A METHOD FOR OBTAINING INITIAL ESTIMATES FOR FITTING LINEAR COMBINATIONS OF EXPONENTIALS

By
UMED SINGH
Department of Statistics, The Ohio State University, Columbus, Ohio, 43210, U.S.A.
(Received in November, 1975)

1. INTRODUCTION

Writing $\rho_j = \exp.\ (-\lambda_j)$, $\lambda_j > 0$, the models we are concerned with have the forms

$$y(t) = \sum_{j=1}^{p} \alpha_j \rho_j^t + \varepsilon(t), \tag{1.1}$$

and

$$y(t) = \alpha_0 + \sum_{j=1}^{p} \alpha_j \rho_j^t + \varepsilon(t), \tag{1.2}$$

where the $\varepsilon(t)$ are random errors assumed to be independent with zero means and the parameters are assumed to satisfy the conditions $\alpha_j \neq 0$ for (1.1) and $\alpha_j \neq 0$, $\alpha_0 > 0$ for (1.2).

For the non-linear models above, application of the Least Squares method results in equations which are in general solvable only by iteration. The Least Squares computations have several unusual features when applied to linear combinations of exponentials. The most unusual aspect is the frequent failure of the iterative computation schemes to converge. Secondly, the iterative process converges but the resulting estimators may not be the least squares estimates. These pitfalls of Least Squares computations have been discussed by Cornfield *et. al.* [1960]. For successful implementation of iterative procedure, one needs 'good' initial estimates of parameters appearing in a non-linear fashion in these models. Sometimes the initial estimates may provide the most consistent estimates of the parameters if facilities for computation of iterative least squares are not available.

Cornell [1962] has proposed a general method which provides a simple and direct procedure for estimating the non-linear parameters in the case of the two general models (1.1) and (1.2). Since the method is based on independent partial totals of the sample observations, it has the disadvantage that the estimators obtained are not of ρ_j, but of some integral power of ρ_j. Agha [1971] has given another method which overcomes the disadvantage of Cornell's method of providing estimators of some integral power ρ_j, but utilizes dependent partial totals of the sample observations. Also it assumes that all $\alpha_j > 0$ in the model (1.1). Foss [1969] has proposed a method which is computer oriented and arrives at the initial estimates by a least square 'peeling-off' technique.

In the following sections we suggest an alternative method which provides estimates of ρ_j rather than some integral power of ρ_j as in the case of Cornell's method. Secondly, it utilizes independent partial totals as against the use of over-lapping partial totals in the case of Agha's method. Also in most of the cases this method yields a smaller residual sum of squares than other methods. In case of log transformation of the observations, it is generally noticed that for large t, the curve is approximately a straight line. In such situations the modified form of the proposed method based on sequential estimation technique performs better over the other methods. The proposed method of course, require equally spaced data as required by Cornell and Agha methods.

2. General Estimation Procedure

Consider first the general model (1.1) with p exponentials,

$$y(t) = \sum_{j=1}^{p} \alpha_j \rho_j^t + \varepsilon(t), \quad t=0, 1, 2, \ldots\ldots, n.$$

Also to implement the procedure, we assume $n=2mp-1$, and observations are specified only at equally spaced values of t. Number of observations is equal to $n+1$, $=2p\,m$, that is, m times the number of parameters in the model.

The estimation procedure is as follows :

Partition the sample values into $2\,p$ sums, s_h, given by

$$s_h = \sum_{i=0}^{(m-1)} y\,(h+2pi), \tag{2.1}$$

$$i=0, 1, 2, \ldots\ldots, m-1,$$

$$h=1, 2, \ldots\ldots, 2p.$$

These partial sums s_h, have expectations, s_h, given by

$$E[s_h]=S_h=\sum_{j=1}^{p} \alpha_j \rho_j^h \frac{1-\rho_j^{2pm}}{1-\rho_j^{2p}} \tag{2.2}$$

Since ρ_j are distinct, it is easy to verify that the polynomial, S_h, satisfies the p difference equations

$$\sum_{i=1}^{p+1} (-1)^{2p+1-i} \Lambda_{p+1i} S_{h+i} =0,$$

$$h=0, 1, 2, \ldots\ldots, p-1. \tag{2.3}$$

where, for $r=1, 2, \ldots\ldots, p$, the elementary symmetric functions Λ_r equals the sum of all possible products, that is,

$$\Lambda_r=\sum (\rho_{j1} \rho_{j2} \ldots \rho_{jr}), \tag{2.4}$$

summation is over (p_r) different combinations. Replacing S_h by the corresponding observed partial sums, s_h, in (2.3), we obtain estimators $\hat{\Lambda}_r$ of the Λ_r from the equations

$$\sum_{i=1}^{p+1} (-1)^{2p+1-i} \hat{\Lambda}_{r\; p+1-i}\, S_{h+1}$$

$$=0, h=0, 1, 2, \ldots, p-1. \tag{2.5}$$

Let $\underset{\sim}{A}$ be a $p \times p$ matrix whose jth column is $(s_j, s_{j+1}, \ldots, s_{j+p-1})^T$ and $\underset{\sim}{A}_j$ be the matrix obtained by replacing the $(p+1-j)$th column of $\underset{\sim}{A}$ by the column vector $(s_{p+1}, s_{p+2}, \ldots, s_{2p})^T$. Then by Cramer's rule we have

$$\hat{\Lambda}_r=(-1)^{r+1} \frac{1\underset{\sim}{A}_j 1}{1A1}, r=1, 2, \ldots, p. \tag{2.6}$$

Since the $\hat{\Lambda}_r$ estimate the elementry symmetric functions of the ρ_j, the estimators r_j of the ρ_j is given by the p roots of the equation

$$x^p - \frac{\underset{\backsim}{1}A_1\underset{\backsim}{1}}{\underset{\backsim}{1}A\underset{\backsim}{1}} x^{p-1} - \frac{\underset{\backsim}{1}A_2\underset{\backsim}{1}}{\underset{\backsim}{1}A\underset{\backsim}{1}} x^{p-2} - \ldots - \frac{\underset{\backsim}{1}A_p\underset{\backsim}{1}}{\underset{\backsim}{1}A\underset{\backsim}{1}} = 0 \tag{2.7}$$

For estimators $\hat{\lambda}$ of the λ_j, we take $\hat{\lambda}_j = -\log_e \hat{r}_j$. The estimators a_j of α_j are then obtained by solving any p-equations of the set

$$\sum_{j=1}^{p} r_j^h \frac{1-r_j^{2pm}}{1-r_j^{2p}} a_j = s_h,\ h=1, 2, \ldots, 2p. \tag{2.8}$$

The method of partial sums may similarly be applied to the model (1.2). Here we assume that there are $n+1=(2p+1)m$ observations. We form the partial sums, s_h^*,

$$s_h^* = \sum_{i=0}^{(m-1)} y\,[h+(2p+1)i], \tag{2.9}$$

$$i=0, 1, 2, \ldots, m-1,$$

$$h=1, 2, \ldots, 2p+1.$$

clearly,

$$E\left[\ s_h^*\ \right] = s_h^* = m\alpha_0 + \sum_{j=1}^{p} \alpha_j \rho_j^h \frac{\left(1-\rho_j^{(2p+1)m}\right)}{\left(1-\rho_j^{2p+1}\right)} \tag{2.10}$$

From the $S_h^{*'}$ we form the differences

$$S_h^* = S_h^* - S_{h+1}^* \tag{2.11}$$

and similarly define

$$\bar{s}_h^{*'} = s_h^* - s_{h+1}^* . \tag{2.12}$$

Utilizing the same procedure as before for $S_h^{*'}$ and $s_h^{*'}$ the solution for the estimators $\hat{\Lambda}_r$ of the Λ_r is thes ame in terms of the $s_h^{*'}$ as that given by (2.6) in terms of the s_h. a_j will be obtained in the same manner as before. Estimator a_0 of α_0 is determined by

$$m\, a_0 = s_h^{*} - \sum_{j=1}^{p} a_j\, r_j^{h} \frac{\left(1 - r_j^{(2p+1)m}\right)}{\left(1 - r_k^{2p+1}\right)} \tag{2.13}$$

There may be situations where exponentials are well separated in time (t), that is, when $\lambda_i >> \lambda_j$, ($i > j$, $i, j = 1, 2, ..., p$), yield data known as 'decay type' data. In such situations a modification in forming partial sums is recommended. Partial sums may be formed sequentially with first $4p$ or $6p$ or $8p$ observations for model (1.1) and with first 2 ($2p+1$), or 3 ($2p+1$) or 4 ($2p+1$) observations for model (1.2). The initial estimates based on modified partial sums are better if error variances are large. In pharmacokinetic studies it is generally not practicable to collect data at equi-spaced time intervals after some stage of collection of data. In such situations the modified sequential estimation procedure of partial sums still works.

3. Examples

In this section we apply the estimation procedure developed in Section 2 to the numerical examples reported by Cornell [1962] and compare it with other methods due to Agha [1971], Cornell [1962] and Foss [1969].

3.1. One Exponential Term

The data on counts describing the decay of the neutron density in a medium-size assembly of beryllium is reported in Table 1. Observations were made at equally spaced time interval of 0.1 milliseconds.

TABLE 1

Decay of the Neutron Density in a Medium-Sized Assembly of Beryllium.

t	0	1	2	3	4	5	6	7	8
$y(t)$	100145	78005	60305	46485	336205	28275	21705	16955	13045
t	9	10	11	12	13	14	15	16	17
$y(t)$	10085	7835	6165	4782	3780	2915	2249	1752	1395

Cornell's estimators of ρ and α are

$$r_c = 0.776{,}06,$$

$$a_c = 100{,}043.$$

Agha's estimators of ρ and α are

$$r_a = 0.77592,$$

$$a_a = 100{,}089.$$

Estimators of ρ and α based on (2.7) and (2.8) are

$$r = 0.77588,$$

$$a = 100{,}156.$$

The residual sum of squares $\Sigma[y(t) - a\, r^t]^2$ for the three estimation methods are in Table 2.

TABLE 2

Method	*Res. S. S.*
A	282,380
C	315,265
S	263,196

It is obvious that the proposed method gives a considerable reduction in the residual sum of squares.

3.2. Two Exponential Terms

We apply the various estimation methods to the data in Table 3. The observations describe the distribution of background pulses generated in a proportional counter by neutron interaction with the walls and gas plus pulses due to circuit noise. The pulse heights t are recorded at equi-spaced intervals.

TABLE 3

Logarithms $y(t)$ of Frequencies of Pulse Heights t Generated in a Proportional Counter.

$t =$	0	1	2	3	4	5	6	7
$y(t) =$	10.430	4.703	2.327	1.140	0.615	0.325	0.170	0.117
$t =$	8	9	10	11	12	13	14	15
$y(t) =$	0.05	0.04	0.046	0.022	0.036	0.021	0.018	0.016

The estimators of the parameters and the residual sum of squares $\Sigma[y(t)-a_1r^t{}_1-a_2r^t{}_2]^2$ are calculated for the various methods and are given in Table 4.

TABLE 4

Method	r_1	r_2	a_1	a_2	*Res. S. S.*
A	0.5490	0.2533	6.9284	3.4825	0.011,060,72
C	0.9961	0.4978	0.0220	9.9030	0.351,092.55
F	0.8596	0.3113	9.2941	1.1341	0.017,480,00
S	0.5046	0.0734	9.1854	1.2640	0.013,094,00

Notice that the reduction in residual sum of squares due to the proposed method is drastic over Cornell's method and it compares very well with Agha's method in this case.

Assuming that the random errors $\varepsilon(t)$ are independently and normally distributed with means zero and common variance, the iterative maximum likelihood estimators of ρ_1, ρ_2, α_1 and α_2 were obtained using the four sets of initial estimates. The results are shown in Table 5.

TABLE 5

Initial Estimates Method	Maximum Likelihood Estimates r_1	r_2	a_1	a_2	No. of Iterations	Res. M. S.
A	0.5188	0.1860	8.3093	2.1207	7	$.35161\times10^{-3}$
C	0.3000	0.5519	4.1305	6.2995	21	$.45862\times10^{-3}$
F	0.5188	0.1860	8.3093	2.1207	32	$.35161\times10^{-3}$
S	0.5188	0.1860	8.3093	2.1207	9	$.35161\times10^{-3}$

Notice that the iterative maximum likelihood estimators of the parameters are the same for the initial values provided by all methods except by Cornell's method. However, the number of iterations with initial estimates by Foss's method are considerably large. Unfortunately Cornell's methods does not seem to provide the right answer inspite of 21 iterations and the residual mean square is also considerably high.

4. Consistency of the Estimators

The partial-totals estimator are not in general unbiased since they are solution of polynomials, they are consistent estimators. The proof of consistency follows along the lines of Cornell [1962] and we outline it below for our case. Consider the model (1.1). Suppose the errors are independently distributed for all t and are identically distributed for all t in the same group. Define group means

$$\bar{y}_h = \frac{s_h}{m}.$$

Replacing the sum s_h by the corresponding $\bar{y}_h$ in equation (2.1) we have by law of large numbers

$$p \lim_{m\to\infty} \bar{y}_h = \lim_{m\to\infty} \frac{1}{m} \sum_{i}^{2p(m-1)+h} \sum_{j}^{p} \alpha_j \rho_j^{h+2pi} \tag{4.1}$$

$$i=0, 1, 2, \ldots, m-1,$$
$$h=1, 2, \ldots, 2p.$$

Letting $m\to\infty$, keeping the domain i constant in length, say T, the m observations included in the hth group are made for

$$i=\frac{(h-1)T}{2pm},\ \frac{(h-1)T}{2pm}+\frac{T}{m},\ \frac{(h-1)T}{2pm}+\frac{2T}{m},\ \ldots,\ \frac{(h-1)T}{2pm}+\frac{m-1}{m}T.$$

Now (4.1) may be written as

$$p\lim_{m\to\infty}\bar{y}_h=\frac{1}{T}\lim_{m\to\infty}\sum_{k=\frac{h-1}{2p}}^{\frac{h-1}{2p}+m-1}\sum_{j=1}^{p}\alpha_j\rho_j^{k\frac{T}{m}}\cdot\frac{T}{m} \tag{4.2}$$

The expression on right is definite integral equal to

$$\psi=\frac{1}{T}\sum_{j=1}^{p}\frac{\alpha_j}{\lambda_j}\left(1-\rho_j^{T}\right) \tag{4.3}$$

At the point ψ, $r_j=\rho_j$ and $\lim_{m\to\infty} a_j=\alpha_j$ for all j.

Then with the ρ_j distinct as specified by our model, derivative of all orders af the estimators r_j and a_j are continuous in the neighborhood of ψ and Slutsky's Theorem as given by Cramer [1946, *p.* 255] is applied to show that r_j and a_j converge in probability to ρ_j and α_j, respectively, for all j. Thus r_j and a_j are consistent estimators of ρ_j and α_j, respectively, for all j when $m\to\infty$. This is also true for the model with constant term α_0 added.

SUMMARY

This paper describes a technique for obtaining the initial estimates for fitting linear combinations of exponentials. The method utilizes independent partial totals and provides a simple and direct procedure for estimating the non-linear and the linear parameters. Modifications are presented that make the estimation procedure more versatile to decay-type data where the exponentials are well separated. The procedure is illustrated with two examples from the literature.

REFERENCES

1. Agha, M. [1971]. : A direct method for fitting linear combinations of exponentials. Biometrics 27, 399-413.
2. Cornell, R.G. [1962]. : A method for fitting linear combination of exponentials. Biometrics 18, 104-13.
3. Cornfield, J., Steinfeld. J., and Greenhouse, S.W. [1960]. : Models for the interpretation of experiments using tracer compounds. Biometrics 16, 212-34.
4. Cramer, H. [1946] : Mathematical Methods of Statistics. Princeton University Press, Princeton, New Jersey.
5. Foss, S. [1969]. : A method for obtaining initial estimates of the parameters in exponential curve fitting. Biometrics 25, 580-4.