

MPRA

Munich Personal RePEc Archive

A note on least squares fitting of signal waveforms

Mishra, SK

North-Eastern Hill University, Shillong (India)

04. September 2007

Online at <http://mpra.ub.uni-muenchen.de/4705/>

MPRA Paper No. 4705, posted 07. November 2007 / 04:09

A Note on Least-Squares Fitting of Signal Waveforms

SK Mishra
 Dept. of Economics
 North-Eastern Hill University
 Shillong, Meghalaya (India)
 E-mail: mishrasknehu@yahoo.com

Introduction: In their recently published paper Han et al. (2006) fit the function

$$s(t) = \sum_{j=1}^{j_{\max}} \frac{a_j}{1-b_j\tau} \left[\exp(-b_j t) - \exp(-t/\tau) \right] \quad \dots \quad (1)$$

to the data of signal waveforms associated with optogalvanic (OG) transitions recorded with a hollow cathode discharge tube containing a mixture of neon (Ne) and carbon monoxide (CO) gases. Before fitting the function to the actual experimental data, they validate the method of curve fitting with the simulated data consisting of 601 points of $[s(t), t]$, where $t=0, 0.25, 0.50, \dots, 150$ at equal intervals. They observe that the regular least squares fitting technique is unstable when used to fit exponential functions to signal waveforms, since such functions are highly correlated. Therefore, they devise a procedure based on Monte Carlo method, utilizing both search and random walk. They report that the proposed procedure gives a stable least squares fitting algorithm that converges very rapidly. For their simulated data the details are given as follows.

Table-1: Details of the Curve Fitting Exercise on the Simulated Data (Han et al.)										
Parameters	τ	a_1	b_1	a_2	b_2	a_3	b_3	χ^2	Computation time (min)	CPU GHz
True	0.50	2.00	0.20	-2.00	0.10	0.50	0.05	-	-	-
Guessed	0.10	0.50	0.30	-0.40	0.05	0.10	0.02	-	-	-
Estimated	0.511	2.07	0.202	-1.90	0.107	0.334	0.0443	3.2E-07	≈ 20 minutes	1.73

A perusal of the results shows that Han et al. have been able to estimate the parameters that are quite close to the true ones (with which the simulated data were generated). For whatever little discrepancies remain, they hold that one reason is that some parameters (a_3 and b_3) are small, so that the small differences between their true values and the fitted values amount to a larger percentage difference. Further, “the correlations between the parameters also contribute to the fact that the fitted values do not equal the true values.” They do not attribute the discrepancies arising out of the power of their algorithm. This is exactly what has attracted us to this work.

Estimation of Parameters by the Differential Evolution based Algorithm: We generate the data (601 points) and fit the equation (1) to it by an algorithm that obtains least squares by the Differential Evolution method of global optimization (Mishra, 2007). This method has shown a great power in fitting nonlinear curves to datasets given by NIST (National Institute of Standards and Technology, USA), CPC-X and others.

We rename $s(t)$ as $y = [y_i; i=1, 2, \dots, 601]$ and t as $x = [x_i; i=1, 2, \dots, 601]$. The (FORTRAN) program for generating the data is given below.

```

parameter (n=601, m=3)
implicit double precision (a-h, o-z)
character *20 fo
dimension a(m),b(m), x(n),y(n)
COMMON /RNDM/IU,IV
fo='dat40'
jmax = 3
tau = 0.5d0
a(1) = 2.d0
b(1) = 0.2d0
a(2) = -2.d0
b(2) = 0.1d0
a(3) = 0.5d0
b(3) = 0.05d0
del=150.d0/600.d0
C generate x(i), y(i), i=1,n
open(7, file=fo)
do i=1,n
x(i)=(i-1)*del
y(i)=0.d0
do j=1,jmax
v=(a(j)/(1.d0-b(j)*tau))*(dexp(-b(j)*x(i))-dexp(-x(i)/tau))
y(i)=y(i)+v
enddo
write(7,1)i, x(i),y(i)
enddo
close(7)
1 format(i5,f10.5,f16.12)
write(*,*) 'over'
end

```

When we estimate the parameters of the model (equation –1) we obtain the results that are much more accurate than those obtained by Han et al. One of the sample results is given in table-3. It may be noted that we have run the program on a very slow computer (7.5 MHz). Even then, the solution has taken about 25 minutes only. Hahn et al. had obtained their results on a 1.7 GHz machine in 20 minutes. Further, our starting points are not as close to the true parameters as set by Han et al. We have given wider range as the domain of parameters. One set of our results is given in Table-2 below (time taken 25 min approx).

Parameters	τ	a_1	b_1	a_2	b_2	a_3	b_3
True	0.50	2.00	0.20	-2.00	0.10	0.50	0.05
Range	0 - 1	0 - 3	0 - 1	0 - 4	0 - 1	0 - 1	0 - 1
Estimated	0.5000374	1.9999816	0.2000151	-1.9993511	0.1000169	0.4993898	0.0499842

For the estimates reported in Table-2 we have obtained $s^2 = 0.2961362927800704E-08$ where $\sum_{i=1}^n e_i^2 = \sum_{i=1}^{601} (y_i - \hat{y}_i)^2 = s^2$ which is a measure of the goodness of fit.

```

SUBROUTINE REGMODEL_40(M,X,F,NQ) /optogalvanic (OG) effct FUNCTION
IMPLICIT DOUBLE PRECISION (A-H,O-Z)
COMMON /RNDM/IU,IV
COMMON /REGDAT40/DMODEL40(601,2),NDM40
INTEGER IU,IV
DIMENSION X(*),ALIM(7), a(3), b(3)
DATA (ALIM(I),I=1,7)/1, 3, 1, 4, 1, 1, 1/
jmax = 3
IF(NQ.NE.0) OPEN(11,FILE='REGRES')
DO J=1,M
IF(X(J).GT.ALIM(J).OR. X(J).LT.0.DO) THEN
CALL RANDOM(RAND)
X(J)=ALIM(J)*RAND
ENDIF
ENDDO
SY=0.DO
SSY=0.DO
S=0.DO
SER=0.DO
N=NDM40 / N IS THE NUMBER OF OBSERVATIONS IN DATA SET 1
C REDEFINE PARAMETERS TO BE ESTIMATED
tau=X(1)
a(1)=X(2)
b(1)=X(3)
a(2)=-x(4) !(negatively signed) Note this
b(2)=x(5)
a(3)=x(6)
b(3)=x(7)
DO I=1,N
X1=DMODEL40(I,1) / REGRESSOR REDEFINED AS X1
Y=DMODEL40(I,2) / REGRESSAND REDEFINED AS Y; EXPECTED Y IS YX
C computing the function
yx=0.d0
do j=1,jmax
v=(a(j)/(1.d0-b(j)*tau))*(dexp(-b(j)*x1)-dexp(-x1/tau))
yx=yx+v
enddo
IF(NQ.NE.0) WRITE(11,*) I,x1,Y,YX
ER=Y-YX / ERROR
SER=SER+ER
S=S+ER**2
SY=SY+Y
SSY=SSY+Y**2
ENDDO
RMS2=S/N
V=SSY/N-(SY/N)**2
F=-(1.D0-RMS2/V) / FUNCTION TO MINIMIZE (IT MAXIMIZES R_SQUARE)
f=s
IF(NQ.NE.0) WRITE(11,1) (1.D0-RMS2/V),f,DSQRT(S/N)
1 FORMAT('R-SQUARE = ',F20.15,' S-SQUARE = ', 2F25.15)
IF(NQ.NE.0) CLOSE(11)
RETURN
END

```

Lest it causes confusion, it is to be noted that we have changed the sign of the parameter a_2 in our program so that all the parameters may become positive and lie in the ranges specified by us. This endeavour has saved a line or two in our program. Nevertheless, it was not necessary to do that. The subroutine that defines the OG function is given above (Subroutine REGMODEL_40). It is called by the Main/DE Program for function evaluation. The DE program may be obtained from the author.

Table-3: Sample Results of Differential Evaluation (Seed = 7117) Time Taken : 24 min.52 sec

POPULATION SIZE [N] AND NO. OF ITERATIONS [ITER] ? SUGGESTED : N -> 100 OR ->10.M; ITER 10000 OR 50 100 100000 CROSSOVER PROBABILITY [PCROS] AND SCALE [FACT] ? SUGGESTED : PCROS ABOUT 0.9; FACT=.5 OR LARGER BUT <=1 .5 .9 .01 RANDOM NUMBER SEED ? A FOUR-DIGIT POSITIVE ODD INTEGER, SAY, 1171 7117 COMPUTING --- PLEASE WAIT 0.776364112 1.40070136 0.311494485 1.13397674 0.100847601 0.354311254 0.0561837315 FBEST UPTO NOW = 0.4950342361374442E-01 TOTAL NUMBER OF FUNCTION CALLS = 100000 0.669387275 1.3818906 0.284143457 1.1110071 0.0918613798 0.317632438 0.044231334 FBEST UPTO NOW = 0.1190228961782318E-01 TOTAL NUMBER OF FUNCTION CALLS = 200000 0.571787804 1.48436152 0.233555314 1.45082412 0.0887824799 0.472085661 0.0474569645 FBEST UPTO NOW = 0.2747202134357340E-02 TOTAL NUMBER OF FUNCTION CALLS = 300000 0.522756529 1.85698911 0.208370622 1.87553186 0.096906218 0.527917476 0.0505768034 FBEST UPTO NOW = 0.1435838012874943E-03 TOTAL NUMBER OF FUNCTION CALLS = 400000 0.502718669 1.95719576 0.201307518 1.98162795 0.0985368408 0.52436911 0.0504926088 FBEST UPTO NOW = 0.4053676471497610E-05 TOTAL NUMBER OF FUNCTION CALLS = 500000 0.499470645 1.98190257 0.200416893 1.9953671 0.0993034449 0.513105171 0.0502843493 FBEST UPTO NOW = 0.7009686872335368E-06 TOTAL NUMBER OF FUNCTION CALLS = 600000 0.500596493 1.98965539 0.200397258 1.9931019 0.0997182231 0.503685049 0.0500790062 FBEST UPTO NOW = 0.1213803813840479E-06 TOTAL NUMBER OF FUNCTION CALLS = 700000 KF-40 OG FUNCTION: 7-VARIABLES M-7 NO. OF VARIABLES = 7 COMPUTATION OVER. THANK YOU					
---	--	--	--	--	--

Our results indicate that neither of the two reasons postulated by Han et al. is responsible for the discrepancies in the true parameters and the estimated parameters obtained by them. If it were not so, the DE based least squares algorithm would not have given us the results that are much more close to the true parameters.

References

- Han, XL, Pozdin, V, Haridass, C and Misra, P (2006) "Monte Carlo Least-Squares Fitting of Experimental Signal Waveforms", *Journal of Information & Computational Science*, 3(4), pp. 1-7. http://www.physics1.howard.edu/~pmisra/publications/137_ISICS06.pdf
- Mishra, SK (2007) "Performance of Differential Evolution Method in Least Squares Fitting of Some Typical Nonlinear Curves", SSRN <http://ssrn.com/abstract=1010508>