On Stochastic Interpolation of Omitted Observation

by

M. Ogawara

Meteorological Research Institute

(Received May 31, 1950)

Abstract

Most series of meteorological observations may be considered as a stationary autoregressive stochastic process of order, say, h. In such a series, it is proved that, for the interpolation of an omitted observation by a linear scheme, the observations more than h terms apart before or after the omitted term need not be taken into consideration. Precision formulae, by which we can indicate the probability that the true value be within any range around the estimated value, are derived for various cases.

The theories and formulae are extended to the case of vector or multivariate time series which may often come out in practice.

1. Introduction

When, in a series of meteorological observations, one of them has been omitted on account of instrument trouble or some other accident, it is often necessary to estimate the value of the quantity which thus escaped measurement. Complexity of meteorological phenomena, roughness of observation network, and deficiency of the number of times of observation all preclude deterministic evaluation of such a lost value; thus, an estimated value, anyhow obtained, is never free from an amount of error. In a scientific interpolation it must be manifestly indicated that what is the probability that the true value would be within any range around the estimated value.

As regards stochastic interpolation in infinite stochastic sequences, some theoretical researches have already been made by A. Kolmogoroff (1) and others (2). We propose, in this paper, theories and formulae that may be directly applicable to such problems of meteorological technique as mentioned above, which concern finite stochastic sequences.

Our problems of stochastic interpolation may be classified as follows:

(a) when we use only the values of observed quantities obtained at that station where the failure took place before and after the hour of the failure of observation;

(1) A. Kolmogoroff: Sur l' interpolation et extrapolation des suites stationnaires, C.R. 208 (1939).

(2) e.g. G. Maruyama: The harmonic analysis of stationary stochastic processes, Mem. Fac. Science., Kyûsyû Univ., 4 (1949).

(b) when we use only the values of observed quantities obtained at the neighbouring stations at the same hour that the failure took place;

(c) when (a) and (b) are combined.

Among these, (a) may be reduced to interpolation by time series analysis and (b) involves multiple correlation, while the most general case is (c) which is to be treated as a vector time series.

In this paper, the stochastic sequence is assumed to be stationary and Gaussian (or normal) process. That which is not stationary may be stationarized, or that which is not normal may be so transformed that its distribution law is normal. Application of our method is then possible in either case.

2. General remarks in the one-dimensional case

Let the variation of a certain meteorological element $X(t)$ be a stationary stochastic process, and its mean value be zero, so that $E\{X(t)\}=0$ [(3)]. Let us assume that the observation of $X(t)$ at the hour t was omitted. In order to interpolate the $X(t)$ by the values obtained at equal intervals before and after that hour,

(2. 1) $$\cdots\cdots, X_{t-m}, \cdots\cdots, X_{t-1}, X_{t+1}, \cdots\cdots X_{t+n}, \cdots\cdots,$$

we adopt a linear scheme

(2. 2) $$Y_{mn}(t)=a_{-m}X(t-m)+\cdots\cdots+a_{-1}X(t-1)+X(t)$$
$$+a_1X(t+1)+\cdots\cdots+a_nX(t+n),$$

and determine the coefficients in such a manner that

$$E\{Y^2_{mn}(t)\}=\text{minimum},$$

so that

(2. 3) $$\begin{pmatrix} a_{-m} \\ \vdots \\ a_{-1} \\ a_1 \\ \vdots \\ a_n \end{pmatrix} = -\begin{pmatrix} 1\cdots\cdots\rho_{m-1} & \rho_{m+1}\cdots\cdots\rho_{m+n} \\ \vdots \qquad \vdots & \vdots \qquad \vdots \\ \rho_{m-1}\cdots\cdots 1 & \rho_2\cdots\cdots\rho_{n+1} \\ \rho_{m+1}\cdots\cdots\rho_2 & 1\cdots\cdots\rho_{n-1} \\ \vdots \qquad \vdots & \vdots \qquad \vdots \\ \rho_{m+n}\cdots\cdots\rho_{n+1} & \rho_{n-1}\cdots\cdots 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_m \\ \vdots \\ \rho_1 \\ \rho_1 \\ \vdots \\ \rho_n \end{pmatrix}$$

where ρ_k is the autocorrelation coefficient of $X(t)$, that is

$$\rho_k=E\{X(t)X(t-k)\}/\sigma_x^2$$

where σ_x^2 is the variance of $X(t)$. If m=n, then obviously $a_{-i}=a_i$ $(i=1,2,\cdots\cdots)$.

The minimized variance of $Y_{mn}(t)$ may be written as

(2. 4) $$\sigma_{Y_{mn}}=\frac{\Delta(m+n)}{\Delta_{m,n}}\sigma_x^2$$

(2. 5) $$=\Big(\sum_{k=-m}^{n} a_k\rho_k\Big)\sigma_x^2, \quad (a_0=1)$$

where

(2. 6) $$\Delta(m+n)=\begin{vmatrix} 1 & \rho_1\cdots\cdots\cdots\rho_{m+n} \\ \rho_1 & 1\cdots\cdots\cdots\rho_{m+n-1} \\ \vdots & \vdots \\ \rho_{m+n} & \rho_{m+n-1}\cdots\cdots 1 \end{vmatrix},$$

(3) E means "Expectation"

(2. 7)
$$\Delta_{m,n}=\begin{vmatrix} 1 \cdots\cdots \rho_{m-1} & \rho_{m+1}\cdots\cdots\rho_{m+n} \\ \vdots \qquad \vdots & \vdots \qquad \vdots \\ \rho_{m-1}\cdots\cdots 1 & \rho_2\cdots\cdots \rho_{n+1} \\ \rho_{m+1}\cdots\cdots\rho_2 & 1\cdots\cdots \rho_{n-1} \\ \vdots \qquad \vdots & \vdots \qquad \vdots \\ \rho_{m+n}\cdots\cdots\rho_{n+1} & \rho_{n-1}\cdots\cdots 1 \end{vmatrix}$$

and where (2.5) is derived from (2.4) by the use of the least square property.

$Y_{mn}(t)$ is uncorrelated with $X(t-k)$, $k=\pm 1, \pm 2, \cdots\cdots, {+m \atop -n}$, and $E\{Y_{mn}(t)\}=0$.[4] Hence, when the values

(2. 8) $$X(t-k)=x_{t-k} \left(k=\pm 1, \pm 2, \cdots\cdots, {+m \atop -n} \right)$$

are given, the value of $X(t)$ at the hour t is not obtained deterministically, but may be a random variable $X_c(t)$ under the condition (2.8);

(2. 9) $$X_c(t)=Y_{mn}(t)-\sum_{k=-m}^{n} a_k x_{t+k} \quad (k \neq 0).$$

We may take as an interpolated value $x^*(t)$ of $X(t)$ the mean value

(2.10) $$x^*(t)=E\{X_c(t)\}=-\sum_{k=-m}^{n}{}' a_k x_{t+k},$$

where $\sum'$ represents the sum for all the values of k except zero. The coefficients a's in (2.10) are given by (2.3), but, in which, for practical purpose, ρ_k are replaced by the serial correlation coefficients r_k.

According to (2.9), the estimated variance s_c^2 of $X_c(t)$ is given by the estimated variance s^2_{ymn} of $Y_{mn}(t)$, and by (2.5)

(2.11) $$s_c^2=s^2_{ymn}=\left(\sum_{k=-m}^{n} a_k r_k\right) s_x^2 \quad (a_0=1),$$

where s_x^2 is the serial variance, i.e. the estimated variance of $X(t)$. This s_c^2 gives the precision of interpolation, that is to say, the probability that the true value of $X(t)$ should be contained in $x^*(t) \pm \delta . s_c$ is approximately given by

(2.12) $$\frac{1}{\sqrt{2\pi}}\int_{-\delta}^{\delta} e^{-\frac{x^2}{2}} dx.$$

When $X(t)$ is a stationary Gaussian process, it is believed that a linear estimation such as (2.10) may have optimum properties.[5]

3. The case of autoregressive process

The variation of a meteorological element $X(t)$ may, in most cases, be assumed to be such that the probabilistic properties is totally determined by the process in the past finite interval of time. Such a process is mathematically de-

(4) While $Y_{mn}(t) \rightarrow Y(t)$ when $m, n \rightarrow \infty$, $Y(t)$ is not non-autocorrelated, and so the theory of i nterpolation is more complicated than in the case of extrapolation.

(5) cf. H. B. Mann and A. Wald : On the Statistical Treatment of Linear Stochastic Difference Equation, Econometrica, 11, No. 3, 4 (1943).

scribed by an autoregressive stochastic process, i. e.

(3. 1) $$X(t)+p_1X(t-1)+\cdots\cdots+p_hX(t-h)=Z(t) \quad (p_h\neq 0),$$

where $Z(t)$ is of non-autocorrelated and is also uncorrelated with $X(t-k)(k=1,2,\cdots\cdots)$, and where h is said to be the order of the process. In this case, the autocorrelation coefficients of $X(t)$ satisfy the following finite difference equations:

(3. 2) $$\rho_k+p_1\rho_{k-1}+\cdots\cdots+p_k\rho_{k-h}=0 \quad (k=1,2,\cdots\cdots).$$

Theorem 1. Let $X(t)$ be an autoregressive process of order h. If $m>h$ and $n>h$, then

(3. 3) $$\sigma^2_{Y_{mn}}=\sigma^2_{Y_{hh}};$$

in other words, the observations at any instants more than h units of time apart need not be taken into consideration.

Proof. Using (3.2) and

(3. 4) $$1+p_1\rho_1+\cdots\cdots+p_h\rho_h=\sigma_z^2/\sigma_x^2=\kappa^2,$$

we get

$$\Delta(m+n)=\Delta(2h)\,(\kappa^2)^{m+n-2h} \qquad (\Delta(2h)\equiv\Delta(h+h)),$$
$$\Delta m,n=\Delta_{h,h}(\kappa^2)^{m+n-2h}$$

which combined with (2.4) gives (3.3).

Theorem 2.

(3. 5) $$\sigma^2_{Y_{hh}}=\sigma_z^2/(1+p_1^2+\cdots\cdots+p_h^2)$$

(3. 5′) $$=\frac{1+p_1\rho_1+\cdots\cdots+p_h\rho_h}{1+p_1^2+\cdots\cdots+p_h^2}\sigma_x^2.$$

In the case of extrapolation of $X(t)$ by $X(t-k)=x_{t-k}$ $(k=1,2,\cdots\cdots)$, we have

$$\sigma^2_{Y_{h0}}=(1+p_1\rho_1+\cdots\cdots+p_h\rho_h)\sigma_x^2,$$

so the precision of interpolation is better than that of extrapolation.

Proof.

(3. 6) $$\Delta(2h)=\Delta(h-1)(\kappa^2)^{h+1}$$

and since $p_h\neq 0$,

$$\rho_k=-\frac{1}{p_h}\rho_{h+k}-\frac{p_1}{p_h}\rho_{h+k-1}-\cdots\cdots-\frac{p_{h-1}}{p_h}\rho_{k+1} \quad (k=h,\ h-1,\cdots\cdots,1,\ -1,\cdots\cdots,-h-1),$$

$$\rho_h-\frac{1}{p_h}\kappa^2=-\frac{1}{p_h}-\frac{p_1}{p_h}\rho_1-\cdots\cdots-\frac{p_{h-1}}{p_h}\rho_{h-1} \quad (\text{for } k=-h).$$

Hence

$$\Delta_{h,h}=p_h^2\Delta(2h-1)+\kappa^2\Delta_{h,h-1}=p_h^2(\kappa^2)^h\Delta(h-1)+\kappa^2\Delta_{h,h-1}.$$

Similarly, if $p_k\neq 0$ $(k=1,2,\cdots\cdots,h-1)$,

$$\Delta_{h,h-1}=p_h^2\Delta(2h-2)+\kappa^2\Delta_{h,h-2}=p^2_{h-1}(\kappa^2)^{k-1}\Delta(h-1)+\kappa^2\Delta_{h,h-2}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\Delta_{h,1}=p_1^2\Delta(h)+\kappa^2\Delta_{h,0}=p_1^2\kappa^2\Delta(h-1)+\kappa^2\Delta(h-1).$$

Therefore

(3. 7) $$\Delta_{h,h}=\Delta(h-1)(\kappa^2)^h(1+p_1^2+\cdots\cdots+p_h^2)$$

From (3.6) and (3.7) we get (3.5).

If $p_n=0$, then $\Delta_{m,n}=\Delta_{m,n-1}\kappa^2$ and (3.5) and (3.5′) are true in this case too.

Corollary 1. If $n\leqq h$, then

$$\sigma^2_{Y h,n}=\sigma_z^2/(1+p_1^2+\cdots\cdots+p_n^2) \tag{3.8}$$

This may be employed in interpolation for any instant within h units of time after the hour of the failure of observation.

Corollary 2. If $0<m<h$, $0<n<h$ and $m+n>h$, then

$$\begin{cases} \sigma^2_{Ymn}=\sigma_z^2\Big/\left\{p^2_{h-m+l}+\cdots\cdots+p^2_{h-m}+\dfrac{\Delta_{n,h-m-1}}{\Delta(h-1)}\kappa^2\right\} \\ \qquad\;=\sigma_z^2\Big/\left\{p^2_{h-n+l}+\cdots\cdots+p^2_{h-n}+\dfrac{\Delta_{n,h-n-1}}{\Delta(h-1)}\kappa^2\right\} \end{cases} \tag{3.9}$$

where $l=m+n-h$.

4. The case of simple Markoff process

In this section, we consider the special case in which $h=1$, but let us assume that the time intervals are not necessarily uniform. The autocorrelation coefficient, in this case, is given by

$$\rho_\tau=\rho^{|\tau|} \qquad (|\rho|<1) \tag{4.1}$$

and the equation corresponding to (2.2) is reduced to

$$a_1X(t-\tau_1)+X(t)+a_2X(t+\tau_2)=Y(t) \qquad (\tau_1,\ \tau_2>0) \tag{4.2}$$

From this, by the method of least squares, we get

$$\begin{cases} a_1+\rho^{\tau_1}+a_2\rho^{\tau_1+\tau_2}=0, \\ a_1\rho^{\tau_1+\tau_2}+\rho^{\tau_2}+a_2=0. \end{cases} \tag{4.3}$$

Hence

$$\begin{cases} a_1=\dfrac{-\rho^{\tau_1}(1-\rho^{2\tau_2})}{1-\rho^{2(\tau_1+\tau_2)}}, \\ a_2=\dfrac{-\rho^{\tau_2}(1-\rho^{2\tau_1})}{1-\rho^{2(\tau_1+\tau_2)}}. \end{cases} \tag{4.4}$$

Besides,

$$(a_1\rho^{\tau_1}+1+a_2\rho^{\tau_2})\sigma_x^2=\sigma_y^2 \tag{4.5}$$

and so

$$\sigma_y^2=\frac{(1-\rho^{2\tau_1})(1-\rho^{2\tau_2})}{1-\rho^{2(\tau_1+\tau_2)}}\sigma_x^2. \tag{4.6}$$

The interpolation formula is, then,

$$x^*(t)=-a_1x(t-\tau_1)-a_2x(t+\tau_2) \tag{4.7}$$

and the precision is given by the estimated value of (4.6);

$$s_y^2=\frac{(1-r^{2\tau_1})(1-r^{2\tau_2})}{1-r^{2(\tau_1+\tau_2)}}s_x^2 \quad (r\equiv r_1). \tag{4.8}$$

Especially, if $\rho\to1\,(r\to1)$, then

$$-a_1\sim\frac{\tau_2}{\tau_1+\tau_2},\quad -a_2\sim\frac{\tau_1}{\tau_1+\tau_2},$$

$$\sigma_y^2\sim O(1-\rho)$$

and (4.7) may be reduced to the formula of proportional division.

5. The case of vector process

Let us consider, for the sake of simplicity, two-dimensional vector process

$$X(t)=\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix}.$$

Corresponding to (2.2), we put, with matrices

$$a^{(k)}=\begin{pmatrix} a_{11}^{(k)} & a_{12}^{(k)} \\ a_{21}^{(k)} & a_{22}^{(k)} \end{pmatrix} \quad (k=1,2,\cdots\cdots),$$

(5. 1) $$a^{(-m)}X(t-m)+\cdots\cdots+a^{(0)}X(t)+\cdots\cdots+a^{(n)}X(t+n)=Y(t),$$

where

(1°) if we want to interpolate only one component of the vector, $X_1(t)$ (the value pertaining to one station), we put

(5. 2) $$a^{(0)}=\begin{pmatrix} 1 & a_{12}^{(0)} \\ a_{21}^{(0)} & 1 \end{pmatrix},$$

especially, in the case of the problem (b) of § 1, we put $m=n=0$;

(2°) if we want to interpolate the whole components of the vector $X(t)$ (pertaining to the whole system of several stations), we put

(5. 3) $$a^{(0)}=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The coefficients $a^{(k)}$'s are to be determined by the method of least squares. Using the estimates of those coefficients, the interpolation formula is given by

(5. 4) $$x^*(t)=(a^{(0)}-\mathrm{I})x(t)-\sum_{k=-m}^{n}{}' a^{(k)}x(t+k),$$

where I stands for the unit matrix. The precision of this interpolation may be given by the (estimated) variance of the component of vector $Y(t)$.

If $X(t)$ is the vector autoregressive process of order h, it satisfies the following equations: (6)

(5. 5) $$p^{(0)}X(t)+p^{(1)}X(t-1)+\cdots\cdots+p^{(h)}X(t-h)=Z(t),$$

(5. 5′) $$q^{(0)}X(t)+q^{(1)}X(t+1)+\cdots\cdots+q^{(h)}X(t+h)=W(t),$$

where p's and q's are matrices (especially $p^{(i)}=q^{(i)}$ for one dimensional case), and

$$p^{(0)}=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad q^{(0)}=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and where $Z(t)$ and $W(t)$ are non-autocorrelated, $Z(t)$ is uncorrelated with $X(t-k)$, and $W(t)$ is uncorrelated with $X(t+k)$, $(k=1,2,\cdots\cdots)$. Besides, the autocor relation matrices

$$\rho_k=\begin{pmatrix} \rho_k^{11} & \rho_k^{12} \\ \rho_k^{21} & \rho_k^{22} \end{pmatrix},$$

(6) M. Ogawara: On the analysis of vector time series, Tōkei Sūri Kenkyū (Bull. Math. Statistics), 3, No. 1 & 2 (1949).

where $\rho_k{}^{ij}=E\{X_i(t)X_j(t-k)\}/\sigma_i\sigma_j$, are such that

$$\rho_k+p^{(1)}\rho_{k-1}+\cdots\cdots+p^{(h)}\rho_{k-h}=0 \quad (k=1,2,\cdots\cdots), \tag{5. 6}$$

$$\rho_k'+q^{(1)}\rho'_{k-1}+\cdots\cdots+q^{(h)}\rho'_{k-h}=0 \quad (k=1,2,\cdots\cdots), \tag{5. 6'}$$

where ρ' is the transposed matrix of ρ. For $k=0$, we have

$$\rho_0+p^{(1)}\rho_{-1}+\cdots\cdots+p^{(h)}\rho_{-h}=u, \tag{5. 7}$$

$$\rho_0'+q^{(1)}\rho'_{-1}+\cdots\cdots+p^{(h)}\rho'_{-h}=v, \tag{5. 7'}$$

where

$$u=\left(\frac{E\{Z_i(t)Z_j(t)\}}{\sigma_i\sigma_j}\right), \quad v=\left(\frac{E\{W_i(t)W_j(t)\}}{\sigma_i\,\sigma_j}\right) \tag{5. 8}$$

σ_i^2 is the variance of $X_i(t)$, $Z_i(t)$ is the ith component of $Z(t)$, and so on.

Theorem 3. The determinant of u is equal to the determinant of v; $\det(u) = \det(v)$.

From this follows:

Theorem 4. In the case (1°), the variance of the component $Y_1(t)$ is given by

$$\sigma_{y_1}{}^2=\frac{\Delta(m+n)}{\Delta'_{m,n}}\sigma_{x_1}{}^2=\frac{\Delta(2h)}{\Delta'_{h,h}}\sigma_{x_1}{}^2(m\leqq h,\ n\leqq h) \tag{5. 9}$$

and in the case (2°),

$$\sigma_{y_1}{}^2+\sigma_{y_2}{}^2=\frac{\Delta^{(1)}(2h)}{\Delta''_{h,h}}\sigma_{x_1}{}^2+\frac{\Delta^{(2)}(2h)}{\Delta''_{h,h}}\sigma_{x_2}{}^2 \quad (m\leqq h,\ n\leqq h), \tag{5.10}$$

where

$$\Delta(m+n)=\begin{vmatrix} \rho_0{}^{11} & \rho_0{}^{12}\cdots\cdots\cdots\rho_{-m}{}^{11} & \rho_{-m}{}^{12}\cdots\cdots\rho^{11}{}_{-(m+n)} & \rho^{12}{}_{-(m+n)} \\ \rho_0{}^{21} & \rho_0{}^{22}\cdots\cdots\cdots\rho_{-m}{}^{21} & \rho_{-m}{}^{22}\cdots\cdots\rho^{21}{}_{-(m+n)} & \rho^{22}{}_{-(m+n)} \\ \vdots & \vdots \qquad\qquad \vdots & \vdots \qquad\qquad \vdots & \vdots \\ \rho_m{}^{11} & \rho_m{}^{12}\cdots\cdots\cdots\rho_0{}^{11} & \rho_0{}^{12}\cdots\cdots\rho_{-n}{}^{11} & \rho_{-n}{}^{12} \\ \rho_m{}^{21} & \rho_m{}^{22}\cdots\cdots\cdots\rho_0{}^{21} & \rho_0{}^{22}\cdots\cdots\rho_{-n}{}^{21} & \rho_{-n}{}^{22} \\ \vdots & \vdots \qquad\qquad \vdots & \vdots \qquad\qquad \vdots & \vdots \\ \rho^{11}{}_{m+n} & \rho^{12}{}_{m+n}\cdots\cdots\rho_n{}^{11} & \rho_n{}^{12}\cdots\cdots\rho_0{}^{11} & \rho_0{}^{12} \\ \rho^{21}{}_{m+n} & \rho^{22}{}_{m+n}\cdots\cdots\rho_n{}^{21} & \rho_n{}^{22}\cdots\cdots\rho_0{}^{21} & \rho_0{}^{22} \end{vmatrix} \tag{5.11}$$

and $\Delta'_{m,n}$ stands for the minor of $\Delta(m+n)$ obtained by deleting the row $\rho_m{}^{11},\cdots\cdots,\rho_{-n}{}^{12}$ and the column $\rho_{-m}{}^{11},\cdots\cdots,\rho_n{}^{21}$, $\Delta^{(1)}(2h)$ is the same as $\Delta'_{h,h}$, $\Delta^{(2)}(2h)$ is the minor of $\Delta(2h)$ obtained by deleting the row $\rho_h{}^{21},\cdots\cdots,\rho_{-h}{}^{22}$ and the column $\rho_{-h}{}^{12},\cdots\cdots,\rho_h{}^{22}$, and lastly $\Delta''_{h,h}$ is the minor of $\Delta(2h)$ obtained by deleting the middle two rows and the middle two columns.

6. Summary

We have derived the formulae ((2.10), (4.7), and (5.4)) to be used for performing stochastic interpolation with a linear approach when one observation was omitted on an occasion, together with the formulae ((3.5), (3.8), (3.9), (4.8), (5.9) and (5.10)) to give the accuracies in various cases.

In cases when several successive observations are omitted, stochastic interpolation of those values may also be treated in a similar manner.

In practical application of our formulae, it is necessary to calculate the para-

meters involved in those formulae (interpolation constants) for each station. The calculation may be laborious, but when once we have prepared them, the interpolation would be easily performed at any time. The example of such calculation will be published some day. Problems of testing statistical hypothesis concerning the interpolation constants are left in future.