韓國數學敎育學會誌<數學敎育>
1984. 12, Vol. XXIII, No. 1

Journal of the Korea Society of Mathematical Education
Dec. 1984, Vol. XXIII, No. 1

Computer를 使用한 時系列 分析

江原大學校 白 淸 鎬

1. 緖 論

企業과 產業 그리고 社會全般, 더 나아가서는 人間生活의 여러面에서, 過去의 經驗을 土臺로 한 未來에 對한 豫測은 그 必要性이 漸次 增大되고 있다.

주어진 資料의 성질에 따라 여러가지 豫測技法이 硏究되어 왔으며 그러한 硏究는 現在도 進行中이다. 本 硏究는 週期性을 갖는 過去의 資料에 의하여, 線型함수와 三角함수를 Base로 하는 線型複合多項式으로서의 時系列 모형을 設定하고, computer를 利用하여 最小自乘法으로 예측함수를 구하고 그 結果를 분석하고자 한다.

2. 豫測함수의 設定

주어진 資料의 全體的인 傾向變動(trend variation)을 고려한 선형함수 $h_1(t)=a_1+a_2t$와 주기성을 고려한 삼각함수 $h_2(t)=\sum_{k=1}^{\infty}\left(b_k\sin\frac{2\pi}{l}kt+c_k\cos\frac{2\pi}{l}kt\right)$ (l은 주기)를 써서 다음과 같은 함수

$$F(t)=a_1+a_2t+\sum_{k=1}^{\infty}\left(b_k\sin\frac{2\pi}{l}kt+c_k\cos\frac{2\pi}{l}kt\right)$$

를 만든다. 그러나 실제적인 이용과 그에 따르는 computer에 의한 계산을 위하여 時系列 모형 즉, 예측을 위한 함수의 모형을

$$y=f(t)=b_1+b_2t+\left(b_3\sin\frac{2\pi}{l}\cdot1\cdot t+b_4\cos\frac{2\pi}{l}\cdot1\cdot t\right)+\left(b_5\sin\frac{2\pi}{l}\cdot2\cdot t+b_6\cos\frac{2\pi}{l}\cdot2\cdot t\right)$$
$$+\cdots+\left(b_{n-1}\sin\frac{2\pi}{l}\cdot N\cdot t+b_n\cos\frac{2\pi}{l}\cdot N\cdot t\right) \qquad (1)$$

(단 n은 짝수, $N=\frac{n-1}{2}$의 정수부분)

과 같이 설정한다.

T개의 資料 (t_i, y_i) $(i=1, 2, \cdots, T)$가 주어졌을 때 (1)식의 n을 變化시키면서 computer를 이용하여 계산하고, 誤差의 제곱의 합 M이 최소로 되는 n을 구하여 時系列 모형을 얻는다. 단, $n\leqq T$이다.

3. 理論的 背景

T개의 資料 (t_i, y_i) $(i=1, 2, \cdots, T)$가 있을 때 예측함수를,

$$y=b_1g_1(t)+b_2g_2(t)+\cdots\cdots+b_ng_n(t)$$

라 할 때,

$$M=\sum_{i=1}^{T}\left[b_1g_1(t_i)+b_2g_2(t_i)+\cdots+b_ng_n(t_i)-y_i\right]^2$$

이 최소가 되려면, M을 $b_1, b_2, \cdots\cdots, b_n$의 선형복합다항식으로 생각하여 $\frac{\partial M}{\partial b_1}=0$, $\frac{\partial M}{\partial b_2}=0, \cdots$, $\frac{\partial M}{\partial b_n}=0$

이 되는 $b_1, b_2, \cdots\cdots, b_n$을 구한다.

$$\begin{cases} \dfrac{\partial M}{\partial b_1}=2\sum_{i=1}^{T}\left[b_1g_1(t_i)+b_2g_2(t_i)+\cdots+b_ng_n(t_i)-y_i\right]g_1(t_i)=0 \\ \dfrac{\partial M}{\partial b_2}=2\sum_{i=1}^{T}\left[b_1g_1(t_i)+b_2g_2(t_i)+\cdots+b_ng_n(t_i)-y_i\right]g_2(t_i)=0 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \dfrac{\partial M}{\partial b_n}=2\sum_{i=1}^{T}\left[b_1g_1(t_i)+b_2g_2(t_i)+\cdots+b_ng_n(t_i)-y_i\right]g_n(t_i)=0 \end{cases}$$

에서 $\sum_{i=1}^{T}$를 $\sum$로, $g_j(t_i)=g_j$로 쓰면

$$\begin{cases} b_1\sum g_1^2+b_2\sum g_1g_2+\cdots\cdots+b_n\sum g_1g_n=\sum g_1y_i \\ b_1\sum g_2g_1+b_2\sum g_2^2+\cdots\cdots+b_n\sum g_2g_n=\sum g_2y_i \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ b_1\sum g_ng_1+b_2\sum g_ng_2+\cdots\cdots+b_n\sum g_n^2=\sum g_ny_i \end{cases}$$

를 얻게되며, 이것을 행렬로 표시하여

$$\begin{pmatrix} \sum g_1^2 & \sum g_1g_2\cdots\cdots\sum g_1g_n \\ \sum g_2g_1 & \sum g_2^2\ \cdots\cdots\sum g_2g_n \\ \cdots\cdots & \\ \sum g_ng_1 & \sum g_ng_2\cdots\cdots\sum g_n^2 \end{pmatrix}\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}=\begin{pmatrix} \sum g_1y_i \\ \sum g_2y_i \\ \vdots \\ \sum g_ny_i \end{pmatrix}$$

로 쓰고, 이것을 다시 $GB=Z$로 표현하면, $B=G^{-1}Z$로 구해진다. 이것은 최소 자승법에 의한 것이다.

여기서 G를 구하는 방법은 다음과 같다.

$$X=\begin{pmatrix} g_1(t_1) & g_2(t_1)\cdots\cdots\cdots\cdots g_n(t_1) \\ g_1(t_2) & g_2(t_2)\cdots\cdots\cdots\cdots g_n(t_2) \\ \cdots\cdots & \cdots\cdots \\ g_1(t_T) & g_2(t_T)\cdots\cdots\cdots\cdots g_n(t_T) \end{pmatrix}$$

다 하면, G는 X의 轉置행렬 X^t와 X와의 곱으로 표현된다. 즉, $G=X^tX$이다. 또 함수값의 列 Vector $(y_1, y_2, \cdots, y_T)$를 Y라 하면 Z는 X^tY로 표현된다. 따라서.

$$\begin{aligned} B&=G^{-1}Z \\ &=(X^tX)^{-1}X^tY \end{aligned}$$

로 구할 수 있고, 이때 예측함수에 의한 Y의 推定值 $\hat{Y}$는,

$$\begin{aligned} \hat{Y}&=XB \\ &=XG^{-1}Z \\ &=X(X^tX)^{-1}X^tY \end{aligned}$$

로 구할 수 있다.

本 연구에서 사용된 時系列 모형에서는 $g_1(t)=1$, $g_2(t)=t$, $g_j(t)=\begin{cases} \sin\dfrac{2\pi}{l}Nt & (j\text{가 홀수}) \\ \cos\dfrac{2\pi}{l}Nt & (j\text{가 짝수}) \end{cases}$

$\left(\text{단 } j=3, 4, \cdots, n,\ n\text{은 짝수},\ N\text{은 } \dfrac{j-1}{2}\text{의 정수부분}\right)$

이다. 이때 $\hat{Y}$의 계산순서는 그림과 같다.

$$\hat{Y}=\underbrace{\underbrace{X\underbrace{(\underbrace{X^tX}_{\textcircled{1}})^{-1}}_{\textcircled{2}}}_{\textcircled{3}}\underbrace{X^tY}_{\textcircled{4}}}_{\textcircled{5}}$$

이러한 回歸法에 의한 예측에서, s시간후의 時系列 y_{T+s}의 예측값은

$$\hat{y}_{T+s}=b_1+b_2\cdot(T+s)+\sum_{j=1}^{n} b_j\cdot g_j(T+s)$$

이다.

4. 結 論

이론적으로는 T개의 資料에 대하여 $b_1, b_2, \cdots, b_{2(T+1)}$을 구하여 예측함수를 만드는 것이 바람직하나, computer를 사용하여 계산할 경우 項數가 많아짐에 따라 演算의 回數가 증가하여 결과적으로 誤差의 累積 및 傳播가 많아지게 되어 結果에 심각한 영향을 미치게 된다.

例를 들어 試驗한 結果에 의하면, 36개의 資料에 대한 週期를 12로 하였던 바, b_{14}까지 구하여 계산된 결과의 M이 最小가 되었다. 따라서 예측함수에서 제 몇 항까지를 使用해야 제일 효율적인가 하는 문제는 資料의 갯수, 資料의 波動형태와 傾向變動, 그리고 computer의 演算回數 等의 事項이 力動的으로 調和되어서 解決되어야 할 것이다.

만일 資料의 傾向變動이 2次곡선의 모양이면 $h_1(t)=b_1+b_2t+b_3t^2$ 等으로 선정할 수도 있으며, 波動형태에 따라서는 $h_2(t)$도 여러가지 다른 함수로 바꾸어 사용할 수도 있다. 特히 行列을 이용하여 계산하고자 할 때는, 예측함수의 Base로서 $g_i(t)$를 직교함수로 設定하면 行列 G가 對角행렬이 되어 연산이 대폭적으로 줄어들어 誤差의 영향을 감소시킬 수 있으나, 그러한 함수의 設定은 資料의 모양과 特性을 충분히 고려해야 한다.

參 考 文 獻

1. Francis B. Hildebrand, *Advanced Calculus for Applications*, Second edition, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1976.
2. Montgomery, D.C. and L.A. Johnson, *Forecasting and Time Series Analysis*, McGraw-Hill, Inc., New York, 1976.
3. 金宇哲外 七人, 現代統計學, 英志文化史, 1984.
4. 朴在年, 수치해석, 正益社, 1983.
5. 金鍾浩, 洪錫强, 數値解析, 集賢社, 1984.