SANKHYĀ

THE INDIAN JOURNAL OF STATISTICS

Editors: P. C. MAHALANOBIS, C. R. RAO

SERIES B, VOL. 30 — JUNE 1968 — PARTS 1 & 2

A SIMPLE METHOD FOR FITTING THE REGRESSION CURVES DERIVABLE FROM A SECOND DEGREE DIFFERENCE EQUATION

By C. G. KHATRI

Indian Statistical Institute

SUMMARY. The solutions of the second degree difference equation $Y_{x+2}-\omega_1 Y_{x+1}+\omega_2 Y_x = \alpha_0+\alpha_1 x+\ldots+\alpha_k x^k$ where $\omega_1, \omega_2, \alpha_0, \alpha_1, \ldots, \alpha_k$ are constants, depend on the roots of the quadratic equation $y^2-\omega_1 y+\omega_2 = 0$. As for example, if the roots are real and distinct, then the regression equation is given by $Y_x = \alpha+\beta_1 x+\ldots+\beta_k x^k+\delta_1\rho_1^x+\delta_2\rho_2^x$. Thus, the fitting of these types of regression curves depends on the estimates of ω_1 and ω_2. Here, a simple method based on orthogonal polynomials is suggested for the estimation of ω_1 and ω_2, and consequently for ρ_1 and ρ_2.

1. Introduction

Let us consider a second degree difference equation given by

$$Y_{x+2}-\omega_1 Y_{x+1}+\omega_2 Y_x = \alpha_0+\alpha_1 x+\alpha_2 x^2+\ldots+\alpha_k x^k \quad \ldots \quad (1)$$

where $\omega_1, \omega_2, \alpha_0, \alpha_1, \ldots, \alpha_k$ are real numbers and the values of x are given for $x = 1, 2, \ldots, n$. The solution of (1) will depend on the roots of the quadratic equation in y

$$y^2-\omega_1 y+\omega_2 = 0. \quad \ldots \quad (2)$$

If the roots of (2) are real and different, say, ρ_1 and ρ_2, then a solution of (1) is given by

$$Y_x = \alpha+\beta_1 x+\beta_2 x^2+\ldots+\beta_k x^k+\delta_1\rho_1^x+\delta_2\rho_2^x \quad \ldots \quad (3)$$

where $\alpha, \beta_1, \ldots, \beta_k, \rho_1$ and ρ_2 are functions of $\omega_1, \omega_2, \alpha_0, \alpha_1, \ldots, \alpha_k$ and δ_1 and δ_2 are constants of integration. If the roots of (2) are real and equal, say, ρ, then a solution of (1) is given by

$$Y_x = \alpha+\beta_1 x+\beta_2 x^2+\ldots+\beta_k x^k+\rho^x(\gamma_1+\gamma_2 x) \quad \ldots \quad (4)$$

and if the roots of (2) are complex numbers, then a solution of (1) is given by

$$Y_x = \alpha+\beta_1 x+\beta_2 x^2+\ldots+\beta_k x^k+\rho^x(\gamma_1 \cos \mu x+\gamma_2 \sin \mu x) \quad \ldots \quad (5)$$

where $\rho = \omega_2^{1/2}$ and $\mu = \cos^{-1}(\omega_1/2\rho)$.

For fitting the data on force of mortality, Khatri and Shah (1959) considered the difference equation (1) when $k = 0$ and 1, and the fitting of the corresponding curves (3), (4), and (5) was done by the extension of the internal least squares method given by Hartley (1948). When $\rho_2 = 0$ (or $\omega_2 = 0$) in (3) (or (1)), method based on linear functions of observations was given by Patterson (1956) for $k = 0$, Shah (1961) for $k = 1$ and Chopra (1965) for any k. In this note, we extend their method to the three situations given in (3), (4) and (5). This method makes use of the difference equation (1) and the estimates of the parameters ω_1 and ω_2 are based on quadratic functions of the observations via similar linear functions occurring in Chopra's method (1965). It may be noted that when ω_1 and ω_2 of (1) (i.e. ρ_1 and ρ_2 of (3), ρ of (4), ρ and μ of (5)) are known (or estimated), then other parameters of (3) or (4) or (5) can be obtained by least squares method. Hence the problem is restricted to the estimation of ω_1 and ω_2.

2. A simple method

Let the n observations on Y_x be denoted by $y_1, y_2, \ldots, y_n$ when the respective values of x are $1, 2, \ldots, n$. Let us assume that they are uncorrelated and have equal variance, σ^2, when x is fixed. Moreover, assume that the expected values of y_x satisfy the difference equation given in (1).

As done by Chopra (1965), let us choose $\mu_1, \mu_2, \ldots, \mu_{n-3}$ depending on x such that μ_i, $i = 1, 2, \ldots, n-3$, are the coefficients of ξ'_{k+1} (orthogonal polynomial of degree $k+1$) corresponding to $(n-3)$. Then, it is fairly obvious that

$$\sum_{i=1}^{n-3} i^j \mu_i = 0 \qquad \text{for} \quad j = 0, 1, 2, \ldots, k. \qquad \ldots \quad (6)$$

Further, let us define

$$T_i = \sum_{j=1}^{n-3} \mu_j y_{j+i-1} \qquad \text{for} \quad i = 1, 2, 3, 4. \qquad \ldots \quad (7)$$

Then, with the help of (6) and (1), it is easy to verify that

$$E(T_4 - \omega_1 T_3 + \omega_2 T_2) = 0 \quad \text{and} \quad E(T_3 - \omega_1 T_2 + \omega_2 T_1) = 0. \qquad \ldots \quad (8)$$

Hence, our approximate estimates $\hat{\omega}_1$ and $\hat{\omega}_2$ will be given by the solution of the equations

$$T_4 = T_3 \hat{\omega}_1 - \hat{\omega}_2 T_2 \quad \text{and} \quad T_3 = \hat{\omega}_1 T_2 - \hat{\omega}_2 T_1 \qquad \ldots \quad (9)$$

or $\hat{\omega}_1$ and $\hat{\omega}_2$ are given by

$$\hat{\omega}_1 = \frac{T_3 T_2 - T_4 T_1}{T_2^2 - T_3 T_1} \quad \text{and} \quad \hat{\omega}_2 = \frac{T_3^2 - T_4 T_2}{T_2^2 - T_3 T_1} \text{ if } T_2^2 \neq T_3 T_1. \qquad \ldots \quad (9')$$

By usual method, the asymptotic variance-covariance matrix of $\hat{\omega}_1$ and $\hat{\omega}_2$ is given by

$$\frac{\sigma^2}{(a_2^2 - a_3 a_1)^2} \begin{pmatrix} -a_1 & a_2 \\ -a_2 & a_3 \end{pmatrix} \begin{pmatrix} e_1 & e_2 \\ e_2 & e_1 \end{pmatrix} \begin{pmatrix} -a_1 & -a_2 \\ a_2 & a_3 \end{pmatrix} \qquad \ldots \quad (10)$$

where $a_i = \sum_{j=1}^{n-3} \mu_j Y_{j+i-1}$ for $i = 1, 2, 3$,

$$e_1 = (1+\omega_1^2+\omega_2^2) \sum_{i=1}^{n-3} \mu_i^2 - 2\omega_1(1+\omega_2) \sum_{i=1}^{n-4} \mu_i \mu_{i+1} + 2\,\omega_2 \sum_{i=1}^{n-5} \mu_i \mu_{i+2}$$

$$e_2 = -(1+\omega_2)\omega_1 \sum_{i=1}^{n-3} \mu_i^2 + (1+\omega_1^2+\omega_2^2+\omega_2) \sum_{i=1}^{n-4} \mu_i \mu_{i+1} - (1+\omega_2)\omega_1 \sum_{i=1}^{n-5} \mu_i \mu_{i+2} + \omega_2 \sum_{i=1}^{n-6} \mu_i \mu_{i+3}.$$

Now, for determining the types of the curve, let us consider the quadratic equation in y :

$$(T_2^2-T_3T_1)y^2-(T_2T_3-T_4T_1)y+(T_3^2-T_4T_2)=0. \quad \ldots \quad (11)$$

The equation (11) will have in general two roots.

2.1. Let the roots of (11) be real and unequal, say, y_1 and y_2. Then, we fit the regression curve given by (3) and take the estimates $\hat{\rho}_1$ and $\hat{\rho}_2$ as y_1 and y_2 respectively. In this case, the asymptotic variances and convariance of $\hat{\rho}_1$ and $\hat{\rho}_2$ are given by

$$V(\hat{\rho}_1)=\frac{\sigma^2}{f_1^2\rho_1^2(\rho_1-\rho_2)^4}[(1+\rho_2^2)e_1-2\rho_2e_2],$$

$$V(\hat{\rho}_2)=\frac{\sigma^2}{f_2^2\rho_2^2(\rho_1-\rho_2)^4}[(1+\rho_1^2)e_1-2\rho_1e_2],$$

and $$\operatorname{cov}(\hat{\rho}_1,\hat{\rho}_2)=\frac{\sigma^2}{f_1f_2\rho_1\rho_2(\rho_1-\rho_2)^4}[(1+\rho_1\rho_2)e_1-(\rho_1+\rho_2)e_2] \quad \ldots \quad (12)$$

where $f_i=\delta_i\sum_{j=1}^{n-3}\mu_j\rho_i^{j-1}$, $i=1,2$ and e_1 and e_2 are the same as defined in (10) after replacing ω_1 and ω_2 by $(\rho_1+\rho_2)$ and $\rho_1\rho_2$ respectively.

2.2. Let the roots of (11) be complex. Then, we fit the regression curve given by (5) and take the estimates $\hat{\rho}$ and $\hat{\mu}$ as

$$\hat{\rho}=\hat{\omega}_2^{1/2} \text{ and } \hat{\mu}=\cos^{-1}(\hat{\omega}_1/2\hat{\rho}). \quad \ldots \quad (13)$$

The asymptotic variance-covariance matrix of $\hat{\mu}$ and $\hat{\rho}$ is given by

$$\frac{\sigma^2}{4(a_2^2-a_3a_1)}\begin{pmatrix} a_4 & -a_5 \\ -a_2 & a_3 \end{pmatrix}\begin{pmatrix} e_1 & e_2 \\ e_2 & e_1 \end{pmatrix}\begin{pmatrix} a_4 & -a_2 \\ -a_5 & a_3 \end{pmatrix} \quad \ldots \quad (14)$$

where e_1 and e_2 are the same as defined in (10) after replacing ω_1 and ω_2 by $2\rho\cos\mu$ and ρ^2 respectively,

$$a_{i+1}=\rho^i[\{\gamma_1\cos(i\mu)+\gamma_2\sin(i\mu)\}b_1+\{-\gamma_1\sin(i\mu)+\gamma_2\cos(i\mu)\}b_2]$$

and $$a_{j+3}=\rho^{j-1}[\{\gamma_1\sin(j\mu)-\gamma_2\cos(j\mu)\}b_1+\{\gamma_1\cos(j\mu)+\gamma_2\sin(j\mu)b_2\}]$$

for $i=0,1,2$ and $j=1,2$, with

$$b_1=\sum_{j=1}^{n-3}\mu_j\rho^{j-1}\cos(j\mu) \text{ and } b_2=\sum_{j=1}^{n-3}\mu_j\rho^{j-1}\sin(j\mu).$$

2.3. Let the roots of (11) be equal. In practice, the two roots will not be exactly equal. Hence if the two real roots are nearly equal (i.e. ratio of the two roots is nearly 0.99 or 1.01) or if the two roots are complex having the imaginary part zero compared to the real part, we shall fit the regression curve given by (4) and shall take the estimate $\hat{\rho}$ as $\hat{\omega}_1/2$. The asymptotic variance of $\hat{\rho}$ can be obtained from (10).

In this case, we can slightly improve the estimate $\hat{\rho}$ given above in the following way :

Let us choose $\eta_1, \eta_2, \ldots, \eta_{n-2}$ depending on x such that γ_i, $i=1,2,\ldots,n-2$, are the coefficients of ξ'_{i+1} corresponding to $(n-2)$. Further, let $t_i=\sum_{j=1}^{n-2}\eta_jy_{j+i-1}$

for $i = 1, 2, 3$. Then as in the general case, the estimate of ρ will be a solution of the quadratic equation in z given by

$$t_1 z^2 - 2t_2 z + t_3 = 0. \qquad \dots \quad (15)$$

Note that in practice there will be two solutions of (15). We expect that the solutions will be real provided the solutions of (11) are nearly equal. We shall choose the solution z of (15) as the estimate of ρ, which is near to $\hat{\omega}_1/2$. The asymptotic variance of the estimate of ρ obtained in this way is

$$\frac{\sigma^2}{4\gamma_2^2\rho^4 f_2^2}\, e_1 \text{ where } f_2 = (\eta_1 + \eta_2\rho + \dots + \eta_{n-2}\rho^{n-3}) \qquad \dots \quad (16)$$

and e_1 is the same as that defined in (10) after replacing μ_i's, ω_1 and ω_2 by η_i's, $2\rho \cos\mu$ and ρ^2 respectively.

3. An illustration

Let us consider the curve given by

$$y = 3 + (0.5)^x(1+x).$$

Then, the data for the above curve are given by

x	0	1	2	3	4	5	6
y	4.00	4.00	3.75	3.50	3.31	3.19	3.11

and we have $T_1 = -1.75$, $T_2 = -2.32$, $T_3 = -1.87$ and $T_4 = -1.29$. The equation (11) becomes

$$2.1099y^2 - 2.0809y + 0.5141 = 0.$$

The two solutions are (2.1739/4.2198, 1.9870/4.2198). The ratio of the two roots is nearly 1.01. Hence, we fit the regression (4). The estimate of ρ can be taken as $2.0809/4.2198 = 0.493$.

Now, for the improvement on this value of $\hat{\rho}$, we have $t_1 = -1.88$, $t_2 = -2.06$ and $t_3 = -1.59$ and this gives the equation (15) as

$$1.88z^2 - 4.12z + 1.59 = 0.$$

The two solutions of z are 0.50 and 1.7. The solution near to .493 is 0.50. Hence the improved estimate of ρ is 0.50 which is the exact value of the data.

References

Chopra, A. S. (1965): Application of orthogonal polynomials in fitting an asymptotic regression curve. *Jour. Ind. Soc. Agri. Stat.*, 17, 76-82.

Hartley, H. O. (1948): The estimation of non-linear parameters by internal least squares. *Biometrika*, 35, 32-45.

Khatri, C. G. and Shah, B. K. (1959): Method of fitting certain curves of force of mortality over the whole range. *Journal of M. S. University of Baroda*, 8, 35-40.

Patterson, H. D. (1956): A simple method for fitting an asymptotic curve. *Biometrics*, 12, 323-329.

Shah, B. K. (1961): A simple method of fitting the regression curve $Y = a + \delta x + \beta\rho^x$. *Biometrics*, 17, 651-653.

Paper received : January, 1968.