformazioni

I modelli sono stati trasformati come segue:

1° modello

$$R = \frac{k_1 q_1}{1 + k_1 F}$$

$$\frac{1}{R} = \frac{1 + k_1 F}{k_1 q_1} = a + bF \quad (4)$$

dove: $a = \frac{1}{k_1 q_1}$ $b = \frac{1}{q_1}$

2° modello

$$R = \frac{k_1 q_1}{1 + k_1 F} + \frac{k_2 q_2}{1 + k_2 F}$$

Se fissiamo il punto di coordinate F_1, R_1 appartenente alla curva, avremo l'identità:

$$\frac{1}{R_1} = \frac{s + uF_1 + F_2}{t + vF_1}$$

dove:

$$\begin{aligned} t &= a_1 q_2 + a_2 q_1 & v &= q_1 + q_2 & s &= a_1 \cdot a_2 \\ u &= a_1 + a_2 & a_1 &= 1/k_1 & a_2 &= 1/k_2 \end{aligned}$$

Con un secondo punto: F_2, R_2 , ugualmente appartenente alla curva, dopo alcuni passaggi avremo la seguente linearizzazione:

$$\frac{F_2 - F}{\frac{1/R - 1/R_1}{F - F_1} \quad \frac{1/R_2 - 1/R_1}{F_2 - F_1}} = a + bF \quad (5)$$

dove:

$$a = \frac{t(t + vF_1)(t + vF_2)}{-t^2 + utv - sv^2}$$

$$b = \frac{v(t + vF_1)(t + vF_2)}{-t^2 + utv - sv^2}$$

Poniamo:

$$z = 1 + \frac{b}{a} F_1 \quad w = 1 + \frac{b}{a} F_2$$

e scriviamo il seguente sistema lineare di quattro equazioni, nelle incognite:

$$v, t, s, u.$$

$$v = \frac{b}{a} t$$

$$t = \frac{-a + ub - sb^2/a}{wz}$$

$$R_1 = \frac{-a + ub - sb^2/a}{(s + uF_1 + F_1^2)w}$$

$$R_2 = \frac{-a + ub - sb^2/a}{(s + uF_2 + F_2^2)z}$$

da cui:

$$s = \frac{NP + Q}{1 - MP} \quad u = \frac{NP + Q}{1 - MP} M - N$$

dove:

$$M = \frac{b^2/a + R_2 z}{R_2 F_2 z - b} \quad N = \frac{R_2 F_2 z - a}{R_2 F_2 z - b}$$

$$P = \frac{b - R_1 F_1 w}{R_1 w + b^2/a} \quad Q = \frac{R_1 F_1^2 w - a}{R_1 w + b^2/a}$$

sono tutti valori noti.

Per le relazioni esistenti tra le radici dell'equazione di 2° grado, sapendo che: $u = a_1 + a_2$ $s = a_1 \cdot a_2$

$$a^2 - ua + s = 0$$

Dalla soluzione dell'equazione:

$$a_1 = 1/k_1 \quad a_2 = 1/k_2$$

si ottengono poi i parametri: k_1 e k_2 .

3° modello

$$R = \frac{k_1 q_1}{1 + k_1 F} + k_3$$

Se fissiamo il punto di coordinate: F_1, R_1 appartenente alla curva, avremo l'identità:

$$R_1 = \frac{k_1 q_1}{1 + k_1 F_1} + k_3$$

quindi:

$$R - R_1 = \frac{k_1 q_1}{1 + k_1 F} - \frac{k_1 q_1}{1 + k_1 F_1}$$

ed anche:

$$\frac{F_1 - F}{R - R_1} = \frac{1 + k_1 F_1}{k_1^2 q_1} + \frac{1 + k_1 F}{k_1 q_1} \cdot F = a + bF$$

da cui:

$$k_1 = b/a \quad q_1 = \frac{1 + k_1 F_1}{k_1 b}$$

Si noti che per il 3° parametro, dalla trasformazione si ottiene:

$$R = R_1 + \frac{F_1 - F}{a + bF}$$

e dal modello:

$$\lim_{F \rightarrow \infty} R = k_3$$

quindi:

$$k_3 = R_1 - \frac{1}{b}$$

4. Applicazioni

Le trasformazioni realizzate muovono dal presupposto che si conoscano le coordinate di un punto per il terzo modello e di due punti per il secondo. Ovviamente non è dato a priori di conoscere tali punti. Perlustrando i dati sperimentali o l'intorno di questi, si assume che il punto o i due necessari appartengono alla curva se per l'ipotesi la distribuzione teorica risultante è quella che meglio si adatta alla distribuzione empirica. L'adattamento è misurato con il test χ^2 .

Esplorazione dei punti sperimentali - I modelli indicati sono tipici dell'interazione di un ormone o di un farmaco (detto genericamente ligando) con il proprio recettore.

In Biochimica il più usato nella identificazione e caratterizzazione dei recettori cellulari è il terzo modello.

L'analisi viene effettuata facendo interagire concentrazioni crescenti di un ligando marcato con i recettori presenti in un preparato biologico.

In breve tempo si stabiliscono condizioni di equilibrio dinamico tra ligando legato ai recettori e ligando rimasto libero. Per ogni concentrazione di ligando usato vengono determinate sperimentalmente le proporzioni di legato e di libero.

Il ligando libero è rappresentato nel modello dalla variabile F , mentre R rappresenta il rapporto tra la parte legata B e quella libera F , cioè $R = B/F$.

Per comodità, saranno riportati nelle tabelle soltanto i valori di F e di B e non quelli di R , che si ottengono dalla relazione suindicata.

I parametri q_1 e k_1 rappresentano rispettivamente il numero dei recettori presenti e la loro *affinità* per il ligando interagente e sono le due caratteristiche più importanti e più frequentemente richieste.

Il parametro k_3 , invece, tiene conto del fatto che l'ormone impiegato per l'analisi non si lega solo alle molecole dei recettori, ma anche, in maniera non specifica, ad altre molecole presenti nel materiale biologico in esame.

Tale legame non specifico può essere valutato sperimentalmente, ma in misura solo approssimata e con l'impiego di una serie supplementare di esperimenti. La valutazione può essere effettuata in maniera rapida, precisa e con risparmio di materiale biologico, con l'analisi matematico-statistica di una sola serie di dati sperimentali, utilizzando il terzo modello.

I dati della Tab. I si riferiscono ad un esperimento effettuato per caratterizzare i recettori beta-adrenergici dei linfociti umani. Come « ligando » radioattivo è stato usato il [3H]-Diidroalprenololo.

Adottando il metodo proposto, si verifica che per il terzo modello il

Tab. I - Distribuzioni relative alla caratterizzazione dei recettori beta-adrenergici dei linfociti umani.

| F (nM) | B speriment. (fmol mg ⁻¹) | B teorici (3° modello) |
|-----------|--|---------------------------|
| 0,402348 | 38,221153 | 39,019871 |
| 0,521520 | 42,357642 | 42,322726 |
| 0,815148 | 49,207666 | 49,207666 |
| 1,850176 | 69,046266 | 69,986463 |
| 3,977826 | 110,758616 | 110,208834 |

miglior punto di appoggio è il terzo punto sperimentale. La conformità della distribuzione teorica a quella empirica è ottima: $\chi^2 = 0,031175$ g.d.l. 5 $P > 0,999$.

Per i parametri le stime che risultano sono le seguenti:

$$k_1 = 14,53471 \quad q_1 = 36,93951 \quad k_3 = 18,57732.$$

Tra le condizioni che si richiedono per la corretta applicazione degli OLS vi è quella fondamentale che gli errori non siano correlati.

Nel nostro caso l'autocorrelazione dei residui è irrilevante. Comunque procedendo con i GLS (C. Cochran *et al.*; 1949) si ottengono le seguenti stime:

$$k_1 = 14,24024 \quad q_1 = 37,09415 \quad k_3 = 18,46982$$

e una distribuzione teorica pressoché uguale alle precedente. Nella Tab. II i dati sperimentali sono gli stessi analizzati da Munson (1979) nell'applicazione del suo programma di calcolo « Ligand » per la stima dei parametri del secondo modello e derivano, anche in questo caso, da esperimenti sulla caratterizzazione di recettori ormonali*. Munson nella parte centrale del suo programma inserisce essenzialmente l'algoritmo di Marquardt-Levenberg (R. Levenberg, 1944; D.W. Marquardt, 1963) Munson utilizza il suo metodo non-lineare con Rodbard nel 1980.

* Il primo modello considera l'interazione di un ligando con una popolazione di recettori specifici. Il secondo modello, per il ligando utilizzato, ipotizza due popolazioni recettoriali specifiche:

- la prima di affinità k_1 e di densità recettoriale q_1
- la seconda di affinità k_2 e di densità recettoriale q_2 .

In entrambi i modelli non figura il parametro k_3 . Ciò non significa che sperimentalmente non è presente alcun tipo di legame specifico, ma che di questo si è tenuto conto in precedenza con una valutazione a sè.

Tab. II - Confronto tra le distribuzioni teoriche del secondo modello, ottenute con il metodo P.P. e con il metodo « Ligand » di Munson (0.1 E-11).

| F | B speriment. | B teorici | |
|---------|--------------|-----------|-----------|
| | | P.P. | "Ligand" |
| 1,3524 | 0,0702 | 0,0528 | 0,0625 |
| 9,6857 | 0,5074 | 0,3781 | 0,4474 |
| 18,0191 | 0,8988 | 0,7033 | 0,8318 |
| 34,6858 | 1,4897 | 1,3533 | 1,5992 |
| 84,6858 | 3,5632 | 3,3001 | 3,8898 |
| 168,019 | 6,2184 | 6,5345 | 7,6695 |
| 501,352 | 17,2019 | 19,3438 | 22,3383 |
| 1668,02 | 57,9676 | 62,6314 | 68,8501 |
| 5001,35 | 169,018 | 174,647 | 174,6296 |
| 33334,7 | 762,406 | 762,406 | 698,1419 |
| 50001,4 | 980,695 | 975,986 | 937,6462 |
| 166668 | 1926,03 | 1926,03 | 2275,5015 |
| 833335 | 6061,45 | 6042,132 | 5853,1950 |

Nella Tab. II sono riportate le distribuzioni teoriche che si ottengono per il secondo modello con il metodo « Ligand » di Munson e con il metodo proposto nel presente lavoro, che per comodità chiameremo metodo P.P..

I due metodi conducono a distribuzioni teoriche che si accostano alla distribuzione empirica in modo molto diverso.

Con il metodo « Ligand » il valore di χ^2 (12 g.d.l.) è 72,37.

Con il metodo P.P. il valore di χ^2 (12 g.d.l.) è 1,005 con $P > 0,999$.

Munson aveva indicato che per i suoi dati sperimentali non era opportuno servirsi del secondo modello. Comunque dal lato statistico è inagibile il miglior adattamento ottenuto con il metodo P.P..

Esplorazione nell'intorno del punto sperimentale - Con il metodo P.P. può convenire talvolta rimuovere l'appoggio sui dati sperimentali, almeno per due ragioni:

- per evitare di consumare dati sperimentali;
- per migliorare l'adattamento della distribuzione teorica a quella empirica, esplorando nell'intorno del punto ottimo sperimentale.

Abbiamo applicato il terzo modello ai dati sperimentali di Munson. Con gli OLS il punto ottimo sperimentale si trova in $F_1 = 33334,7$ da cui deriva una distribuzione teorica che si accosta a quella empirica con un valore di $\chi^2 = 1,78$.

Se abbandoniamo il punto sperimentale ed esploriamo nel suo intorno, troviamo che in $F_1 = 33335,7$ possiamo migliorare ulteriormente la bontà dell'accostamento (vedi Tab. III).

Tab. III - Confronto fra le distribuzioni teoriche del terzo modello, ottenute con il metodo P.P. e con il metodo « Ligand » (0.1 E-11).

| F | B speriment. | B teorici | |
|---------|--------------|-----------|-----------|
| | | P.P. | "Ligand" |
| 1,3524 | 0,0702 | 0,0534 | 0,0612 |
| 9,6857 | 0,5074 | 0,3829 | 0,4387 |
| 18,0191 | 0,8988 | 0,7122 | 0,8159 |
| 34,6858 | 1,4897 | 1,3704 | 1,5696 |
| 84,6858 | 3,5632 | 3,3416 | 3,8256 |
| 168,019 | 6,2184 | 6,6161 | 7,5680 |
| 501,352 | 17,2019 | 19,5791 | 22,3217 |
| 1668,02 | 57,9676 | 63,3246 | 71,4065 |
| 5001,35 | 169,018 | 176,1199 | 193,3754 |
| 33334,7 | 762,406 | 762,4032 | 759,9493 |
| 50001,4 | 980,695 | 975,5512 | 962,8689 |
| 166668 | 1926,03 | 1938,9900 | 1890,9234 |
| 833335 | 6061,45 | 6070,9555 | 6248,7208 |

Ma in questo caso abbiamo anche trovato una correlazione positiva tra i residui risultanti dalla regressione con gli OLS. Infatti l'indice D.W. (Durbin-Watson's test) è uguale a 0,3061. Abbiamo pertanto ritenuto di abbandonare gli OLS e di agire con i GLS.

I risultati sono indicati nella Tab. IV.

Con il metodo « Ligand » il valore di χ^2 (12 g.d.l.) è 13,66.

Con il metodo P.P. il valore di χ^2 è 4,21 cui corrisponde per lo stesso numero di g.d.l. $P = 0,98$. Tale metodo si presenta anche in questo caso più vantaggioso dell'altro.

Tab. IV - Confronto fra le distribuzioni teoriche del terzo modello, ottenute con il metodo P.P. applicando i GLS e con il metodo « Ligand » di Munson (0.1 E-11).

| F | B speriment. | B teorici | |
|---------|--------------|-----------|-----------|
| | | P.P. | "Ligand" |
| 1,3524 | 0,0702 | 0,0569 | 0,0612 |
| 9,6857 | 0,5074 | 0,4075 | 0,4387 |
| 18,0191 | 0,8988 | 0,7579 | 0,8159 |
| 34,6858 | 1,4897 | 1,4582 | 1,5696 |
| 84,6858 | 3,5632 | 3,5551 | 3,8256 |
| 168,019 | 6,2184 | 7,0363 | 7,5680 |
| 501,352 | 17,2019 | 20,7944 | 22,3217 |
| 1668,02 | 57,9676 | 66,9530 | 71,4065 |
| 5001,35 | 169,018 | 184,1466 | 193,3754 |
| 33334,7 | 762,406 | 762,4059 | 759,9493 |
| 50001,4 | 980,695 | 966,1890 | 962,8689 |
| 166668 | 1926,03 | 1897,8137 | 1890,9234 |
| 833335 | 6061,45 | 6015,4241 | 6248,7208 |

Le stime dei parametri risultano come segue:

| metodo Munson | metodo P.P. |
|----------------------|----------------------|
| $k_1 = 4117860$ E-11 | $k_1 = 3414930$ E-11 |
| $q_1 = 944,996$ | $q_1 = 1056,856$ |
| $k_3 = 639655$ E-8 | $k_3 = 599332$ E-8 |

5. Conclusioni

Quasi sempre le trasformazioni riducono le relazioni non lineari nei dati originali a relazioni lineari nei dati trasformati. Le trasformazioni ragionevoli possono coprire un'ampia varietà di relazioni non lineari e rendere in pratica utili servizi al problema della stima dei parametri.

Le trasformazioni usate in questo studio per alcune funzioni omografiche, consentono più di un vantaggio:

- la stima dei parametri è semplice e si realizza con una modesta calcolatrice programmabile;

— la distribuzione teorica è migliore di quella che si ottiene con metodo non lineare.

Comunque il rischio di violare alcune delle ipotesi fondamentali, alla base delle tecniche lineari, non va sottovalutato. Che ciò si verifichi con le trasformazioni non è infrequente. Accertata la violazione almeno per le ipotesi della eteroscedasticità e della autocorrelazione dei residui, è opportuno sostituire gli OLS con i GLS. Forse in questo caso si perde un po' della speditezza di cui si è detto, ma si ottengono ancora distribuzioni teoriche migliori.

Riferimenti bibliografici

- Bard Y., Comparison of gradient methods for the solution of non-linear parameter estimation problems. *Numer. Anal.*, 7, 1970, 157-186.
- Broyden C.G., A new method for solving non-linear simultaneous equations. *Comput. J.*, 12, 1969, 95-100.
- Cochrane C., Orcutt G.H., Application of least squares regression to relationships containing autocorrelated error terms. *J. A. Statis. Ass.*, 44, 1949, 32-61.
- Gallant A.R., Seemingly unrelated non-linear regression. *Journal of Econometrics*, 3, 1975, 35-50.
- Graybill F.A., *An introduction to linear statistical models*, Vol. 1, New York, McGraw-Hill, 1961.
- Hartley H.O., Booker A., Non-linear least squares estimation. *The Annals of Mathematical Statistics*, 36, 1965, 638-643.
- Johnston J., *Econometrica*, F. Angeli Editore, Milano, 1978.
- Levenberg K., A method for the solution of certain non-linear problems in least squares. *Quart. Appl. Math.*, 2, 1944, 164-168.
- Marquardt D.W., An algorithm for least squares estimation of non-linear parameters. *J. Soc. Indust. Appl. Math.*, 11, 1963, 431-441.
- Munson P.J., *A program system for fitting multiple ligand, multiple binding site data*. Programmers guide to «Ligand», 1979. Programma ed esempio associati all'articolo:
- Munson P.J., Rodbard D., Ligand, A versatile computerized approach for characterization of ligand-binding system. *Anal Biochem.*, 107, 1980, 220-239.
- Rodbard D., Feldman H.A., *Theory of protein-ligand interaction*. In: Hormone action. Part A: Steroid hormones, pp. 3-16. Editors: B.W. O'Malley and J.G. Hardman. *Methods in Enzymology*, Vol. XXXVI, Editors: S.P. Colowick and Kaplan. Academic Press, New York, 1975.
- Ortega J.M., Rheinboldt, *Iterative solution of non-linear equations in several variables*, Academic Press, New York, 1970.
- Sparz M.A., A review of minimization technique for non-linear functions. *SIAM Review*, 4, 1962, 343-363.

Summary

The estimation of parameters represents a paramount aspect in the theory fitting mathematical functions to empirical distributions of data. Whilst the estimation is fairly easy for linear models, solutions may sometimes result rather wearisome for non-linear ones. In the present piece of work homographic functions are analysed by using linear methods following suitable transformations.

Two examples are given showing the convenience of the transformation used. Results were obtained better than those derived by an iterative procedure based on non-linear regression analysis. The presence of error correlation found in one of the given examples was taken into account by Durbin-Watson's test.