Confidence intervals for nonlinear regression: A BASIC program

STEPHEN JOSÉ HANSON
Arizona State University, Tempe, Arizona 85281

In psychological experiments linearity is desirable and usually realized. Some models, however, cannot avoid the nonlinearity of the underlying processes, or, for descriptive models, the obvious coordinate systems (log-log, semilog, probit, etc.) which leave the data nonlinear have been tried and failed. In any case, nonlinear regression becomes appropriate.

Recent examples of parameters that appear in intrinsically nonlinear equations are arousal constants (Killeen, 1975), expectancy thresholds (Gibbon, 1977), transfer parameters (Davis & Levine, 1977), intellectual growth parameters (Zajonc & Markos, 1975), generalization factors (Blough, 1975), coupling constants (Staddon, 1977), cost parameters (Staddon, in press), consolidation parameters (Wickelgren, 1976), salience parameters (Rescorla & Wagner, 1972), and search constants (Norman & Rumelhart, 1970). Unless a large amount of data variance is accounted for or confidence intervals are established, the relationships or critical differences between parameters cannot be assessed.

This paper briefly outlines a method to obtain confidence intervals from nonlinear regression and demonstrates the use of a BASIC program employing this algorithm for computing confidence intervals for any nonlinear function regardless of the regression method.

Confidence Intervals for Nonlinear Regression. Suppose observations from an experiment arise as (y_1, t_1), $(y_2, t_2) \ldots (y_n, t_n)$. The standard regression model is of the form

$$y_i = F(t_i, \theta) + \epsilon_i,$$

where, as usual, the ϵ_i are assumed to be normally distributed with mean zero and common variance.

If $F(\theta)$ is an n by p matrix with elements

$$= [\delta F(t_i, \theta_i)] / \delta \theta_i,$$

then the matrix $(F'F)^{-1}_{pxp}$ contains in the diagonals the variances of the p parameters. Denote this inverse matrix as $\hat{C}$; then the 95% confidence interval may be found for the ith parameter given the .025 critical value of a t variate with $n - p$ degrees of freedom as

$$\theta_i \pm t_{.025} \sqrt{\hat{c}_{ii}} \; s.$$

Similarly, a t value for each parameter may be constructed by computing the ratio

$$\theta_i / \sqrt{\hat{c}_{ii}} \; s,$$

where s is the $\sqrt{1/(n-p) \Sigma(y_j - F(t_i, \theta_i))^2}$ and $\hat{C}_{ii}$ is the ith diagonal element from $\hat{c}$ (cf. Gallant, 1975; also see for convergence properties of these statistics).

Program Parameters. The following input information is either specified in the initial queries or documented in various points in the program CONF. Initially, the user-defined function FNA should be input at line 1 of the program (see Figure 1). All computations later will be made on FNA. Any nonlinear function may be input into the program this way as long as the variable is T and the parameters are written as P(1), P(2), ... P(i). These parameter positions in the function should be noted as they must be input into the same positions as they appear in this function.

The program asks first for the number of parameters and the number of data points. Next, it needs the sum of the squares residual (s); if only error variance is output, s may be computed by

$$s = \sqrt{\text{error variance}} \quad n/(n-p).$$

Finally, input each estimated value of the parameter from the regression into the P vector. The program asks for each parameter in the equation sequentially to be input into P(i), $i = 1, 2, \ldots p$ (see Figure 2).

The program begins by numerically computing the partials of the user-defined function and filling the F matrix, with these partials evaluated at the estimated values of the parameters and observation points t_i.[1] Next, an inversion is performed on the $F'F$ matrix through a Gauss-Jordan reduction. The diagonal elements of this matrix are the variances of the parameters. Thus, the standard error (SE), a t value, and confidence intervals for the parameters may all be defined.

The statistical properties of the parameters are useful in that they can be given some interpretation. For example, the t value gives some indication of the "significance" of that particular parameter. Although the model may be a good description of the data, the amount of "work" done by any particular parameter may be insignificant. Specifically, a parameter whose t value is less than $t_{.025}$ (with $n - p$ df) is considered not different[2] from zero. Standard errors of the parameters give some indication of how well the parameters are determined. For example, if $SE \, \theta_j > SE \, \theta_i$, a parameter-parameter confidence region maps out an elliptical area where the major axis corresponding to θ_j is larger than that corresponding to θ_i (Draper & Smith, 1966). That is, for any particular value of θ_i there is a larger range of values of θ_j which holds the regression constant. This statistic indicates from the data peculiarities of the function and the importance of the kind or "quality" of work done by that parameter.

Figure 1

```
0      REM INPUT MODEL HERE (LINE 1)
1      DEF FNA(T)=P(1)*(EXP(-T/P(2))-EXP(-T/P(3)))
2      DIM Z(50,50),Z1(50,50)
4      PRINT "NO. OF PARAMETERS, NO. OF DATA PTS.?"
5      INPUT N,M\N1=M
6      PRINT"SS RES.";\INPUT W
7      PRINT "INPUT "N" PARAMETERS"
8      FOR J=1 TO N
9      PRINT "P("J")=";\INPUT P1(J)
10     P(J)=P1(J)\NEXT J
11     REM CONSTANTS LINES 15-17
12     REM LINE 15 =CRITICAL T VALUE
13     REM LINE 16 =LABLE FOR LEVEL
14     REM LINE 17 =SCALE VALUE FOR Y
15     Q=2.306
16     A=95
17     V=1
20     FOR J=1 TO N
25     T=-1/(2*M)*V
30     K=1
31     REM DO PARTIALS
35     FOR I=1 TO M
40     T=T+1/M*V
45     P=P(J)
50     P(J)=.999*P(J)
55     GOSUB 1000
60     P(J)=1.002*P(J)
65     GOSUB 1000
70     Z(I,J)=(D(2)-D(1))/2.00000E-03/P
75     P(J)=.999*P(J)
80     NEXT I
85     NEXT J
100    REM TRANSPOSE Z
105    FOR I=1 TO N
110    FOR J=1 TO M
115    Z1(I,J)=Z(J,I)
120    NEXT J
125    NEXT I
130    REM ZT*Z
140    FOR I=1 TO N
150    FOR J=1 TO N
155    FOR K=1 TO M
160    A(I,J)=A(I,J)+(Z1(I,K)*Z(K,J)
165    NEXT K
166    NEXT J
168    NEXT I
169    REM COMPUTE INVERSE A
170    I=1
171    M=N
175    J=M+1
180    B=I+M
185    IF B=J THEN 198
190    A(I,J)=0
195    GO TO 200
198    A(I,J)=1
```

Figure 1 Continued

```
200   J=J+1
215   IF J<=2*M THEN 185
216   I=I+1
217   IF I<=M THEN 175
230   REM BEGIN GAUSS–JORDAN ELIMINATION
235   K=1
240   J=K
245   S=A(K,K)
250   A(K,J)=A(K,J)/S
255   J=J+1
260   IF J<=2*M THEN 250
265   I=1
270   IF I=K THEN 400
275   J=K
280   S=A(I,K)
350   A(I,J)=A(I,J)-S*(A(K,J))
360   J=J+1
380   IF J<=2*M THEN 350
400   I=I+1
405   IF I<=M THEN 270
410   K=K+1
415   IF K<=M THEN 240
450   PRINT "CONFIDENCE INTERVALS FOR "N1 – N" DF AT "A"% LEVEL"
451   PRINT "AND T FOR HO: P(I)=0"
460   S=SQR(W/(N1-N))
500   FOR I=1 TO M
501   K=M+I
502   P(I)=P1(I)
504   PRINT "S.E.(P("I"))="SQR(A(I,K))
505   PRINT P(I)"+OR-"Q*SQR(A(I,K))*S
507   T=P(I)/SQR(A(I,K))*S)
508   PRINT "T="T
509   IF T>Q THEN 511
510   PRINT "NOT SIGNIFICANT"
511   NEXT I
550   END
1000  IF K=2GO TO 1015
1014  K=1
1015  D(K)=FNA(T)
1020  K=2
1025  RETURN
```

These statistical properties are output by CONF in a convenient format in the order in which the parameters were input (see Figure 2).

An Example. Killeen (1975) suggested a nonlinearizable function for the time course of general activity of pigeons between periodic brief feedings:

$$R(t) = A(e^{-t/C} - e^{-t/I}),\ C > I,\ t > 0.$$

This function is bitonic for C and I > 0, and rate of growth for decay from the maxima is governed by the two time constants I and C. The time constants are indicants of two interacting exponential processes which arise between feedings. "A" was taken as a measure of the motivation or arousal of the organism. These parameters were estimated on a PDP-11 minicomputer using a random leap-optimization algorithm (cf. Curry, 1975).

This model has been input as the user-defined function FNA at line 1 (see Figure 1), and two cases of the activity data are considered. The first case arises under one of the shortest interfood intervals and is quite peaked; that is, the data rise to and fall from its maxima quite rapidly. Nonetheless, this particular model de-

```
RUNNH
NO. OF PARAMETERS, NO. OF DATA PTS.?
?3
?10
SS RES.?9.7
INPUT 3 PARAMETERS
P( 1 )=?134
P( 2 )=?.30
P( 3 )=?.13
CONFIDENCE INTERVAL FOR 7 DF & 95 % LEVEL
AND T FOR H0: P(I)=0
S.E.(P( 1 ))= 11.8454
 134 +OR- 32.1549
T= 9.60985
S.E.(P( 2 ))= .024406
 .3 +OR- .066251
T= 10.4421
S.E.(P( 3 ))= 1.22117E-03
 .13 +OR- 3.31492E-03
T= 90.4336

RUNNH
NO. OF PARAMETERS, NO. OF DATA PTS.?
?3
?10
SS RES.?2.17
INPUT 3 PARAMETERS
P( 1 )=?10
P( 2 )=?1.63
P( 3 )=?.02
CONFIDENCE INTERVAL FOR 7 DF & 95 % LEVEL
AND T FOR H0: P(I)=0
S.E.(P( 1 ))= 2.35737
 10 +OR- 3.02668
T= 7.6189
S.E.(P( 2 ))= 1.9574
 1.63 +OR- 2.51316
T= 1.49564
NOT SIGNIFICANT
S.E.(P( 3 ))= .0194156
 .02 +OR- .0249283
T= 1.85011
NOT SIGNIFICANT
READY
```

Figure 2

scribes the data quite well (ω^2 = .99). The least squares estimates of Parameters A, C, and I were, respectively, 134, .30, and .13. This information was input into CONF and standard errors, T values, and confidence intervals were computed for all three parameters. The sample run appears in the top of Figure 2.

Two aspects of the parameters are notable. First, as indicated by the t values, all parameters are doing a "significant" amount of work. Second, due to the nature of the function, some parameters seem more constrained than others. For example, comparing the standard error of I to either C or A indicates that the allowable range of the latter two parameters is much greater than that of I. Thus, changes in I have a much more significant effect on the fit than either of the other two parameters.

The second case is also interesting. The data were collected at one of the longest intervals, 200 sec. The

data were nearly flat (see Killeen, 1975, Figure 1), although the best fit was that with a nonzero I value (increasing limb). The fit here too is quite good ($\omega^2 = .97$). The input parameters into CONF were 10 (A), 1.63 (C), and .02 (I). The sample run is shown at the bottom of Figure 2.

The results of the run indicate that neither I nor C is necessary for the goodness of fit. Both t values are less than that for a critical .025 t value[3] with 7 df. Also, the confidence intervals for these parameters are larger than the parameters themselves. "A" on the other hand is quite constrained and appears to be the only hardworking parameter in this fit.

Across runs the parameter changes over some independent variable, and, in this case, interfood interval can be accompanied by corresponding confidence limit changes. Changes in confidence limits might be an important consideration when attempting to determine the kind of functional form the parameters take.

Computer. CONF was written and developed on a PDP-11/05 under the RT-11 operating system.

Availability. A listing of CONF with sample output is available at no cost upon request from the author at Arizona State University, Department of Psychology, Tempe, Arizona 85281.

REFERENCES

BLOUGH, D. S. Steady state data and a quantitative model of operant generalization and discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, 1975, **104**, 3-22.

CURRY, R. E. A random search algorithm for laboratory computers. *Behavior Research Methods & Instrumentation*, 1975, 7, 369-376.

DAVIS, J. D., & LEVINE, M. W. A model for the control of ingestion. *Psychological Review*, 1977, **84**, 379-412.

DRAPER, N. R., & SMITH, H. *Applied regression analysis.* New York: Wiley, 1966.

GALLANT, A. R. Nonlinear regression. *American Statistician*, 1975, **29**, 73-81.

GIBBON, J. Scalar expectancy theory and Weber's Law in animal timing. *Psychological Review*, 1977, **84**, 279-325.

KILLEEN, P. R. On the temporal control of behavior. *Psychological Review*, 1975, **82**, 89-115.

NORMAN, D., & RUMELHART, D. A system for perception and memory. In D. Norman (Ed.), *Models of human memory.* New York: Academic Press, 1970.

RESCORLA, R., & WAGNER, A. R. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory.* New York: Appleton-Century-Crofts, 1972.

STADDON, J. E. R. Behavioral competition in conditioning situations: Notes toward a theory of generalization and inhibition. In H. Davis & H. M. B. Hurwitz (Eds.), *Operant-Pavlovian interaction.* New York: Wiley, 1977.

STADDON, J. E. R. Operant behavior as adaptation to constraint. *Journal of Experimental Psychology: General*, in press.

WICKELGREN, W. Network strength theory of storage and retrieval dynamics. *Psychological Review*, 1976, **86**, 466-478.

ZAJONC, R. B., & MARKOS, G. B. Birth order and intellectual development. *Psychological Review*, 1975, **82**, 74-88.

NOTES

1. The variable T is normalized over (0,1) for ready comparison of different data spans. However, note that if data are fit over a nonnormalized range, this variable must be altered in the program at line 17: Simply set V equal to the upper limit of the range of the independent variable.

2. For time constants as in the present case (or any parameter which appears in a denominator), a test against either zero or infinity would be equivalent, since both values reduce the exponential terms to constants independent of T. However, one must be careful of the interpretation of the kind of "nonwork" the parameter is doing. Time constants close to zero indicate the process is almost over once it starts. On the other hand, time constants close to infinity indicate that the process is quite persistent.

3. The probability level may be altered at line 15 by setting Q equal to the appropriate critical t value (adjust A at line 16 also).