

비선형 회귀분석을 위한 소프트웨어 NLIN90의 소개¹⁾

강 근 석²⁾

요 약

기존의 몇몇 통계 패키지들(SAS, SPSS등)에서 제공하는 비선형 회귀분석을 위한 프로그램들은 그 사용방법이 불편하고, 최근에 개발된 중요한 통계치는 계산을 해주지 못하고 있다. 여기서 소개하고자 하는 비선형 회귀분석을 위한 소프트웨어 NLIN90은 비선형 회귀모형들의 데이터베이스화 사용이 편리하고, 기본 통계치외에 모형 전체와 모수 개개에 대한 곡률분석, 변환된 모수에 관한 곡률분석, 실험계획분석등 정확한 분석에 필요한 많은 통계치를 제공하는 소프트웨어로 개발되었다.

1. 서 론

비선형 회귀분석에 있어서는 그 특성상 정확한 통계이론에 의한 분석이 힘들고 계산상의 어려움으로 인하여 선형근사화(linear approximation)에 의존하는 분석이 주로 사용되며 그 계산을 위해서는 컴퓨터의 사용이 필수적이다. 선형근사화는 가정된 모형의 종류, 모수의 형태, 관측된 자료에 따라 정확도가 변화하기 때문에 (Bates와 Watts, 1980) 항상 분석하고자하는 개개의 문제에 대해 그 정확도를 조사하여야 한다. 이 선형근사화의 정확도를 나타내는 척도로서 모형의 비선형성에 대한 곡률측도(curvature measures of nonlinearity)를 사용할 수 있다. 모형의 비선형성에 대한 연구는 Bates와 Watts(1980) 이후부터 본격화되어 현재 많은 새로운 결과들이 나오고 있는 중이다. 이중 Bates와 Watts의 모형전체의 비선형성 척도(global curvature)와 Clarke (1987)의 각 모수에 대한 비선형성 척도(marginal curvature)는 비선형 회귀분석의 정확도 측정에 중요한 역할을 한다고 알려져 있다. 그러므로, 비선형 회귀문제의 정확한 분석을 위해서는 이들 척도의 계산이 필수불가결하다. 이를테면, 모수 θ 의 신뢰구간이 (a, b) 로 주어졌을 때 선형회귀에서는 이것이 정확한 것이지만, 비선형회귀

¹⁾ 이 논문은 1991년도 한국과학재단 연구비지원에 의한 결과임.
과제번호 : 913-0105-005-1

²⁾ (156-743) 서울시 동작구 상도 5동 1-1 숭실대학교 통계학과

에서는 비선형성 측도의 정도를 계산해보기 전에는 그 정확도에 대해 알 수가 없다. 또한 모수변환 (parametrization)에 따른 비선형성 측도값의 변화 측정도 중요하다고 할 수 있다 (Kang and Rawlings, 1989). 그리고, 각 모수의 비선형성을 그래프로서 나타내기 위한 방법으로는 Bates와 Watts (1988)의 profile analysis, Cook 과 Weisberg (1990)의 confidence curve등을 들 수 있다. 이와 같은 최근의 연구 결과들은 컴퓨터의 발달로 인해 가능해 졌다고 할 수 있다.

이와같은 분석을 위한 컴퓨터 소프트웨어의 국내 개발은 전무한 실정이고, 외국에서 개발된 기존의 몇몇 통계 패키지(SAS, BMDP)등에 비선형 회귀분석을 위한 프로그램들이 있지만 그 사용방법이 불편하고 (선형회귀에서와 달리 비선형회귀에서는 모형식과 모형식의 모수에 대한 일차 또는 이차 도함수를 사용자가 입력해 주어야 한다) 또한, 앞에서 언급된, 최근에 개발된 중요한 통계치는 계산을 해주지 못하기 때문에 사용이 더욱 편리하고 정확한 분석에 필요한 많은 통계치를 제공하는 컴퓨터 소프트웨어를 개발하고자 하였다.

2. 주요 개발내용

여기에서 소개되는 비선형 회귀분석을 위한 프로그램 NLIN90(a nonlinear regression program for 90's)에서는 다음의 내용들을 수행할 수 있다.

모수추정(parameter estimation)

독립변수와 반응변수의 측정값이 주어지면 Gauss-Newton 방법에 의한 모수의 최소제곱추정치들 구할 수 있다. 또한 이 추정치들에 대한 표준오차, 상관계수와 95% 신뢰구간등을 계산하여준다. 이 모수추정은 선택적이다. 곡률분석이나 실험계획분석에 있어서는 반응변수의 측정값이 없고 단지 실험점의 값만 가지고 분석할 필요가 있는데, 이 경우는 사용자가 입력하는 모수의 값을 가지고 분석할 수 있도록 하였다.

곡률분석(curvature analysis)

선형근사화에 바탕을 둔 분석의 정확도를 나타내는 두가지 비선형성 측도를 계산하여준다.

(1) 전체 모형의 비선형성 측도(global curvature)

여기서는 Bates와 Watts(1980)의 고유곡률(intrinsic curvature)과 모수효과곡률(parameter-effects curvature)을 계산하여준다.

(2) 각 모수의 비선형성측도(marginal curvature)

Cook 과 Witmer(1985)는 전체모형에 대한 비선형성측도의 값(특히 모수효과 곡률)이 크다하더라도 각 모수 개개에 대한 비선형성의 정도가 반드시 큰 것은 아님을 몇가지 예제의 경우에서 보였다. 그러므로, 선형근사화를 이용한 개별 모수에 대한 분석(예:t-검정 또는 신뢰구간)을 위해서는 각 모수 개개에 대한 비선형성의 정도를 조사할 필요가 있다. 이를 위해 NLIN90에서는 Clarke(1987)이 제안한 각 모수의 비선형성측도를 Kang과 Rawlings (1989)가 개발한 계산공식을 이용하여 계산하여 준다.

변환된 모수에 관한 곡률분석(curvature analysis for transformed parameters)

비선형 모형이 가지는 비선형성은 한 모형내에서도 그 모형에 포함된 모수의 형태에 따라 그 정도가 달라진다. 그러므로, 관측된 자료가 있는 경우 비선형성의 정도가 가장 작은 형태의 모수를 가지는 모형을 사용하는 것이 좋다. 이런 모형을 찾기 위해서는 여러가지 모수형태에 대한 곡률분석이 필요한 데, NLIN90에서는 두가지 방법으로 이 작업을 할 수가 있다. 첫번째로는 다른 모수형태를 가지는 비선형모형들을 직접 모형함수(model function)로 사용하여 곡률을 계산하는 것이고, 두번째로는 사용된 모형함수가 가지는 모수형태와의 관계를 이용하는 것이다. 즉, 분석하고자하는 모수형태가 모형함수의 것과 다를 때는 그 변환식과 그 식의 모수에 대한 도함수를 사용자가 입력하므로써 두가지 모수형태에 대한 곡률을 구할 수 있다(4절의 예제 참조).

실험계획분석(experimental design analysis)

회귀분석에서 만족할 만한 분석결과를 얻기 위해서는 그 분석목적에 맞는 데이터를 관찰하기 위한 실험계획의 문제를 생각하여야한다. 비선형회귀분석을 위한 실험계획의 목적은 여러가지가 있을 수 있지만 크게 나누어 모수의 정확한 추정을 위한 것(precise estimation of parameters)과 모형의 비선형성을 작게하기 위한 것(reduction of nonlinearity)으로 구분할 수 있다. 이와같은 실험계획의 분석을 위해 NLIN90에서는 Box와 Lucas(1959)의 D-optimal design criterion을 사용자가 지정한 모수의 조합에 대해 계산하여준다. 사용자는 각 모수의 조합을 0 과 1로 간단하게 표시하여 원하는 결과를 얻을 수 있다.

잔차분석(residual analysis)

회귀분석에서의 잔차분석은 사용되는 통계치의 다양해짐과 더불어 그 중요성을 더해가고 있다. NLIN90에서는 이 잔차분석에 필요한 통계치들을 계산하여 별도의 컴퓨터 화일에 저장하여준다. 이 화일은 모든 팩키지에 입력이 가능한 ASCII 코드로 저장되므로 사용자는 이를 가지고 좋은 그래픽을 제공하는 팩키지(예:S)에서 원하는 잔차분석을 할 수 있다. 현재로는 이 화일에 다음의 내용들이 저장된다: 독립변수와 반응변수의 측정값, 반응변수의 예측치, 잔차, 표준화 잔차, Cook의 영향측도(influence measure).

비선형모형의 데이터베이스화

사용자가 일반 통계팩키지에서 비선형회귀분석을 위한 프로그램을 사용할 때 제일 어려운 점은 분석하고자하는 모형의 각 모수에 대한 일차도함수 또는 이차도함수 식을 입력해야한다는 것이다. 이와같은 작업은 귀찮을 뿐만 아니라 입력과정에서도 많은 실수가 발생하여 사용자를 곤혹스럽게한다. 특히 사용자가 비통계전공자인 경우에는 이와같은 작업을 요구하기가 무리이다. 이를 위해 일부 팩키지들은 도함수를 쓰지않는 방법을 제공하고 있으나, 정확한 분석을 위해서는 도함수를 사용하는 것이 중요하고 또한 곡률분석의 경우에는 이차도함수의 식이 없이는 분석이 불가능하다.

비선형회귀분석에서 쓰이는 모형의 종류는 무한개 있을 수 있지만 실제로 각 학문분야에서 흔히 사용되는 비선형모형들의 갯수는 한정되어 있다. 예를 들어, 비선형회귀분석에 대한 이론과 응용을 기술한 Ratkowsky(1983)의 책에는 20여개의 모형이 소개되어 있다. 그러므로, 이들 대표적인 비선형모형들에 대해 모형함수식과 함수식의 각 모수에 대한 정확한 일차 및 이차도함수를 제공하는 부프로그램(subroutine)들을 작성하여 일종의 데이터베이스를 만듦으로써 사용자가 모형식이나 모형식의 모수에 대한 도함수를 구하거나 입력할 필요가 없이 단지 그가 원하는 모형식을 지정하여 쓸 수 있도록 한다. 현재는 20개 정도의 대표적인 비선형모형에 대한 부프로그램이 작성되어 있으며 앞으로 문헌조사를 통하여 계속 첨가할 예정이다.

3. 개발방법및 결과

비선형회귀분석을 위한 프로그램 NLIN90은 모두 FORTRAN언어를 사용하여 작성

되었으며, 여기에는 새로이 작성된 주 프로그램과 20여개의 부프로그램, 그리고, 계산의 정확도를 높이기 위해 개방 프로그램(public domain program)인 LINPACK (Dongarra 외, 1979)의 DTRDI, DTRSL, DQRDC, DQRSL과 BLAS(basic linear algebra subroutines)의 코드가 일부 이용되었다. 사용의 보편성을 위해 FORTRAN 77의 표준명령문만을 사용하였으므로 기존의 거의 모든 FORTRAN Compiler하에서 이용할 수 있다. 그리고, 모든 계산은 double precision으로 수행된다. 개발된 프로그램은 IBM PC 286급과 386급에서 MS-FORTRAN(Ver.4.0 과 5.0)을 이용하여 테스트되었으며, 대형 컴퓨터에서는 UNIX를 OS로 하는 CONVEX에서 테스트 되었다.

현재 본 프로그램내에는 작업배열의 크기가 약 32K 바이트로 설정되어 있어 분석 능력범위는 모수와 관측치의 갯수에 따라 달라진다. 예를 들면, 모수의 개수가 2개일 때는 관측치 300개 정도, 3개일 때는 120개 이내, 그리고, 5개일 때는 65개이내 정도에서 분석할 수 있다. 만약 문제가 분석능력범위를 넘을 때는 경고문과 함께 필요한 배열크기를 계산하여준다. 이 경우 사용자는 프로그램내의 작업배열크기를 설정하는 상수를 자신이 사용하고 있는 시스템의 범위 내에서 증가시켜 사용하면 된다.

4. 사용예제

본 프로그램의 실행 예제를 보여주기 위해 Clarke(1987)에 사용된 다음의 Michaelis-Menten 모형을 이용한다.

$$f = \theta_1 X / (\theta_2 + X) \quad (1)$$

이 모형에는 하나의 독립변수 X와 두 개의 모수 θ_1 , θ_2 가 있다. 이 예제에는 6개의 관측치가 있다. 위 모형의 또 다른 혼한 형태는

$$f = X / (\beta_1 + \beta_2 X) \quad (2)$$

이다. 즉, 모형식 (1) 과 (2)에서 모수들은 관계식

$$\beta_1 = \theta_2 / \theta_1, \quad \beta_2 = 1 / \theta_1 \quad (3)$$

을 가진다. 여기에서 우리는 분석에 (1)식의 모형을 사용하고, 모수 θ_1 과 θ_2 에 대한 곡률분석과 아울러 (2)식의 모수 β_1 과 β_2 에 대해서도 곡률분석을 원한다고 가정하자. 또한 모수 θ_1 과 θ_2 의 모든 조합에 대해 D-최적계획측도값을 원한다고 하자.

위의 분석들을 위한 입력 데이터파일은 그 내용이 모수의 개수와 모수추정의 선택을 나타내는 코드, 모수의 초기치, 독립변수의 개수, 관측치 수, 반응변수의 값의 제공

여부를 나타내는 코드, 그리고 관측자료등으로 입력되는 내용이 거의 모두 출력에 나타나므로 입력화일의 구조는 여기서 생략한다. 위의 예제에 대한 출력결과는 <표 1>에 주어져 있다.

출력의 첫부분은 입력자료, 모수의 초기치, 모수의 추정결과인 추정치, 표준오차, 상관계수의 값을 보여주고, 다음으로 전체모형에 대한 곡률분석의 결과를 보여준다. 첨부된 설명에 의해 이 모형과 자료의 조합에 대해서는 비선형성 정도가 그리 크지 않음을 알 수 있다. 그러나, 다음의 각 모수 개개에 대한 곡률분석에서는 선형근사화를 이용한 신뢰구간등이 정확하지 않음을 보여주고있다. 그리고, 계속되는 변환된 모수에 대한 곡률분석의 결과를 보면 이 데이터의 경우는 모형식 (1) 보다 모형식 (2)를 사용하는 것이 더욱 좋음을 알 수 있다. 즉, 모수 β_1 과 β_2 의 비선형성이 모수 θ_1 과 θ_2 의 비선형성 보다는 상대적으로 작다. 그리고, 마지막으로, D-최적계획측도의 값을 구하는 단계에서, 첫번째 값은 모수 (θ_1, θ_2)에 대한 D-최적계획측도값, 두번째는 모수 θ_2 에 대한 D-최적계획측도값, 세번째는 모수 θ_1 에 대한 D-최적계획값을 나타낸다.

<표 1> Michaelis-Menten 예제의 출력결과

TITLE: Michaelis-Menten Model, $f = p_1 * x / (p_2 + x)$

Input Data: X's and Y

.13800	.14800
.22000	.17100
.29100	.23400
.56000	.32400
.76600	.39000
1.4600	.49300

Parameter estimation requested. Initial parameter values are

1	2
.10000	.10000

Para.	Estimate	Std.Error	Correlation
-------	----------	-----------	-------------

1	.690398	3.682442E-02	1.00
2	.596545	6.825672E-02	.95 1.00

Variance estimate 1.8400E-04 based on 4 degrees of freedom

Global curvature (rms)

Intrinsic = .03070 Parameter = .12547

Compare these values with the upper 0.05 quantile of F dist. of (2, 4) d.f. Both values are considered small if $c * \sqrt{F} \ll 1$, where c is a curvature. Eg., if we accept a deviation of no more than 15%(10%), then we will declare this analysis as having unacceptable curvature if at least one value of $c * \sqrt{F}$ is greater than 0.3 (0.2) (Bates & Watts (1980), JRSSB, pp.1-25)

Marginal curvature for each parameter

Para.	Curvature	$ c * t $
1	.04826	.10280
2	.06321	.13463

where c is the marginal curvature and t is the upper 0.05 quantile of t distribution with 4 d.f.

If $|c * t| < 0.1$, curvature effects may be ignored and linear approximation will suffice.

If $0.1 < |c * t| < 1/3$, curvature effects should be compensated for as suggested by Clarke. (Eg., use adjusted conf. intervals)

If $|c * t| > 1/3$, curvature effects are serious enough to make any simple analysis risky. (Clarke (1987), JASA, pp.844-850)

95% Confidence Intervals

Para.	Est.	Linear Approx.		Adjusted by Curvature	
		Lower	Upper	Lower	Upper
1	.69040	.61196	.76883	.62003	.77689
2	.59654	.45116	.74193	.47074	.76150

Marginal curvature of the transformed parameters

No. of transformation = 2

Para.	Value	Std. Error	Curvature	$ c * t $
1	.86406	5.73255E-02	.02461	.05242
2	1.4484	7.72569E-02	.00507	.01081

95% Confidence Intervals

	Linear Approx.		Adjusted by Curvature	
	Lower	Upper	Lower	Upper
	.74196	.98616	.74836	.99256
	1.2839	1.6130	1.2857	1.6148

D-optimal design criterion : evaluated for the subsets of parameters whose index is 1

Value of criterion =	.22395	INDEX =	1	1
Value of criterion =	.19873	INDEX =	0	1
Value of criterion =	.36836	INDEX =	1	0

위의 출력의에 NLIN90은 잔차분석에 필요한 통계치를 수록하는 파일 fort5.lis 와 Gauss-Newton 방법의 중간과정 결과들을 수록한 파일 fort6.lis를 생성한다. 추정에 문제가 발생하는 경우 이 파일의 내용을 참조하여 해결할 수 있다.

5. 결 론

비선형 회귀분석의 사용범위는 매우 광범위하다고 할 수 있다. 대표적인 예로는 화학이나 화학공학에서의 화학 반응속도 추정, 동식물학에서의 성장 곡선 추정, 의학이나 약학에서의 투약 반응 속도 추정, 농학에서의 공해의 농수산물 수확량에 미치는 영향분석, 계량 경제학에서의 각종 비선형 모형의 모수 추정등을 들수 있다. 이들의 각 경우가 비선형회귀분석을 이용하여 분석해야 함에도 불구하고, 좋은 소프트웨어의 미비로 많은 사람들이 정확하지 않은 근사방법을 쓰고 있는 것이 현재의 상황이다. 여기에 소개된 비선형 회귀분석을 위한 컴퓨터 소프트웨어인 NLIN90은 외국에서 개발된 기존의 몇몇 통계 패키지(SAS, BMDP)등에 있는 비선형 회귀분석을 위한 프로그램들 보다 사용이 편리하고, 정확한 분석에 필요한 많은 통계치를 제공하는 소프트웨어로서 이들 분야에서의 연구에 필요한 공헌을 할 것이라 생각한다.

또한, 한글 통계패키지의 개발이 시급한 실정에서 본 연구의 결과는 이를 위한 기술축적의 단계로서 평가할 수 있다(이 정진외, 1992). 한글 통계패키지가 개발되면 본 연구에서 개발된 컴퓨터 루틴들을 그대로 첨가할 수 있을 것이다. 그리고, 이 프로그램은 앞으로 개발할 예정인 다변량 비선형 회귀분석(multiresponse regression analysis)을 위한 프로그램의 기초가 될 것이다.

참 고 문 헌

- [1] Bates, D. M. and Watts, D. G. (1980), "Relative Curvature Measures of Nonlinearity (with discussion)", *Journal of the Royal Statistical Society*, B. 42, 1-25.
- [2] Bates, D. M. and Watts, D. G. (1988), *Nonlinear Regression Analysis and its Applications*, Wiley, New York.
- [3] Box, G. E. P. and Lucas, H. L. (1959), "Design of experiments in non-linear

- situation", *Biometirka*, 46, 77-90.
- [4] Clarke, G. P. Y. (1987), "Marginal Curvatures and Their Usefulness in the Analysis of Nonlinear Regression Models", *Journal of the American Statistical Association*, 82, 844-850.
- [5] Cook, R. D. and Weisberg, S. (1990), "Confidence Curves in Nonlinear Regression", *Journal of the American Statistical Association*, 85, 544-551.
- [6] Cook, R. D. and Witmer, J. A. (1985), "A Note on Parameter-effects Curvature", *Journal of the American Statistical Association*, 80, 872-878.
- [7] Dongarra, J. J., Bunch, J. R., Moler, C. B., and Stewart, G. W. (1979), *Linpack Users' Guide*. Philadelphia: SIAM
- [8] Kang, G. and Rawlings, J. (1989), "Marginal Curvature for Functions of Parameters in Nonlinear Regression", *Journal Series of the North Carolina Agricultural Research*.
- [9] Ratkowsky, D. A. (1983), *Nonlinear Regression Modelling: A Unified Practical Approach*. New York: Marcel Dekker.
- [10] 이 정진, 강근석, 이 윤오 (1992), "통계학 교육용 한글 소프트웨어 개발 연구", *응용통계연구*, 5권 1호, 81-90.

Introduction of NLIN90, a Software for Nonlinear Regression Analysis¹⁾

Gunseog Kang²⁾

Abstract

A computer software for nonlinear regression analysis, NLIN90, was developed to provide easy access and useful information for more precise analysis which can be obtained from the newly developed theory. Together with the elementary statistics, it provides statistics for curvature analysis of model function and of each parameter, for curvare analysis of transformed parameters, for experimental design analysis, and for residual analysis. Easy access is obtained by utilizing a database of nonlinear models.

¹⁾ This research was supported by Korea Science & Engineering Foundation, Grant No. 913-0105-005-1

²⁾ Department of Statistics, Soongsil University, Sangdo 5 Dong 1-1, Dongjakku, Seoul, 156-743, Korea