Time Series Analysis and Stochastic Prediction (II)

Ogawara, Masami
Meteorological Research Institute

https://doi.org/10.5109/12984

出版情報：統計数理研究. 8 (3/4), pp.55-72, 1959-03. Research Association of Statistical Sciences
バージョン：
権利関係：

TIME SERIES ANALYSIS AND STOCHASTIC PREDICTION (II)

By

Masami OGAWARA

(*Received* September 1, 1958)

Contents

Chapter IV. Stochastic Prediction

§ **4.1. General principles.** Let $x_t\ (t \in T')$ be a time series, that is a realization of a stochastic process $x(t)\ (t \in T)$, on $T' \subset T$, and let $y = \varphi_s(x(\cdot))$ be a measurable (P) functional of $x(t)\ (t \in S \subset T - T')$ which may take values in Euclidian space R^k.

If the probability law of the $x(t)$ process is perfectly known to us, the prediction of the y by the observed sample function $x_t(t \in T')$ may be defined by its conditional probability distribution given $x_t(t \in T')$.

Even when the probability law is not known, if the process is stationary and ergodic and if the sample function x_t is given on $(-\infty, t_0)$, it is well-known that the mean value and the covariance function of the process are almost surely known and we can get the linear least square prediction of $x(t_0+s)\ (s>0)$ and that for a normal process the linear least square prediction and its prediction error is nothing but the conditional mean value and the conditional variance of the $x(t_0+s)$ given $x(t)=x_t\ (-\infty < t \leqq t_0)$ respectively. (Wiener [31], Doob [26])

In most practical cases, however, even if the type of the probability law of the process is approximately known, the parameters specifing the probability law, including the spectral distribution or the autocovariance function, are all unknown, and moreover we are only given a time series on a limited range $T' \subset T$, consequently we cannot rigorously apply the current prediction theories to it.

As an exact prediction method for such time series we have introduced an idea of stochastic prediction (Ogawara [10, 11]), which is considered to be an extension of R. A. Fisher's concept of fiducial prediction (Fisher [32]) to a time series and which is, so to speak, a conditional fiducial prediction. In the following we assume that the type of the probability law defining a stochastic process with a finite parameter scheme is known but all of the parameters are unknown.

Definition. *Suppose that a* k' ($\leq k$) *dimensional function of random variables* $y=\varphi_s(x(\cdot))$ $(S\subset T-T')$ *and* $x=(x_t, t\in T_1\subset T')$ *and fixed variates* $x'=(x_t, t\in T_2\subset T')$ $(T_1\cap T_2=0)$,

$$u=u(y, x; x') \tag{4.4.1}$$

is measurable (*P*) *and has a probability distribution independent of unknown parameters. If we then fix the* x *at the observed value,* y *is a* (*many-valued*) *function of random variable* u, *where we assume that the* u *distribution is unchanged by fixing* x. *Suppose further that, when* $k'<k$, *for a fixed* x, *the* y *has constant distribution densities on the subspace of* y *values such that* $u(y, x; x')=$ *constant, then a distribution of* y *is derived from the* u *distribution. Thus obtained* (*formal*) *distribution of* y *we shall call the stochastic prediction of* y.

In most cases of stochastic prediction for a normal process, as we shall see later on, we can find one dimensional statistic $u(y, x; x')$ which has a known distribution independent of the population parameters and

$$u(y, x; x')=u \quad \text{(constant)} \tag{4.1.2}$$

is a hyperellipsoidal surface of y from which we get

$$\Pr(y\in W(x, x'; p)\,|\,x')=p, \quad 0\leq p\leq 1, \tag{4.1.3}$$

where the $W(x, x'; p)$ is a k dimensional random ellipsoid. If we fix the x at the observed value, (4.1.3) gives the stochastic prediction of y, where it is assumed, by the definition, that the densities of y is constant on the ellipsoidal surface. The meaning of the 'formal' distribution of y may be explained, in this case, as follows. Suppose that we repeat the prediction infinitely many times for the same predictand y saying that $y\in W(x, x'; p)$ with probability p, then $100\,p\%$ of them will be true.

The following theorem about two samples with a multivariate regression is fundamental in constructing stochastic prediction for a normal process.

Theorem 17. (Ogawara [11]) *Let* $(y_{\alpha i\nu}-\sum_{p=1}^{r} a_{ip}x_{\alpha p\nu}; i=1, 2, \cdots, k)$ $(\nu=1, 2, \cdots, n_\alpha)$, $\alpha=1, 2$ *be two independent samples of size* n_1 *and* n_2 *respectively drawn from the same* k *dimensional normal population with* *zero mean, where* $(x_{\alpha p\nu};\ p=1, 2, \cdots, r, \nu=1, 2, \cdots, n_\alpha)$ *is a set of fixed variates corresponding to the* α*-th sample* $(\alpha=1, 2)$, $n_1-k-r>0$, *and let us define the following quantities.*

$$\left.\begin{aligned}
&c_{\alpha pq} \equiv \sum_{\nu=1}^{n_\alpha} x_{\alpha p\nu} x_{\alpha q\nu} \qquad (p, q=1, 2, \cdots, r;\ \alpha=1, 2)\\
&c'_{1iq} \equiv \sum_{\nu=1}^{n_1} y_{1i\nu} x_{1q\nu}, \quad \sum_{p=1}^{k} \hat{a}_{1ip} c_{1pq} = c'_{1iq} \qquad \begin{pmatrix} q=1, 2, \cdots, r\\ i=1, 2, \cdots, k\end{pmatrix}\\
&c_{pq} \equiv c_{1pq}+c_{2pq}, \qquad (c^{pq}) \equiv (c_{pq})^{-1}\\
&a_{1ij} \equiv \sum_{\nu=1}^{n_1}(y_{1i\nu}-\sum_{p=1}^{r} \hat{a}_{1ip} x_{1p\nu})\,(y_{1j\nu}-\sum_{q=1}^{r} \hat{a}_{1jq} x_{1q\nu})\\
&\qquad\qquad\qquad\qquad\qquad (i, j=1, 2, \cdots, k)\\
&(a_1^{ij}) \equiv (a_{1ij})^{-1}.
\end{aligned}\right\} \qquad (4.1.4)$$

If we then set

$$\begin{aligned}
Y \equiv \sum_{\nu=1}^{n_2}\Bigg\{&\Big[1-\sum_{p,q=1}^{r} c^{pq} x_{2p\nu} x_{2q\nu}\Big]\\
&\cdot\Big[\sum_{i,j=1}^{k} a_1^{ij}\Big(y_{2i\nu}-\sum_{p=1}^{r} \hat{a}_{1ip} x_{2p\nu}\Big)\Big(y_{2j\nu}-\sum_{q=1}^{r} \hat{a}_{1jq} x_{2q\nu}\Big)\Big]\Bigg\}, \qquad (4.1.5)
\end{aligned}$$

the h-th moment of a random variable $U=1/(1+Y)$ *is given by*

$$E\{U^h\} = \frac{\prod_{i=1}^{k} \Gamma\left(\frac{n_1+n_2-r+1-i}{2}\right)\Gamma\left(\frac{n_1-r+1-i}{2}+h\right)}{\prod_{i=1}^{k} \Gamma\left(\frac{n_1+n_2-r+1-i}{2}+h\right)\Gamma\left(\frac{n_1-r+1-i}{2}\right)} \qquad (h=1, 2, \cdots)\,. \qquad (4.1.6)$$

Proof. Adding to (4.1.4), we set

$$\begin{aligned}
&\left.\begin{aligned}
&c'_{2iq} \equiv \sum_{\nu=1}^{n_2} y_{2i\nu} x_{2q\nu}\\
&c'_{iq} \equiv c'_{1iq}+c'_{2iq}, \quad \sum_{p=1}^{k} \hat{a}_{ip} c_{pq} = c'_{iq}
\end{aligned}\right\} \begin{pmatrix} q=1, 2, \cdots, r\\ i=1, 2, \cdots, k\end{pmatrix}\\
&\bar{a}_{\alpha ij} \equiv \sum_{\nu=1}^{n_\alpha}\Big(y_{\alpha i\nu}-\sum_{p=1}^{r} \hat{a}_{ip} x_{\alpha p\nu}\Big)\Big(y_{\alpha j\nu}-\sum_{q=1}^{r} \hat{a}_{jq} x_{\alpha q\nu}\Big) \qquad \begin{pmatrix} i, j=1, 2, \cdots, k\\ \alpha\ \ =1, 2\end{pmatrix}\\
&\bar{a}_{ij} \equiv \bar{a}_{1ij}+\bar{a}_{2ij}.
\end{aligned}$$

Then, since

$$\hat{a}_{ip}-\hat{a}_{1ip} = \sum_{s=1}^{r} c^{ps} \sum_{\nu=1}^{n_2}\Big(y_{2i\nu}-\sum_{q=1}^{r} a_{1iq} x_{2q\nu}\Big) x_{2s\nu},$$

we have

$$\bar{a}_{ij} = a_{1ij} + T_{ij},$$

where

$$T_{ij} \equiv \sum_{\nu=1}^{n_2}\Big[1-\sum_{p,q=1}^{r} c^{pq} x_{2p\nu} x_{2q\nu}\Big] \cdot \Big(y_{2i\nu}-\sum_{p=1}^{r} \hat{a}_{1ip} x_{2p\nu}\Big)\Big(y_{2j\nu}-\sum_{q=1}^{r} \hat{a}_{1jq} x_{2q\nu}\Big),$$

and

$$|\bar{a}_{ij}| = |a_{1ij}| \cdot \Big[1+\sum_{i,j=1}^{k} a_1^{ij} T_{ij}\Big] = |a_{1ij}| \cdot [1+Y].$$

Now, since $n_1-r-k>0$, the a_{1ij} and the $\bar{a}_{ij}$ are distributed according to Wishert distribution

$$W_{n_1-r,k}(a_{1ij}\,;\,\sigma^{ij}) \quad \text{and} \quad W_{n_1+n_2-r,k}(\bar{a}_{ij}\,;\,\sigma^{ij})$$

respectively, where

$$W_{m,k}(b_{ij}\,;\,\sigma^{ij}) = \frac{(|\sigma^{ij}|/2^k)^{m/2}}{\pi^{k(k-1)/4}} \frac{|b_{ij}|^{(m-k-1)/2}}{\prod_{i=1}^{k} \Gamma\left(\frac{m+1-i}{2}\right)} e^{-\frac{1}{2}\sum_{i,j=1}^{k}\sigma^{ij} b_{ij}} \prod_{i\leq j=1}^{k} db_{ij},$$

and the T_{ij} are quadratic forms of kn_2 random variables which are distributed according to kn_2 dimensional normal distribution and, since $y_{1i\nu}$ and $y_{2j\nu}$ are independent and the a_{1ij} are independent of $\hat{a}_{1ip}$, the T_{ij} are independent of a_{1ij}. (Wilks [33]) Therefore, by Wilks' technique (Wilks [34]), we can find the moments of $U=|a_{1ij}|/|\bar{a}_{ij}|$, getting (4.1.6).

This theorem involves important special cases. For instance, when $n_1 = n$, $n_2=1$, we have

$$|c_{pq}| = |c_{1pq}|\Big[1+\sum_{p,q=1}^{r} c_1^{pq} x_{2p1} x_{2q1}\Big], \qquad ((c_1^{pq}) = (c_{1pq})^{-1})$$

and

$$\sum_{p,q=1}^{r} c_1^{pq} x_{2p1} x_{2q1} = \sum_{p,q=1}^{r} c_1^{pq} x_{2p1} x_{2q1} \Big/ \Big(1+\sum_{p,q=1}^{r} c_1^{pq} x_{2p1} x_{2q1}\Big).$$

Hence, if we rewrite

$$c_{1pq} \equiv c_{pq}, \qquad (c_{1pq})^{-1} \equiv (c^{pq}), \qquad a_1^{ij} \equiv a^{ij}$$

$$x_{2p1} \equiv x_{p0}, \qquad y_{2i1} \equiv y_{i0} \quad \text{and} \quad \hat{a}_{1ip} \equiv \hat{a}_{ip}$$

in (4.1.5), we observe that

$$F^k_{n+1-r-k} = \frac{\sum_{i,j=1}^{k} a^{ij}\Big(y_{i0}-\sum_{p=1}^{r}\hat{a}_{ip}x_{p0}\Big)\Big(y_{j0}-\sum_{q=1}^{r}\hat{a}_{jq}x_{q0}\Big)}{1+\sum_{p,q=1}^{r} c^{pq} x_{p0} x_{q0}} \cdot \frac{n+1-r-k}{k} \tag{4.1.7}$$

is distributed according to F distribution with k and $n+1-r-k$ degrees of freedom. Moreover, if $k=1$ and if we write $y_{i0} \equiv y_0$, $y_{1i\nu} \equiv y_\nu$, $x_{1p\nu} \equiv x_{p\nu}$ and $\hat{a}_{ip} \equiv \hat{a}_p$, we get

$$F^1_{n-r} = \frac{\Big(y_0-\sum_{p=1}^{r}\hat{a}_p x_{p0}\Big)^2 (n-r)}{\sum_{\nu=1}^{n}\Big(y_\nu-\sum_{p=1}^{r}\hat{a}_p x_{p\nu}\Big)^2\Big(1+\sum_{p,q=1}^{r} c^{pq} x_{p0} x_{q0}\Big)} \tag{4.1.8}$$

$$= \frac{\left(y_0 - \sum_{p=1}^{r} \hat{a}_p x_{p0}\right)^2 (n-r)}{\left(c'_0 - \sum_{p=1}^{r} \hat{a}_p c'_p\right)\left(1 + \sum_{p,q=1}^{r} c^{pq} x_{p0} x_{q0}\right)}, \tag{4.1.8'}$$

where $c'_0 = \sum_{\nu=1}^{n} y_\nu^2$ and $c'_p = \sum_{\nu=1}^{n} y_\nu x_{p\nu}$.

For a stochastic prediction of future value, we use exclusively *the second sample schem* of time series (cf. § 1.1) and the cast of variables in the formulas (4.1.5), (4.1.7) or (4.1.8) for a time series will be given in the following paragraphs. Some remarks on the practice of stochastic prediction are given in other papers. (Ogawara [5], [8~11], [13~15], [17], [20], [21])

It may be noticed that the idea in this paragraph is also applied to stochastic interpolation of time series, where *the first sample scheme* is adopted. Moreover, our method is applied to conditional rejection tests of W. R. Thompson's type for a time series which may have a trend of the type (3.5.3), where either of the two samle schemes is used according to practical circumstances. (Ogawara [18], [26])

In the following paragraphs of this chapter we shall concern exclusively with the normal process, however if a non-normal stochastic process $\xi(t)$ can be transformed into a normal $x(t)$ process by a topological mapping $x(t) = g(\xi(t))$, the stochastic prediction of $x(t)$ is inversely transformed into that of $\xi(t)$.

§ 4.2. **Autoregression process.*** Let $x(t)$ $(t = 0, \pm 1, \pm 2, \cdots)$ be a normal autoregression process of order h, and let $x_{t-\tau}$ $(\tau = 0, 1, \cdots, L)$ be an observed time series, t being the present time.

(a) *One dimensional case.* For the prediction of $x(t+s)$ $(s>0)$, we set

$$\left.\begin{aligned}
y_\nu &\equiv x(t-\nu(h+s)+s), \quad \nu = 0, 1, \cdots, N \\
\bar{y} &= \sum_{\nu=1}^{N} y_\nu / N, \quad \text{where} \quad N(h+s)+h-1 \leq L \\
x'_{p\nu} &\equiv x(t-\nu(h+s)-p+1) - \bar{x}_p, \quad p = 1, 2, \cdots, h;\ \nu = 0, 1, \cdots, N \\
\bar{x}_p &= \sum_{\nu=1}^{N} x(t-\nu(h+s)-p+1)/N \\
c_{pq} &\equiv \sum_{\nu=1}^{N} x'_{p\nu} x'_{q\nu}, \quad c_{p0} \equiv \sum_{\nu=1}^{N} x'_{p\nu} y_\nu \\
&\sum_{q=1}^{h} c_{pq} \hat{c}_q = c_{p0} \quad (p = 1, 2, \cdots, h), \quad (c_{pq}) = (c_{pq})^{-1},
\end{aligned}\right\} \tag{4.2.1}$$

where $x'_{p\nu}$ is adopted instead of $x_{p\nu} = x(t-\nu(h+s)-p+1)$ in (4.1.8). Then the random variables y_ν $(\nu = 0, 1, \cdots, N)$ are independent of each other

* This and the next paragraph are the outline of and the supplement to Ogawara [11].

when the variates $x'_{p\nu}$ are fixed (Theorem 2), where $x(t-\nu(h+s)+u)$ $(u=1, 2, \cdots, s-1;\ \nu=0, 1, \cdots, N)$ are *free variates*. Thus, by (4.1.8) and the definition, the transformed t distribution of

$$y=\bar{y}+\sum_{p=1}^{h}\hat{c}_p x'_{p0}$$
$$+t_{N-h-1}\left[\frac{1}{N-h-1}\sum_{\nu=1}^{N}\left(y_\nu-\bar{y}-\sum_{p=1}^{h}\hat{c}_p x'_{p\nu}\right)\cdot\left(1+\frac{1}{N}+\sum_{p,q=1}^{h}c^{pq}x'_{p0}x'_{q0}\right)\right]^{1/2} \tag{4.2.2}$$

gives the stochastic prediction for $x(t+s)$, where t_{N-h-1} is the random variable distributed according to t distribution with $N-h-1$ degrees of freedom and the other variates are all fixed. We can also predict the $x(t+s)$ from $x_{t-s'-\tau}$ $(\tau=0, 1, \cdots;\ s'>0)$.

Next, for the joint prediction of $(x(t+s), x(t+s+1), \cdots, x(t+s+k-1))$ $(k, s>0)$, we set

$$\left.\begin{array}{ll}
y_{1i\nu}\equiv x(t-\nu(h+s)+s+i-1) & i=1, 2, \cdots, k;\ \nu=1, 2, \cdots, N\\
y_{i0}\equiv x(t+s+i-1) & i=1, 2, \cdots, k\\
x_{1p\nu}\equiv\begin{cases}1 & p=1\\ x(t-\nu(h+s)-p+2) & p=2, 3, \cdots, h+1\end{cases} & \nu=1, 2, \cdots, N\\
x_{p0}\equiv\begin{cases}1 & p=1\\ x(t-p+2) & p=2, 3, \cdots, h+1,\end{cases} &
\end{array}\right\} \tag{4.2.3}$$

where $y_{1i\nu}$ and y_{i0} are random variables, $x_{1p\nu}$ and x_{p0} are fixed variates and the other vaiiables are free. Then, from (4.1.7), putting $r=h+1$, we get

$$(N-h-k)\sum_{i,j=1}^{k}a^{ij}\left(y_{i0}-\sum_{p=1}^{h+1}\hat{a}_{ip}x_{p0}\right)\left(y_{j0}-\sum_{q=1}^{h+1}\hat{a}_{jq}x_{q0}\right)$$
$$=kF^{k}_{N-h-k}(P)\left(1+\sum_{p,q=1}^{h+1}c^{pq}x_{p0}x_{q0}\right)\qquad(0<P<1) \tag{4.2.4}$$

which defines a stochastic prediction, where $F^{k}_{N-h-k}(P)$ denotes the $100\,P\%$ point of F distribution; the probability that the k dimensional random point $(x(t+s), x(t+s+1), \cdots, x(t+s+k-1))$ falls in the random hyperellipsoid (4.2.4) is $1-P$, when the values of $x_{1p\nu}$ and x_{p0} are given. The distribution defined by (4.2.4) includes that of (4.2.2) as a special case where $k=1$.

(b) *k dimensional case.* We get the same formula as (4.2.4) for the prediction of $x(t+s)=(x_1(t+s), x_2(t+s), \cdots, x_k(t+s))$, if we only interpret the notations as follows:

$$\left.\begin{array}{ll}
y_{1i\nu}\equiv x_i(t-\nu(h+s)), & \nu=1, 2, \cdots, N\\
y_{i0}\equiv x_i(t+s) &
\end{array}\right\}\ i=1, 2, \cdots, k \tag{4.2.5}$$

and assigning a through number $p=1, 2, \cdots, k(h+1)$ to the pairs (i, p') $(i=1, 2, \cdots, k;\ p'=0, 1, \cdots, h)$, we set

$$\left.\begin{aligned} x_{1p\nu} &\equiv \begin{cases} 1 & \begin{pmatrix} p'=0 \\ i=1, 2, \cdots, k \end{pmatrix} \quad p=1, 2, \cdots, k \\ x_i(t-\nu(h+s)-p'+1) & \begin{pmatrix} p'=1, 2, \cdots, h \\ i=1, 2, \cdots, k \end{pmatrix} \\ & \quad p=k+1, \cdots, k(h+1) \end{cases} \quad \nu=1, 2, \cdots, N \\ x_{p0} &\equiv \begin{cases} 1 & \begin{pmatrix} p'=0 \\ i=1, 2, \cdots, k \end{pmatrix} \quad p=1, 2, \cdots, k \\ x_i(t-p'+1) & \begin{pmatrix} p'=1, 2, \cdots, h \\ i=1, 2, \cdots, k \end{pmatrix} \quad p=k+1, \cdots, k(h+1) \end{cases} \end{aligned}\right\} \tag{4.2.6}$$

If $x(t)$ is an $n\times n'$ matrix, we can assign a number to each element of the $x(t)$ from 1 to $k=nn'$, thus we may suppose that the $x(t)$ is a k dimensional vector process and apply the above prediction formula. For example, this is the case where we would predict the pattern of a meteorological element represented by the values for a grid of intersections of latitude and longitude.

If we want to get a joint stochastic prediction of $(x(t+s), x(t+s+1), \cdots, x(t+s+l-1))$ for a k dimensional autoregression process $x(t)=(x_1(t), x_2(t), \cdots, x_k(t))$ of order h, the following interpretation of the formula (4.2.4) will do.

$$\left.\begin{aligned} & y_{1i\nu} \equiv x_j(t-\nu(h+s+l-1)+s+u-1), \quad \nu=1, 2, \cdots, N \\ & y_{i0} \equiv x_j(t+s+u-1) \\ & \qquad \begin{pmatrix} u=1, 2, \cdots, l \\ j=1, 2, \cdots, k \end{pmatrix} \quad i=1, 2, \cdots, kl \\ & x_{1p\nu} \equiv \begin{cases} 1 & \begin{pmatrix} p'=0 \\ j=1, 2, \cdots, k \end{pmatrix} \quad p=1, 2, \cdots, k \\ x_j(t-\nu(h+s+l-1)-p'+1) & \begin{pmatrix} p'=1, 2, \cdots, h \\ j=1, 2, \cdots, k \end{pmatrix} \\ & \quad p=k+1, \cdots, k(h+1) \end{cases} \quad \nu=1, 2, \cdots, N \\ & x_{p0} \equiv \begin{cases} 1 & \begin{pmatrix} p'=0 \\ j=1, 2, \cdots, k \end{pmatrix} \quad p=1, 2, \cdots, k \\ x_j(t-p'+1) & \begin{pmatrix} p'=1, 2, \cdots, h \\ j=1, 2, \cdots, k \end{pmatrix} \quad p=k+1, \cdots, k(h+1). \end{cases} \end{aligned}\right\} \tag{4.2.7}$$

The formula of the form (4.2.4) will be useful for more general cases such that $y_{1i\nu}$ and y_{i0} are certain functions (of the same form) of $x_j(t-\nu(h+s+l-1)+s+u-1)$ and $x_j(t+s+u-1)$ $(u=1, 2, \cdots, l;\ j=1, 2, \cdots, k)$ respectively.

(c) *Autoregression process with trend.* If the trend of mean have the form (3.5.3), the process $x(t)$ follows a linear model which can be written, by Theorem 15, in the form

$$\begin{aligned} x(t+s)+b_1x(t)+\cdots+b_hx(t-h+1)\\ -\beta_1v_1(t)-\cdots-\beta_rv_r(t)=z(t) \end{aligned} \tag{4.2.8}$$

Thus the stochastic prediction of $x(t+s)$ in one dimensional case, for instance, can be given by (4.2.2) if only we add to (4.2.1)

$$x'_{h+j,\nu}\equiv v_j(t-\nu(h+s))-\bar{v}_j,\quad j=1,2,\cdots,r;\ \nu=0,1,\cdots,N, \tag{4.2.9}$$

where $\bar{v}_j=\sum_{\nu=1}^{N}v_j(t-\nu(h+s))/N$.

More general stochastic prediction for the autoregression process with a trend of the form (3.5.3) will be carried out by the similar modifications.

Now, let us consider the limit case when $N\to\infty$. Since a stationary normal autoregression process is metrically transitive, each side of (4.2.4) converges to the each side of

$$\sum_{i,j=1}^{k}\sigma^{ij}\Big(y_{i0}-\sum_{p=1}^{h+1}\alpha_{ip}x_{p0}\Big)\Big(y_{j0}-\sum_{q=1}^{h+1}\alpha_{jq}x_{q0}\Big)=\chi_k^2(P) \tag{4.2.10}$$

with probability 1 as $N\to\infty$ respectively, where the α_{ip} are regression coefficients of y_{i0} on x_{p0}, (σ^{ij}) the inverse matrix of the covariance matrix of $y_{i0}-\sum\alpha_{ip}x_{p0}$ $(i=1,2,\cdots,k)$ and $\chi_k^2(P)$ the $100\,P\%$ point of χ^2 distribution with k degrees of freedom. Therefore, owing to the definition of stochastic prediction, the left hand side of (4.2.10) is the exponent of the conditional k dimensional normal distribution given x_{p0} $(p=1,2,\cdots,h+1)$ and

$$\int_D\frac{\sqrt{|\sigma^{ij}|}}{(2\pi)^{k/2}}\exp\Big[-\frac{1}{2}\sum_{i,j=1}^{k}\sigma^{ij}(y_{i0}-\sum\alpha_{ip}x_{p0})(y_{j0}-\sum\alpha_{jq}x_{q0})\Big]dy_{10}\cdots dy_{k0}=P, \tag{4.2.11}$$

where D is the outside of the hyperellipsoid (4.2.10). Thus we have

Theorem 18. *A stochastic prediction defined by* (4.2.4) *for a stationary normal autoregression process converges to the conditional normal distribution of* y_{i0} $(i=1,2,\cdots,k)$ *given* x_{p0} *with probability* 1 *as* N *tends to infinity.*

Accordingly, if we let $N\to\infty$ and then $h\to\infty$ in (4.2.4), we get Wiener's linear least square prediction for a normal linear process. (Wiener [31], Doob [26])

§ **4.3. Moving averages.** Let $x(t)$ $(t=0,\pm1,\cdots)$ be a moving averages of order h and let $x_{t-\tau}$ $(\tau=0,1,\cdots,L)$ be a time series observed in the past.

In one dimensional case, the method of stochastic prediction for $x(t+s)$ $(s>0)$ is as follows.

(1) When $s \leq h$, we put

$$\left.\begin{aligned} y_\nu &\equiv x(t-\nu(l+h+s)+s) \\ x'_{p\nu} &\equiv x(t-\nu(l+h+s)-p+1)-\bar{x}_p, \\ & \qquad\qquad p=1,2,\cdots,l \end{aligned}\right\} \nu=0,1,\cdots,N \quad (4.3.1)$$

$$\bar{y} = \sum_{\nu=1}^{N} y_\nu/N, \quad \bar{x}_p = \sum_{\nu=1}^{N} x(t-\nu(l+h+s)-p+1)/N,$$

then the $N+1$ sets of $l+1$ random variables

$$\{y_\nu, x(t-\nu(l+h+s)-p+1) \ (p=1,2,\cdots,l)\} \qquad \nu=0,1,\cdots,N$$

are independent and the prediction formula of $x(t+s)$ is given by (4.2.2), provided that we write l instead of h. Since $L=N(l+h+s)+l-1$ is given, if we take l large, N becomes small. Therefore we should select the value of l (≥ 1) so that the length of the interval

$$\bar{y}+\sum_{p=1}^{l} c_p x'_{p0}$$
$$\pm\left[F^1_{N-l-1}(P)\frac{1}{N-l-1}\sum_{\nu=1}^{N}\Big(y_\nu-\bar{y}-\sum_{p=1}^{l}\hat{c}_p x'_{p\nu}\Big)^2\cdot\Big(1+\frac{1}{N}+\sum_{p,q=1}^{l}c^{pq}x'_{p0}x'_{q0}\Big)\right]^{1/2}$$

is minimum for a P in which we are most interested. If it is longer than the length of the interval

$$\bar{y}\pm\left[F^1_{N-1}(P)\frac{N+1}{N-1}\cdot\frac{1}{N}\sum_{\nu=1}^{N}(y_\nu-\bar{y})^2\right]^{1/2},$$

which corresponds to (4.3.2) in the following section, we should rather choose the unconditional stochastic prediction given below.

(2) When $s>h$, we cannot do any effective conditional stochastic prediction. In this case, however, $y_\nu\equiv x(t-\nu(h+1))$ ($\nu=0,1,\cdots,N-1$) are independent of each other and the distribution of

$$y=\bar{y}+t_{N-1}s_y\sqrt{(N+1)/(N-1)} \qquad (4.3.2)$$

may be called an *unconditional stochastic prediction* of $x(t+s)$ for an arbitrary $s>0$, where $\bar{y}=\sum y_\nu/N$, $s_y^2=\sum(y_\nu-y)^2/N$ and t_{N-1} is a random variable distributed according to t distribution with $N-1$ degrees of freedom and the other variates are fixed. The distribution of (4.3.2), which is a transformed t distribution, corresponds to an estimation of the absolute probability distribution function of $x(t)$.

The prediction formulas for the joint prediction of

$$x(t+s), x(t+s+1), \cdots, x(t+s+r-1) \qquad (s>0, r>1)$$

and for the prediction of

$$x(t+s)=(x_1(t+s), x_2(t+s), \cdots, x_k(t+s))$$

or

$$\begin{pmatrix} x(t+s) \\ x(t+s+1) \\ \vdots \\ x(t+s+l-1) \end{pmatrix}$$

$$= \begin{pmatrix} x_1(t+s), & x_2(t+s), & \cdots, & x_k(t+s) \\ x_1(t+s+1), & x_2(t+s+1), & \cdots, & x_k(t+s+1) \\ \cdots & \cdots & \cdots & \cdots \\ x_1(t+s+l-1), & x_2(t+s+l-1), & \cdots, & x_k(t+s+l-1) \end{pmatrix}$$

in the case of k dimensional process, and moreover for the prediction in the case of moving averages with trend are similar to those for an autoregression process, except for the distinction of two cases $s \leqq h$ and $s > h$, so we may cut off them here.

We have also a limit theorem quite similar to Theorem 18.

Theorem 18′. *As $N \to \infty$, a stochastic prediction with* (4.3.1) *for a stationary normal moving averages converges to the conditional probability distribution of the predictand given $x(t-\tau) = x_{t-\tau}$ ($\tau = 0, 1, \cdots, l-1$) with probability* 1. *If we then let $l \to \infty$, the conditional mean value and the conditional variance tends to the linear least square prediction and the prediction error in Wiener's sense respectively.*

§ 4.4. **Some continuous parameter processes.** We consider the process of the type (2.3.10) satisfing (2.3.11),

$$(x(t) - m) + a_1 x'(t) + \cdots + a_h x^{(h)}(t) = z'(t) \quad \text{(formally)}, \qquad (4.4.1)$$

where $z(t)$ is a Brownian motion on $(-\infty, \infty)$. Let $x_{t-\tau}$ $(0 \leqq \tau \leqq T)$ be a sample function. The method of stochastic prediction for a measurable (P) functional $y_0 \equiv \varphi_{(t+s,\, t+s+u)}(x(\cdot))$ $(s, u > 0)$ of such process is also reduced to that given in § 4.2. Namely, let

$$t_\nu \equiv t - \nu(s+u), \quad \nu = 0, 1, \cdots, N \quad (t_N \geqq t - T) \qquad (4.4.2)$$

and let

$$\left.\begin{aligned} y_\nu &\equiv \varphi_{(t_\nu+s,\, t_\nu-1)}(x(\cdot)) \\ x'_{p\nu} &\equiv x_{t_\nu}^{(p-1)} - \sum_{\nu=1}^{N} x_{t_\nu}^{(p-1)}/N, \quad p = 1, 2, \cdots, h \end{aligned}\right\} \nu = 0, 1, \cdots, N, \qquad (4.4.3)$$

then, by Theorem 6 and Corollary to it, the conditional random variables y_ν given $x'_{p\nu}$ $(p = 1, 2, \cdots, h)$ are independent and the stochastic prediction formula (4.4.2) is applied.

Particularly, if we put

$$y_\nu \equiv x_{t-\nu(s+u)+s}\,, \quad \bar{y} = \sum_{\nu=1}^{N} y_\nu/N\,, \quad x'_{p\nu} \equiv x^{(p-1)}_{t-\nu(s+u)}\,, \quad \bar{x}_p = \sum_{\nu=1}^{N} x'_{p\nu}/N\,,$$

the prediction formula for $y = x(t+s)$ becomes

$$y = \bar{y} + \sum_{p=1}^{h} \hat{c}_p (x'_{p0} - \bar{x}_p) + t_{N-h-1}\left[\frac{1}{N-h-1}\sum_{\nu=1}^{N}\left(y_\nu - \bar{y} - \sum_{p=1}^{h}\hat{c}_p(x'_{p\nu} - \bar{x}_p)\right)^2 \cdot \left(1 + \frac{1}{N} + \sum_{p,q=1}^{h} c^{pq}(x'_{p0} - \bar{x}_p)(x'_{q0} - \bar{x}_q)\right)\right]^{1/2}, \qquad (4.4.4)$$

where

$$c_{pq} = \sum_{\nu=1}^{N}(x'_{p\nu} - \bar{x}_p)(x'_{q\nu} - \bar{x}_q)\,, \qquad c_{p0} = \sum_{\nu=1}^{N}(x'_{p\nu} - \bar{x}_p)(y_\nu - y)$$

$$\sum_{q=1}^{h} c_{pq}\hat{c}_q = c_{p0} \quad (p = 1, 2, \cdots, h)\,, \qquad (c^{pq}) = c_{pq})^{-1}$$

and t_{N-h-1} is a random variable distributed according to t distribution with $N-h-1$ degrees of freedom and the other variates are fixed.

Now, since

$$\rho^{(\nu)}(\tau) = \int_{-\infty}^{\infty} e^{i\tau\lambda}(i\lambda)^\nu F'(\lambda)\,d\lambda \qquad (\nu = 0, 1, \cdots, 2h-2)$$

and, according to (2.3.3),

$$Ex^{(\nu)}(t)\,x^{(\nu')}(t-\tau) = (-1)^{\nu'}\rho^{(\nu+\nu')}(\tau) \qquad (\tau > 0)\,,$$

the spectral distributions of normal stationary processes $x^{(p-1)}(t)$ are absolutely continuous and the discrete parameter processes $x^{(p-1)}(t - \nu(s+u))$ $(\nu = 0, 1, \cdots)$ are metrically transitive $(p = 1, 2, \cdots, h)$. Therefore, as $N \to \infty$, (4.4.4) tends to

$$m + c_1(x_t - m) + \sum_{p=2}^{h} c_p x_t^{(p-1)} + \xi\sigma\left[1 - \sum_{p=1}^{h}(-1)^{p-1}c_p\rho^{(p-1)}(s+u)\right]^{1/2} \qquad (4.4.5)$$

(in the sense of convergence in law), where the c_p are the same as the coefficients in (3.2.28) provided that $t_0 = t$, $t - t_0 = s + u$, and ξ is a random variable distributed according to the standard normal distribution $N(0, 1)$. Thus we have

Theorem 18″. *The stochastic prediction given by* (4.4.4) *converges to the conditional normal distribution of* $x(t+s)$ *given* $x^{(p-1)}(t) = x_t^{(p-1)}$ $(p=1, 2, \cdots, h)$, *with probability* 1, *as* $N \to \infty$, *the conditional mean value of which is the linear least square prediction of* $x(t+s)$ *given* $x_{t-\tau}$ $(0 \leqq \tau < \infty)$.

Wiener's predicting operator for this case is given by

$$k(\lambda) = \sum_{p=1}^{h} c_p (i\lambda)^{p-1}.$$

The stochastic prediction for

$$\int_{t+s}^{t+s+u} x(t)\,dt \qquad \text{or} \qquad \frac{1}{u}\int_{t+s}^{t+s+u} x(t)\,dt$$

is also quite similar.

Lastly, we only remark that the stochastic prediction of moving averages of the type (3.4.2), which may have a trend of the form (3.5.3), can be also reduced to the discrete parameter case.

Chapter V. Fiducial Prediction*

§ **5.1. Introduction.** It seems difficult to apply the stochastic prediction defined in the preceding chapter to some discontinuous processes. In his recent publication, R. A. Fisher [32] has given a concept of fiducial prediction which is, in the special case of random samples, essentially equivalent to the idea of the author's stochastic prediction, and he has also refered to in his book that his fiducial prediction is not available in some discontinuous cases, such as binomial distribution.

In this chapter, we shall define another concept of fiducial prediction, from which Fisher's fiducial prediction should be distinguished, which is a mean conditional distribution of a predictand with respect to the fiducial distribution of unknown parameters, and we shall show that the fiducial prediction coinsides with the stochastic prediction, if they exist, under some conditions.

§ **5.2. Fiducial prediction.** The concept of fiducial probability put forward and stressed by R. A. Fisher may be the most disputed technical concept of modern statistics. That seems to be due to the term 'probability,' so we shall use here the terminology 'fiducial distribution.' Firstly, let $x=(x_1, x_2, \cdots, x_n)$ be a random sample, $t=t(x)$ a set of h statistics and $\theta=(\theta_1, \theta_2, \cdots, \theta_h)$ a set of population parameters. Then, a necessary and sufficient condition that the t is a sufficient set of statistics for θ is that the likelihood function for (θ, θ'), $L(x;\theta,\theta')$, is written in the form

$$L(x;\,\theta,\theta')=g(t;\,\theta)\,g_1(x;\,\theta')\,,\qquad g, g_1\geqq 0\,, \tag{5.2.1}$$

where θ' is the set of nuisance parameters and $g_1=0$ for the x such that $L=0$. In the following we assume that the $g(t;\theta)$ is a continuous function of θ almost every-where on the space Θ of θ.

Definition. *If* $\int_\Theta g(t;\theta)\,d\theta<\infty$,

$$\phi(\theta;\,t)=g(t;\,\theta)\Big/\int_\Theta g(t;\,\theta)\,d\theta \tag{5.2.2}$$

is said to be the fiducial distribution or fiducial density function of θ.

* Ogawara [23], provided Definitions and Theorems are improved and extended.

Obviously $\phi(\theta; t)$ is proportional to the likelihood of (θ, θ') for a fixed x and the maximum likelihood estimate of the θ is the value of θ which maximizes the fiducial density function $\phi(\theta; t)$. Some examples are as follows.

The case of binomial distribution:

$$t = x_1 + x_2 + \cdots + x_n \qquad (x_i = 0 \text{ or } 1)\,,$$
$$g(t; p) = p^t (1-p)^{n-t} \qquad (t = 0, 1, \cdots, n)\,,$$
$$\phi(p; t) = p^t (1-p)^{n-t} / B(t+1, n-t+1) \qquad (0 < p < 1)\,. \tag{5.2.3}$$

The case of multinomial distribution:

$$t = (t_1, t_2, \cdots, t_k) = \sum x_i\,, \qquad x_i = (x_{1i}, x_{2i}, \cdots, x_{ki})\,,$$
$$g(t; p) = p_1^{t_1} p_2^{t_2} \cdots p_k^{t_k} \qquad (t_1 + t_2 + \cdots + t_k = n),$$
$$\phi(p; t) = \frac{(n+k-1)!}{\prod_{j=1}^{k} t_j!} p_1^{t_1} p_2^{t_2} \cdots p_k^{t_k} \qquad (p_1 + p_2 + \cdots + p_k = 1)\,. \tag{5.2.4}$$

The case of Poissonian distribution:

$$t = x_1 + x_2 + \cdots + x_n\,,$$
$$g(t; m) = e^{-nm} (nm)^t / t! \qquad (t = 0, 1, \cdots)\,,$$
$$\phi(m; t) = \frac{n}{t!} (nm)^t e^{-nm} \qquad (0 < m < \infty)\,. \tag{5.2.5}$$

The case of two dimensional normal distribution:

$$\bar{x} = \sum x_i / n\,, \qquad s^2 = \sum (x_i - x)^2 / n\,,$$
$$g(\bar{x}, s^2; m, \sigma^2)$$
$$= (2\pi\sigma^2)^{-n/2} \exp[-\{n(\bar{x}-m)^2 + ns^2\}/2\sigma^2] \quad \begin{pmatrix} -\infty < \bar{x} < \infty \\ 0 < s^2 < \infty \end{pmatrix},$$
$$\phi(m, \sigma^2; \bar{x}, s^2)\, dm d\sigma^2$$
$$= \frac{\sqrt{n}}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{n(m-\bar{x})^2}{2\sigma^2}\right] \left\{\Gamma\left(\frac{n-1}{2}\right)\right\}^{-1} \left(\frac{ns^2}{2\sigma^2}\right)^{(n-3)/2}$$
$$\cdot \exp\left[-\frac{ns^2}{2\sigma^2}\right] dm \left| d\left(\frac{ns^2}{2\sigma^2}\right)\right| \quad \begin{pmatrix} -\infty < m < \infty \\ 0 < \sigma^2 < \infty \end{pmatrix}. \tag{5.2.6}$$

Now, as we have already seen, for some finite parameter models of time series we can introduce conditional independence into a set x of some members of the time series by fixing another set x' of variates. If then the conditional likelihood function $L(x; \theta, \theta' | x')$ have the form

$$L(x; \theta, \theta' | x') = g(t; \theta | x') g_1(x; \theta' | x')\,, \quad g, g_1 \geqq 0\,, \tag{5.2.7}$$

we may define a conditional fiducial distribution $\phi(\theta; t | x')$ from the g function in the same way as before.

Definition. *Let $\phi(\theta; t|x')$ be a conditional fiducial distribution of θ and let $F(y; \theta|x'')$ be the conditional probability distribution function of a predictand y, given the values of a set of condition variables x'', and assume that the $F(y; \theta|x'')$ is dependent only on the set of parameters θ. Then, a fiducial prediction for the y given t, x' and x'' may be defined by*

$$\Psi(y; t, x', x'') = \int_{\Theta} F(y; \theta|x'') \phi(\theta; t|x') d\theta . \qquad (5.2.8)$$

The density function $\psi(y; t, x', x'')$ of (5.2.8) in the case when it is absolutely continuous and the saltus $\psi(y; t, x', x'')$ of (5.2.8) in the case when it is discrete, both of them we shall call the fiducial prediction of y.

Fiducial prediction by a random sample will be able to do without the condition variables in the above definitions.

Example 1. *Binomial case.* For the prediction of y, the number of an event which will occur in m trials, we get from (5.2.3) and (5.2.8)

$$\psi(y; t) = \frac{(n+1)!\, m!\, (t+y)!\, (n+m-t-y)!}{(n+m+1)!\, y!\, (m-y)!}, \quad y = 0, 1, \cdots, m \qquad (5.2.9)$$

which is identical with the Bayesian prediction of y under the assumption that a priori distribution of p is uniform.

Example 2. *Multinomial case.* The prediction of $y_1, y_2, \cdots, y_k$ ($\sum y_j = m$), the numbers of k events in m trials is found, from (5.2.4), to be

$$\psi(y_1, \cdots, y_k; t_1, \cdots, t_k) = \frac{(n+k-1)!\, m! \prod_{j=1}^{k} (t_j + y_j)!}{(n+m+k-1)! \prod_{j=1}^{k} t_j! \prod_{j=1}^{k} y_j!}, \quad \sum y_j = m . \qquad (5.2.10)$$

This is also coincident with Bayesian prediction based on the uniform a priori probability distribution of parameters.

As an objective method of classified weather forecasting, the predictand is classified into k classes e_j ($j = 1, 2, \cdots, k$), the combinations of causes (predictors) are supposed in l levels c_i ($i = 1, 2, \cdots, l$) to each of them corresponds a set of multinomial probabilities p_{ij} ($j = 1, 2, \cdots, k$), $\sum_j p_{ij} = 1$ ($i = 1, 2, \cdots, l$) and the empirical relative frequencies are hitherto used instead of the p_{ij}. (e.g. J. C. Thompson [35]) For such case we may use (5.2.10) with $m = 1$: the fiducial prediction for '$y_j = 1$' based on a predictor is given by

$$(t_j + 1)/(n + k) \qquad (j = 1, 2, \cdots, k) \qquad (5.2.11)$$

Example 3. *Poissonian case.* Let y be the value to be observed in future, then by (5.2.5) we get

$$\psi(y; t) = \left(\frac{n}{n+1}\right)^{t+1} \binom{t+y}{y} \left(\frac{1}{n+1}\right)^{y}, \quad y = 0, 1, 2, \cdots . \qquad (5.2.12)$$

Now, as to the relation between the stochastic prediction given in the preceding chapter and the fiducial prediction defined above, we have

Theorem 19. *Let $t = t(x^{\circ})$ be a sufficient set of statistics for a set of parameters θ, let $y = y(x)$ be a set of statistics to be observed in future whose probability distribution depends only on the θ and assume that the x° and the x are conditionally independent subsets of random variables in a stochastic process when the values of the members of another subset x' of the sample time series are fixed. If a stochastic prediction of the y is derived from a statistic*

$$u = u(y, t; x') \tag{5.2.13}$$

the distribntion of which is independent of the population parameters, if the distribution function of t, $G(t; \theta|x')$, and the fiducial distribution function of θ, $\Phi(\theta; t|x') = \int^{\theta}\phi(\theta; t|x')\,d\theta$, both have the same absolute increment as a function of $\xi = \xi(t, \theta; x')$,

$$d_tG(t; \theta|x') = |d_\xi H(\xi; x')|, \quad d_\theta\Phi(\theta; t|x') = |d_\xi H(\xi; x')|, \tag{5.2.14}$$

if

$$\{\xi; \theta(\xi, t; x') = \theta, t \in R_t\} = \Xi(x') \quad \textit{for any fixed } \theta,$$
$$\{\xi; t(\xi, \theta; x') = t, \theta \in \Theta\} = \Xi(x') \quad \textit{for any fixed } t, \tag{5.2.15}$$

where R_t, Θ and $\Xi(x')$ is the space of all values of t, θ and ξ respectively, and if u and y have the same dimension, then the stochastic prediction of y coincides with the fiducial prediction of y.

Proof. We may omit the letter x' of the set of condition variates in the expressions below without loss of generality of our proof. According to the definition, the stochastic prediction of y is derived from the distribution function $K(u)$ of u by changing variable, $u = u^*(y, t_0)$, where the $t = t(x^{\circ})$ is fixed at its observed value t_0. Since y and t are conditionally independent, if we denote the (conditional) distribution function (given x') of the y by $F(y; \theta)$,

$$K(u) = \int_{Du(y,t)} d_yF(y; \theta)\, d_tG(t; \theta),$$

where

$$D_u(y, t) = \{(y, t); u(y, t) \leq u\}.$$

In this and in the following, an inequality for two vector quantities stands for the inequalities for their each components. Changing variables from (y, t) into (y, ξ), we have

$$K(u) = \int_{Du,\theta(y,\xi)} d_yF(y; \theta)|dH(\xi)|, \tag{5.2.16}$$

which is independent of θ, where

$$D_{u,\theta}(y, \xi) = \{(y, \xi)\,;\, u(y, t(\xi, \theta)) \leqq u\} \qquad (5.2.17)$$

and the stochastic prediction of y is given by $K(u^*(y, t_0))$ (symbolically) by changing variable u into y through $u = u^*(y, t_0)$.

On the other hand, if we denote the cumulative fiducial prediction of the y by $\Psi(y; t_0)$, owing to the definition (5.2.8), we get

$$\int_{u^*(y,t_0)\leqq u^*} d_y\Psi(y\,;\,t_0) = \int_{D_{u^*,t_0}(y,\theta)} d_yF(y\,;\,\theta)\, d_\theta\Phi(\theta\,;\,t_0)$$
$$= \int_{D_{u^*,t_0}(y,\xi^*)} d_yF(y\,;\,\theta)\,|dH(\xi^*)|\,, \qquad (5.2.18)$$

where

$$D_{u^*,t_0}(y, \theta) = \{(y, \theta)\,;\, u^*(y, t_0) \leqq u^*,\ \theta \in \Theta\}$$
$$D_{u^*,t_0}(y, \xi^*) = \{(y, \xi^*)\,;\, u^*(y, t_0(\xi^*, \theta)) \leqq u^*,\ \theta(\xi^*, t_0) \in \Theta\}\,. \qquad (5.2.19)$$

Since $u^*(y, t) \equiv u(y, t)$, $\{\xi^*; \theta(\xi^*, t_0) \in \Theta\} \equiv \Xi$, and according to (5.2.15),

$$D_{u,\theta}(y, \xi) = D_{u,t}(y, \xi)\,.$$

Hence

$$\int_{u^*(y,t_0)\leqq u^*} d_y\Psi(y\,;\,t_0) = K(u^*)\,, \qquad (5.2.20)$$

where $u^* = u(y, t_0)$. The theorem was thus proved.

In usual cases we meet with where a stochastic prediction is possible, $G(t; \theta)$ and $\Phi(\theta; t)$ have the properties assumed in this theorem.*

The following theorem is obvious.

Theorem 20. *If the set of sufficient statistics $t = t(x^\circ)$ is consistent, the fiducial prediction* (5.2.8) *converges to the conditional probability distribution function of y given x'' in probability for every y, as the size of x° tends to infinity.*

Example 4. Let $\bar{x}$ and s^2 be the sample mean and the sample variance of a random sample of size n drawn from a normal population $N(m, \sigma^2)$ respectively. Then, $(\bar{x}, s^2)$ is a sufficient set of statistics for the set of parameters (m, σ^2) and the joint distribution element of the (m, σ^2) is given by (5.2.6) :

---

* If the y is another set of sufficient statistics for θ and independent of the t, the hypothesis to Theorem 19 seems to correspond to the conditions for *additivity* of independent sufficient statistics with respect to the parameter θ and then the u is a *relative statistic* between y and t in Kitagawa's theory. (Kitagawa [36])

$$\phi(m, \sigma^2; x, s^2)\, dm\, d\sigma^2$$
$$= \sqrt{\frac{n}{2\pi}} \exp\left[-\frac{n(m-x)^2}{2\sigma^2}\right] \frac{1}{\Gamma\left(\frac{n-1}{2}\right)} \left(\frac{ns^2}{2\sigma^2}\right)^{(n-3)/2} \exp\left[-\frac{ns^2}{2\sigma^2}\right] d\left(\frac{m-x}{\sigma}\right) d\left(\frac{ns^2}{2\sigma^2}\right)$$
$$(-\infty < m < \infty,\ 0 < \sigma^2 < \infty) .$$

In this case, we may set $\xi = (\sqrt{n}\,(x-m)/\sigma,\ ns^2/2\sigma^2)$ in Theorem 19. The distribution densities of a variable y to be observed in future is

$$f(y; m, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y-m)^2}{2\sigma^2}\right].$$

Therefore, the fiducial prediction of the y is found to be

$$\psi(y; \bar{x}, s^2) = \int_{-\infty}^{\infty} dm \int_0^{\infty} f(y; m, \sigma^2)\, \phi(m, \sigma^2; \bar{x}, s^2)\, d\sigma^2$$
$$= \frac{1}{B\left(\frac{1}{2}, \frac{n-1}{2}\right)} \{(n+1)s^2\}^{-1/2} \left[1 + \frac{(y-\bar{x})^2}{(n+1)s^2}\right]^{-n/2} \qquad (5.2.21)$$
$$(-\infty < y < \infty) .$$

On the other hand, the stochastic prediction of y is given by the distribution of

$$y = \bar{x} + t_{n-1} s \sqrt{(n+1)/(n-1)}\,, \qquad (5.2.22)$$

where t_{n-1} denotes a random variable distributed according to the t distribution with $n-1$ degrees of freedom and $\bar{x}$ and s^2 are constants. Thus obtained stochastic prediction is obviously identical with (5.2.21). In this example, condition variables are absent, and the prediction tends to $N(m, \sigma^2)$ in probability at each point of y, when $n \to \infty$; in fact, according to Theorem 18, it converges with probability 1.

§ 5.3. An application to Markov sequence. Let $x(n)$ $(n = 0, \pm 1, \cdots)$ be a (strict sense) stationary multiple Markov process of order h which can only assume values in a finite set $\{e_1, e_2, \cdots, e_k\}$, and let $x_{n-\nu}$ $(\nu = 0, 1, \cdots, L)$ be a sample sequence. In order to predict $y = x(n+m)$ $(m > 0)$, we put

$$y_j \equiv x_{n-j(h+m)+m}, \qquad j = 1, 2, \cdots, N,$$
$$x'_{jp} \equiv x_{n-j(h+m)-p+1}, \qquad p = 1, 2, \cdots, h;\ j = 0, 1, \cdots, N,$$

then, according to Theorem 2, if we fix the values of x'_{jp} $(p = 1, 2, \cdots, h;$ $j = 0, 1, \cdots, N)$, $N+1$ random variables y, y_j $(j = 1, 2, \cdots, N)$ are conditionally independent.

If $(x'_{01}, x'_{02}, \cdots, x'_{0h}) = (e'_1, e'_2, \cdots, e'_h)$, where e'_p is one of the e_q $(q = 1, 2, \cdots, k)$, if the number of set $(x'_{j(i)1}, x'_{j(i)2}, \cdots, x'_{j(i)h})$ which is equal to $(e'_1, e'_2, \cdots, e'_h)$ out of $(x'_{j1}, x'_{j2}, \cdots, x'_{jh})$ $(j = 1, 2, \cdots, N)$ is N' and if the number of e_q among $y_{j(i)}$ $(i = 1, 2, \cdots, N')$ is N_q $(q = 1, 2, \cdots, k;\ \sum_q N_q = N')$, then in accordance with (5.2.11), the (conditional) fiducial prediction that $y=e_q$ $(q=1, 2, \cdots, k)$, given x'_{jp} $(p=1, 2, \cdots, h; j=0, 1, \cdots, N)$ is given by

$$(N_q+1)/(N'+k) \qquad (q=1, 2, \cdots, k)\,. \tag{5.3.1}$$

Although such method is exact in the meaning of the definition, as well as the stochastic prediction in the last chapter, provided the assumptions upon which the stochastic process is based are correct, it involves the sacrifice of some proportion of the data available and the sample size is considerably diminished. This seems unavoidable owing to the dependence of members of a sample. On the other side, however, it will still be an important problem left to study more effective methods.

References *(continued)*

[31] N. WIENER; *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, Wiley (1949)

[32] R. A. FISHER; *Statistical Methods and Scientific Inference*, Oliver & Boyd (1956)

[33] S. S. WILKS; *Mathematical Statistics*, Princeton Univ. Press (1944)

[34] S. S. WILKS; *Certain generalizations in the analysis of variance*, Biometrika, **24** (1932), 471–494.

[35] J. C. THOMPSON; *A numerical method for forecasting rainfall in the Los Angeles Area*, Month. Weath. Rev., 78 (1950), 113–124.

[36] T. KITAGAWA; *Successive process of siatistical inference associated with an additive family of sufficient statistics*, Bull. Math. Stat, 7 (1957), 92–112.