Time Series Analysis and Stochastic Prediction (I)

Ogawara, Masami
Meteorological Research Institute

https://doi.org/10.5109/12983

出版情報：統計数理研究. 8 (1/2), pp.8-53, 1958-03. Research Association of Statistical Sciences
バージョン：
権利関係：

TIME SERIES ANALYSIS AND STOCHASTIC PREDICTION (I)

By

Masami OGAWARA

(*Received* February 20, 1958)

Contents

Introduction The purpose of this paper is to prepare some ideas and methods for practices of time series analysis and prediction.

Although most of the present methods seem to be based on the large sample theory and they may offer powerful tools for some problems in special fields such as electric communications, almost every time series we meet with in the natural or social sciences or in manufactories is a small sample which needs an exact method of statistical inference. In order to infer the probability theoretic characters of a time series and to make a prediction, on the basis of small number of observations, it seems to be inevitable to confine ourselves to a suitably selected family of finite parameter schemes, except some nonparametric methods for testing randomness.

In this paper, we shall thus treat rather special stochastic models (Chapter II), using them we shall introduce a conditional test schemes (Chapter III), and we shall define an idea of stochastic prediction which is represented in a form of distribution function (Chapter IV and V) and is useful for operations researches related to a prediction (Chapter VI).

Although our method may not be optimal for a small sample, it will asymptotically show optimal properties if the size of time series tends to infinity. Practical applications of the methods are seen in other papers by the author [1 ~ 23].

The author wishes to express his hearty thanks to Professor T. Kitagawa for his continual encouragement throughout this work.

Chapter I. Time Series

§ **1.1 Sample of a stochastic process.** Time series is a series of observed values which is assumed to be a sample or a realization of a stochastic process. In the majority of practices, however, it seems reasonable to assume that the observation has been made only about a part of the sample function. In this section we shall consider the circumstances like this a little further.

Let ω denote a sample point of a sample space Ω, where a probability measure P is defined on the Borel field $\mathfrak{B}$ on Ω. Then a stochastic process is a one parameter family of random variables $\{x(t, \omega)\ ;\ t \in T\}$, where T is a set of real numbers. A real valued stochastic process is understood as a random vector or a random function $x(\omega) = \{x(t, \omega)\ ;\ t \in T\}$, which takes the values in the set R^T of real functions defined on T, where a probability measure is defined on the Borel field $\mathfrak{B}^T$ on the R^T. Thus we may suppose that ω is a point of $\Omega \equiv R^T$ and P is the probability measure P_T on $\mathfrak{B}^T$ and we may denote the probability field by $R^T(\mathfrak{B}^T, P_T)$ or simply by $R^T(P_T)$.

If $\{x(t, \omega)\ ;\ t \in T\}$ is a stochastic process on $R^T(P_T)$, then $\{x(t, \omega)\ ;\ t \in T'\}\ (T' \subset T)$ is a stochastic process on $R^{T'}(P_{T'})$ which may be provisionally called a *partial process* of the primary one, where $P_{T'}$ is a 'marginal probability distribution' of P_T.

Two *random functions* $x_1(\omega) = \{x_1(t, \omega)\ ;\ t \in T_1\}$ on $R^{T_1}(P_{T_1})$ and $x_2(\omega) = \{x_2(t, \omega)\ ;\ t \in T_2\}$ on $R^{T_2}(P_{T_2})$ are said to be equivalent on $T = T_1 \cap T_2$ if $P\{\omega\ ;\ x_1(\omega) \neq x_2(\omega)\} = 0$ on $R^T(P_T)$ and then we write $x_1(\omega) = x_2(\omega)$ (a.s) on T. Two *stochastic processes* $\{x_1(t, \omega)\ ;\ t \in T_1\}$ on $R^{T_1}(P_{T_1})$ and $\{x_2(t, \omega)\ ;\ t \in T_2\}$ on $R^{T_2}(P_{T_2})$ are said to be equivalent on $T = T_1 \cap T_2$ if $P\{\omega\ ;\ x_1(t, \omega) \neq x_2(t, \omega)\} = 0$ for every $t \in T$ and we denote this equivalency by $x_1(t, \omega) = x_2(t, \omega)$ (a.s.) $(t \in T)$. Evidently, if $x_1(\omega) = x_2(\omega)$ (a.s.) then $x_1(t, \omega) = x_2(t, \omega)$ (a.s.), the converse is not generally true, except the case where both processes are almost surely continuous on T.

Let $\{x_1(t, \omega')\ ;\ t \in T'\}$ and $\{x_2(t, \omega'')\ ;\ t \in T''\}\ (T' \cap T'' = 0)$ be two stochastic processes on the probability fields $R^{T'}(P_{T'})$ and $R^{T''}(P_{T''})$ respectively. We can then define a product space $R^T = R_{T'} \otimes R^{T''}$ $(T = T' \cup T'')$ of $\omega = (\omega', \omega'')$ and Borel field $\mathfrak{B}^T$ on it, and we can introduce a probability measure P_T on the $\mathfrak{B}^T$ such that $\int_{E' \otimes R^{T''}} P_T(d\omega) = P_{T'}(E')$ for every $E' \in \mathfrak{B}^{T'}$ and $\int_{R^{T'} \otimes E''} P^T(d\omega) = P_{T''}(E'')$ for every $E'' \in \mathfrak{B}^{T''}$. Thus a stochastic process $\{x(t, \omega)\ ;\ t \in T\}$ is defined on $R^T(P_T)$ and equivalent to $x_1(t, \omega')$ on T' and to $x_2(t, \omega'')$ on T''. Such process may be called a *combination* of the two stochastic processes.

If $P_T(E' \otimes E'') = P_{T'}(E') P_{T''}(E'')$ for every $E' \in \mathfrak{B}^{T'}$ and $E'' \in \mathfrak{B}^{T''}$, then $\{x_1(t, \omega'); t \in T'\}$ and $\{x_2(t, \omega''); t \in T''\}$ are independent. The combination of three or more processes can be similarly defined

The above stated definitions and concepts about stochastic processes can be generalized to the case where Ω is the space of finite dimensional vector functions on T, to the case where T is the n dimendional Euclidian space and to the case where $x(t, \omega)$ is a set function of a Borel set t. Therefore let us write a probability field generally as $\Omega_T(P_T)$.

Now, the sampling from a stochastic process $\{x(t, \omega); t \in T\}$ on $\Omega_T(P_T)$ consists of the sampling from T and the sampling from Ω_T, and the general type of sampling may be formulated as follows. Let $\{x(t, \omega'); t \in T'\}(\omega' \in \Omega_{T'}(P'_{T'}), T' \subset T)$ be a partial process of our primary process. We take a subset A from T and a subset $\{\omega'_1, \omega'_2, \cdots, \omega'_k\}$ from $\Omega_{T'}$, thus we get a set of time series

$$x_i(t) \equiv x(t, \omega'_i) \ (t \in A) \ \omega'_i \in \Omega_{T'} \quad (i = 1, 2, \cdots, k).$$

Thus the functions (or sequences) $x_i(t)$ $(i = 1, 2, \cdots, k)$ are definite on T' and observed on A, they are definite but unknown on $T' - A$ and they are not only unknown but also indefinite on the domain $T - T'$. More generally, we may draw different T'_i $(i = 1, 2, \cdots, k)$ from T and different A_i $(i = 1, 2, \cdots, k)$ from the T'_i $(i = 1, 2, \cdots, k)$ respectively and ω samplings may be made for each sample space Ω'_{T_i}, thus we get a set of observed sample functions $x_i(t) \equiv x(t, \omega'_{ij})$ $(t \subset A_i)$ $(j = 1, 2, \cdots, k_i; i = 1, 2, \cdots, k)$, where $A_i \subset T'_i \subset T$, $\omega'_{ij} \in \Omega_{T'_i}$. The samplings in the case where $T'_i \equiv T$ and $\Omega_{T'} \equiv \Omega_T$ are discussed in detail by Kitagawa [24].

When $\{x(t, \omega); t \in T\}$ is a one dimensional real process, if we take systematically a set of discrete points $A = \{t_1, t_2, \cdots, t_n\}$ from T and if $\{\omega_1, \omega_2, \cdots, \omega_k\}$ $(k > n \geqq 1)$ is a random sample drawn from Ω_T, $\{x_i(t_1), x_i(t_2), \cdots, x_i(t_n)\}$ $(i = 1, 2, \cdots, k)$ constitutes a random sample of size k from an n dimensional population and we may apply the multivariate analysis to our statistical inferences.

However, when T is the passage of historical time, we can only observe at most one sample function $x(t)$, which is corresponding to a single ω $(k = 1)$, on a part of the set T. In this paper, we shall treat mainly with such case and suppose that a time series $x(t)$ $(t \in A)$ has been got by one of the following two types of sampling schemes.

Type I. $x(t) \equiv x(t, \omega')$ $t \in A \subset T' \subset T$, $\omega' \in R^{T'}(P_{T'})$,
where T is the real line or the set of integral numbers.

Type II. We consider k partial processes $\{x(t, \omega'_i), t \in T'_i\}$, $\omega'_i \in R^{T'_i}(P_{T'_i})$ $(i = 1, 2, \cdots, k)$, where $T'_i \subset T$ and $T'_i \cap T'_j = 0$ $(i \neq j)$, and suppose that k sample functions

$$x(t, \omega'_i) \qquad t \in A_i \subset T'_i \qquad (i = 1, 2, \cdots, k)$$

constitute an observed function $x(t)$ on $A = \bigcup_{i=1}^{k} A_i$, *i.e.*

$$x(t) \equiv x(t, \omega_i') \quad \text{for} \quad t \in A_i \qquad (i = 1, 2, \cdots, k) .$$

The partial processes are not independent in general.

Hereafter, in this paper, we shall use simple notations such as $x_t(t \in A)$ for a sample function or sample sequence and $x(t)$ $(t \in T)$ for a stochastic process, however sometimes we shall not distinguish them if not necessary.

§ **1.2 Stationarity and trend of a time series.** Let $x(t)$ $(t \in T)$ be a one dimendional real stochastic process. The process $x(t)$ is called stationary (in the wide sense) if the mean value function $m(t) = Ex(t)$ and the auto-covariance function $\Gamma(t, t+s) = E\{(x(t) - m(t))(x(t+s) - m(t+s))\}(t, t+s \in T)$ are both finite and do not depend on t. A time series is stationary if it is a sample of a stationary stochastic process.

Definition *A time series is called to have a trend, if it is not stationary in the wide sense, $m(t) < \infty$ and $\Gamma(s, t) < \infty (s, t \in T)$. If $m(t)$ depends on t, it is called the trend of mean value, if $\sigma^2(t) \equiv \Gamma(t, t)$ depends on t it is the trend of variance and if $\rho(\tau; t) \equiv \Gamma(t, t+\tau)/\sigma(t)\sigma(t+\tau)$ $(\sigma(t), \sigma(t+\tau) > 0)$ depends on t the autocorrelation function is said to have a trend.*

Thus the definition of stationarity of a time series is clearly given. We have, however, no criterion to decide wheather a given time series is stationary or not unless we have some physical evidences about the variation of the time series, that is, in any case we cannot statistically reject the hypothesis that the given time series is stationary. For instance, let a time series x_t observed on a closed interal A be a sample of a stochastic process of the type

$$x(t) = m(t) + z(t)(t \in A), \tag{1.2.1}$$

that is $x_t = m(t) + z_t$, where z_t is a sample of a stationary process $z(t)(-\infty < t < \infty)$ with zero mean and $m(t)$ is a continuous trend defined on A for the mean value of $x(t)$. Then, from a different point of view, we can suppose for the same time series $x_t(t \in A)$ that the $m(t)(t \in A)$ is a sample function of a stationary process $y(t)$ $(-\infty < t < \infty)$. Thus the $x_t(t \in A)$ can be also assumed to be a sample function of a stationary process which has the structure

$$x^*(t) = y(t) + z(t)(-\infty < t < \infty) . \tag{1.2.2}$$

As a special case we may take an almost periodic function $y_t(-\infty < t < \infty)$ which coincides with $m(t)$ on A. Then the y_t is a sample function of a stationary non-regular (or deterministic) process $y(t)$ $(-\infty < t < \infty)$.

In the following cases we can draw statistical inferences about the trend $m(t)$ $(t \in A)$ in (1.2.1).

1) If the $z(t)$ $(t \in A; A = \{1, 2, \cdots, N\})$ is a independent stationary normal stochastic sequence and the $m(t)$ is assumed to have the form

$$m(t)=\sum_j c_j\varphi_j(t)\,, \tag{1.2.3}$$

where the c's are unknown parameters and $\{\varphi_j(t)\}$ is a set of known functions, then we can apply the normal regression theory. Particularly, if the $m(t)$ is a step function such that

$$\begin{aligned} m(t)=m_i \qquad & t=n_i+1,\ n_i+2,\ \cdots,\ n_{i+1} \qquad (i=1,2,\cdots,k) \\ & n_1=0,\ n_i+1<n_{i+1} \quad (i=1,2,\cdots,k)\,,\ n_{k+1}=N\,, \end{aligned} \tag{1.2.4}$$

then the usual methods of testing or estimating mean values are used.
2) If the $z(t)$ process comply with an autoregressive scheme or a moving averages and if the functions $\varphi_j(t)$ in (1.2.3) are of special types we have a method of inference which will be described in Chapter III.

So far as we treat the model (1.2.2) in the form $x(t)=y_t+z(t)$, there is no formal difference between (1.2.1) and (1.2.2), though they are essentially different.

In general, however, when we lack information concerning the structure of the process, the matter is complicated.

Let the sample mean, sample variance and the sample autocovariance of a time series x_t $(t=1,2,\cdots,N)$ be

$$\bar{x}=\sum_{t=1}^{N}x_t/N\,,\qquad s^2=\sum_{t=1}^{N}(x_t-\bar{x})^2/N$$

and

$$c(k)=\sum_{t=1}^{N-k}(x_t-\bar{x})(x_{t+k}-\bar{x})/(N-k)$$

respectively, and let

$$\bar{m}=\sum_{t=1}^{N}m(t)/N\,,\qquad \bar{\sigma}^2=\sum_{t=1}^{N}\sigma^2(t)/N\,, \tag{1.2.5}$$

$$\sigma_m^2=\sum_{t=1}^{N}(m(t)-\bar{m})^2/N\,. \tag{1.2.6}$$

Then we have

$$E(\bar{x})=\bar{m}\,,\qquad V(x)=\frac{1}{N^2}\sum_{s=1}^{N}\sum_{t=1}^{N}\Gamma(s,t)\,, \tag{1.2.7}$$

$$E(s^2)=\bar{\sigma}^2+\sigma_m^2-V(\bar{x})\,, \tag{1.2.8}$$

$$E(c(k))=\Gamma(k)+\bar{\Gamma}_m(k)-V(x)\,, \tag{1.2.9}$$

where

$$\bar{\Gamma}(k)=\frac{1}{N-k}\sum_{t=1}^{N-k}\Gamma(t,t+k)\,, \tag{1.2.10}$$

$$\Gamma_m(k)=\frac{1}{N-k}\sum_{t=1}^{N-k}(m(t)-\bar{m})(m(t+k)-\bar{m}) \tag{1.2.11}$$

and where the end effects are neglected.

Thus, for a family of stochastic processes for which

$$E\left\{\frac{1}{N-k}\sum_{t=1}^{N-k}(x(t)-\overline{m})(x(t+k)-\overline{m})\right\}\equiv\overline{\Gamma}(k)+\overline{\Gamma}_m(k) \quad (1.2.12)$$

and

$$E\left\{\frac{1}{N}\sum_{t=1}^{N}(x(t)-\overline{m})^2\right\}=\overline{\sigma}^2+\sigma_m^2 \quad (1.2.13)$$

are constants, if a remarkable trend $m(t)$ is assumed, σ_m^2 and $\overline{\Gamma}_m(k)$ $(1\leqq k\leqq N-1)$ are large, consequently $\overline{\sigma}^2$ and $\overline{\Gamma}(k)$ are small and sample mean $\bar{x}$ can be well estimated, but the extrapolation of such $m(t)$ will be difficult in general, on the other hand if there exist no trend the precision of the estimation of $\bar{x}$ becomes comparatively bad.

In conclusion, *the trend should be supposed and analysed according to our purpose of statistical research; it is a relative concept to how to use the statistical results.* For example, if we want to analyse the character of the time series in the observed period in order to research a relation to other phenomena, or if we want to predict the immediate future, it will be generaly relevant to adopt the model (1.2.1). While, if we are to apply the statistical results to a relatively long period of future, the model (1.2.2) may be rather useful.

The method of moving average may be useful for the estimation of various scales of trend for mean value, and for the decomposition of a given time series to various scales of stationary components, according to our research purposes.

Definitions and discussions stated above can be easily extended to multi-dimensional processes and to the trend of other parameters.

Chapter II Finite Parameter Schemes

§ 2.1 Introduction. If a stochastic process $x(t)$ is stationary in the wide sense and if the autocorrelation function $\rho(\tau)$ (or the autocovariance function $\Gamma(\tau)$) is continuous at $\tau=0$, it is a well known fact that the $\rho(\tau)$ and the spectral distribution function $F(\lambda)$ are equivalent through the relation

$$\rho(\tau)=\int e^{i\tau\lambda}dF(\lambda)\,. \quad (2.1.1)$$

Roughly speaking, there are two ways in the analysis of stationary time series, methods based on finite parameter schemes which seem to attack importance mainly to the autocovariance function, and non-parametric approaches which are rather concerned with the inference of spectral distribution. In either case, most of the current methods of practical time series analysis seem to be based on the large sample theory. In this paper, we shall treat an exact method which is applicable to small samples. For that purpose, and especially for the prediction by small sample, it will be inevitable to adopt a finite parameter scheme. In this chapter we shall outline some finite parameter stochastic models which will be used in the later chapters.

§ **2.2 Finite parameter schemes with discrete time parameter.** According to usual definition, a stochastic process $x(t)$ with integral parameter t is said to constitute a *multiple Markov process of order h*, if there is a integer h such that for each λ and t

$$Pr\{x(t) \leqq \lambda \mid x(t-k) = x_{t-k},\ k=1, 2, \cdots\}$$
$$= Pr\{x(t) \leqq \lambda \mid x(t-k) = x_{t-k}k,\ =1, 2, \cdots, h\} \quad \text{(a.s.)} \quad (2.2.1)$$

If $h=1$ the process is called a *simple Markov process.*

Theoretically, any one dimensional multiple Markov process of order h is only a component of an h dimensional simple Markov process. As we shall see later, however, multiple Markov process is not mere extension of simple Markov process, and the direct treatments of the former in its primary form seem to be rather convenient for practice.

For a simple Markov process we have the following theorem, but *it should be noticed that the corresponding theorem for a multiple Markov process does not generally hold.*

Theorem 1. *If $x(t)$ is a simple Markov process with integral parameter, then for each λ and for arbitrary $t_n < t_{n-1} < \cdots < t_1 < t$ $(n > 1)$*

$$Pr\{x(t) \leqq \lambda \mid x(t_j) = x_{t_j},\ j=1, 2, \cdots, n\}$$
$$= Pr\{x(t) \leqq \lambda \mid x(t_1) = x_{t_1}\} \qquad \text{(a.s.)} \qquad (2.2.2)$$

This is the defining equation of a simple Markov process with continuous parameter t and (2.2.2) may be also adopted for the definition in the integral parameter case.

Proof. For brevity, let us denote the probability distribution of a random variable x given the value of other random variable y by $P(x \mid y)$, $Pr(x \in X, x' \mid y)$ by $\int P(x, x' \mid y)\,dx$, where X is the whole space of x, and a set of (random) variables $(x_1, x_2, \cdots, x_n)$ by x_j $(j=1, 2, \cdots, n)$. Then by (2.2.1)

$$P(x_t \mid x_{tj}\ \ (j=1, 2, \cdots, n)) = \frac{P(x_t, x_{t_1}, \cdots, x_{t_{n-1}} \mid x_{t_n})}{P(x_{t_1}, x_{t_2}, \cdots, x_{t_{n-1}} \mid x_{t_n})}\,.$$

The numerator is equal to

$$\int P(x_{t-j}\ \ (j=1, 2, \cdots, t-t_n-1) \mid x_{t_n}) \prod_{j=1}^{t-t_1-1} dx_{t-j} \prod_{j=1}^{t_1-t_2-1} dx_{t_1-j} \cdots \prod_{j=1}^{t_{n-1}-t_n-1} dx_{t_{n-1}-j}$$

$$= \int \prod_{j=0}^{t-t_{n-1}-1} P(x_{t-j} \mid x_{t-j-1}) \prod_{j=1}^{t-t_1-1} dx_{t-j} \prod_{j=1}^{t_1-t_2-1} dx_{t_1-j} \cdots \prod_{j=1}^{t_{n-1}-t_n-1} dx_{t_{n-1}-j}$$

$$= \int \prod_{j=0}^{t-t_1-1} P(x_{t-j} \mid x_{t-j-1}) \prod_{j=1}^{t-t_1-1} dx_{t-j}$$

$$\cdot \int \prod_{j=0}^{t_1-t_n-1} P(x_{t_1-j} \mid x_{t_1-j-1}) \prod_{j=1}^{t_1-t_2-1} dx_{t_1-j} \cdots \prod_{j=1}^{t_{n-1}-t_n-1} dx_{t_{n-1}-j}$$

$$= \int P(x_{t-j}\ (j=0, 1, \cdots, t-t_1-1) \mid x_{t_1}) \prod_{j=1}^{t-t_1-1} dx_{t-j}$$

$$\cdot \int P(x_{t_1-j} (j=0, 1, \cdots, t_1-t_n-1) \mid x_{t_n}) \prod_{j=1}^{t_1-t_2-1} dx_{t_1-j} \cdots \prod_{j=1}^{t_{n-1}-t_n-1} dx_{t_{n-1}-j}$$

$$= \ P(x_t \mid x_{t_1}) P(x_{t_1}, x_{t_2}, \cdots, x_{t_{n-1}} \mid x_{t_n}) \quad \text{(a.s.)}.$$

For a multiple Markov process of oder $h(\geqq 1)$, we have the following lemmas.

Lemma 1. *When the value of $x(0)$ is given, the stochastic sequence $x(t)$ $(t=1, 2, \cdots)$ is also a multiple Markov process of oder h for the inversely directed time scale, that is, for each λ and t and for arbitrary $n > h$*

$$Pr\{x(t) \leqq \lambda \mid x(t+k) = x_{t+k},\ k=1, 2, \cdots, n;\ x(0) = x_0\}$$
$$= Pr\{x(t) \leqq \lambda \mid x(t+k) = x_{t+k},\ k=1, 2, \cdots, h;\ x(0) = x_0\} \quad \text{(a.s.)}.$$

Lemma 2. *For each λ and for arbitrary $t_1 < t < t_2$, $m, n > h$*

$$Pr\{x(t) \leqq \lambda \mid x(t_1 - i) = x_{t_1-i}\ (i=1, 2, \cdots, m);$$
$$x(t_2+j) = x_{t_2+j}\ (j=1, 2, \cdots, n)\}$$
$$= Pr\{x(t) \leqq \lambda \mid x(t_1 - i) = x_{t_1-i}\ (i=1, 2, \cdots, h);$$
$$x(t_2+j) = x_{t_2+j}\ (j=1, 2, \cdots, h)\} \quad \text{(a.s.)}.$$

From these lemmas, we get the following

Theorem 2. *Let $x(t)$ be a multiple Markov process of oder h. When the values of $x(i(h+k)+j)$ $(i=0, 1, \cdots, n;\ j=1, 2, \cdots, h)$ $(h\geqq 1, k\geqq 1)$ are given, n sets of k random variables $x((i-1)(h+k)+h+j)$; $j=1, 2, \cdots, k$ $(i=1, 2, \cdots, n)$ are mutually independent.*

For a normal stationary process $x(t)$ with integral parameter, we have the following theorem.

Theorem 3. *Let $x(t)$ be a one dimensional normal stationary process with integral parameter. Either of the following three conditions is necessary and sufficient in oder that the $x(t)$ is a non-deterministic Markov process of oder h.*

1) *$x(t)$ satisfies a stochastic finite difference equation*

$$\sum_{j=0}^{h} a_j(x(t-j)-m)=y(t) \quad \text{(a.s.)}, \quad a_0 a_h \neq 0 \tag{2.2.3}$$

where the coefficients a's are such that

$$\left.\begin{array}{c} \textit{all roots of the equation} \\ \sum_{j=0}^{h} a_j z^{h-j}=0 \\ \textit{lie within the uuit circle,} \end{array}\right\} \tag{2.2.4}$$

and where $y(t)$ is a normal stationary indepeudent process with zero mean.

2) *The autocorrelation function $\rho(\tau)$ satisfies a finite difference equation*

$$\begin{gathered} \sum_{j=0}^{h} a_j \rho(\tau-j)=0 \quad (\tau=1,2,\cdots), \quad a_0 a_h \neq 0, \\ \rho(0)=1, \quad \rho(-\tau)=\rho(\tau) \quad (\tau=1,2,\cdots), \end{gathered} \tag{2.2.5}$$

where the coefficients have the property (2.2.4).

3) *$x(t)$ has a spectral density function of the form*

$$F'(\lambda)=1/\left|\sum_{j=0}^{h} a_j e^{i(h-j)\lambda}\right|^2 (-\pi \leqq \lambda \leqq \pi), \quad a_0 a_h \neq 0 \tag{2.2.6}$$

with the condition (2.2.4).

The proofs should be refered to Ogawara [1] and Doob [26].

A stationary stochastic process which is not necessarily normal and has the properties stated above, where $y(t)$ is a non-autocorrelated stationary process, is nothing but the so-called *autoregressive scheme* or *autoregression process.*

A *moving averages of order k* is defined by either of the following three equivalent conditions.

1) A stationary process $x(t)$ with zero mean and with integral parameter is expressed by the form

$$x(t)=\sum_{j=0}^{k} b_j y(t-j) \quad \text{(a.s.)}, \quad b_0 b_k \neq 0, \tag{2.2.7}$$

where $y(t)$ is a non-autocorrelated statonary process, $Ey(t)=0$, and the b's are constants, consequently

$$\begin{aligned} \rho(\tau) &= c^2(b_0 b_\tau + b_1 b_{\tau+1} + \cdots + b_{k-\tau} b_k) & \tau &= 0, 1, \cdots, k \\ &= 0 & \tau &= k+1, k+2, \cdots, \end{aligned} \tag{2.2.8}$$

where $c^2=1/\sum_{j=0}^{k} b_j^2$.

2) The $x(t)$ process has the autocorrelation function such that

$$\rho(k) \neq 0 \quad \text{and} \quad \rho(\tau) \equiv 0 \quad \text{for} \quad \tau > k > 0. \tag{2.2.9}$$

3) The $x(t)$ process has a spectral density function of the form

$$F'(\lambda) = (c^2/2\pi) \left| \sum_{j=0}^{k} b_j e^{i(k-j)\lambda} \right|^2 (-\pi \leqq \lambda \leqq \pi), \quad b_0 b_k \neq 0. \quad (2.2.10)$$

If all roots of

$$\sum_{j=0}^{k} b_j z^{k-j} = 0 \quad (2.2.11)$$

are within the unit circle

or, even if some of them are of modulus one, they are of even multiplicity, then the moving averages is said to be *regular* (Wold [27]). In regular case the autocorrelation coefficients $\rho(\tau)$ $(\tau = 1, 2, \cdots, k)$ satisfing the condition (2) determine unique moving averages.

Autoregression process and moving averages are the special cases of a stationary process $x(t)$ defined by either of the following equivalent conditions, which has been treated by several authors and we shall provisionally call a *generalized autoregression process of orders* (h, k).

1) $$\sum_{j=0}^{h} a_j x(t-j) = \sum_{j=0}^{k} b_j y(t-j) \quad \text{(a.s.)} \quad a_0 a_h b_0 b_k \neq 0, \quad (2.2.12)$$

where $y(t)$ is a non-autocorrelated stationary process and the coefficients a's satisfy the condition (2.2.4).

2) $$\sum_{j=0}^{h} a_j \rho(\tau-j) \neq 0 \quad \tau = 1, 2, \cdots, k$$
$$= 0 \quad \tau = k+1, k+2, \cdots, \quad a_0 a_h \neq 0 \quad (2.2.13)$$

with the coefficients satisfing (2.2.4).

3) $$F'(\lambda) = \left| \sum_{j=0}^{k} b_j e^{i(k-j)\lambda} \right|^2 \Big/ \left| \sum_{j=0}^{h} a_j e^{i(h-j)\lambda} \right|^2 (-\pi \leqq \lambda \leqq \pi) \quad a_0 a_h b_0 b_k \neq 0, \quad (2.2.14)$$

where the a's satisfy (2.2.4).

For a generalized autoregressive scheme (2.2.12)

$$x^*(t) = \sum_{j=0}^{h} a_j x(t-j) \quad (2.2.15)$$

is a moving averages whose autocorrelation function is given by

$$\rho^*(\tau) = \frac{\sum_{i=0}^{h} a_i \sum_{j=0}^{h} a_j \rho(i-j+\tau)}{\sum_{i=0}^{h} a_i \sum_{j=0}^{h} a_j \rho(i-j)} = \frac{\sum_{j=0}^{k-\tau} b_j b_{\tau+j}}{\sum_{j=0}^{k} b_j^2} \quad \tau = 1, 2, \cdots, k$$
$$= 0 \quad \tau = k+1, k+2, \cdots \quad (2.2.16)$$

and

$$b_0 \prod_{j=0}^{k} (z - \beta_j) \equiv \sum_{j=0}^{k} b_j z^{k-j}, \quad (2.2.17)$$

where β_j and $1/\beta_j$ $(j=1, 2, \cdots, k)$ are the roots of

$$\sum_{\tau=-k}^{k} \rho^*(\tau) z^\tau = 0 . \qquad (2.2.18)$$

When $|\beta_j| \leqq 1$ $(j=1, 2, \cdots, k)$, the $x(t)$ may be called a *regular* generalized autoregression process (regular g.a.p.).

For a regular g a.p., if the autocorrelation function $\rho(\tau)$ is given, the coefficients a_j's can be determined by (2.2.13) and $\rho^*(\tau)$ $(\tau=1, 2, \cdots, k)$ are calculated by (2.2.16), consequently the coefficients b_j's are determined by (2.2.17) and (2.2.18), where we may set $a_0=b_0=1$. As regards of the variances, we have

$$\left[\sum_{i=0}^{\min(h,k)} a_i \sum_{j=0}^{h} a_j \rho(i-j)\right] V(x) = \left[\sum_{j=0}^{k} b_j^2\right] V(y) . \qquad (2.2.19)$$

Next, the g.a.p. is expressed in the form

$$x(t) = \sum_{j=0}^{\infty} c_j y(t-j) \qquad (c_0=1) , \qquad (2.2.20)$$

where c_j is the solution of

$$\sum_{i=0}^{j} a_i c_{j-i} = b_j \qquad (j=1, 2, \cdots) \qquad (2.2.21)$$

with the boundary conditions

$$a_0=b_0=c_0=1 , \quad a_i=0 \ (i>h) , \quad b_j=0 \ (j>k) , \quad c_j=0 \ (j<0)$$

and, by (2.2.4), $\sum_{j=0}^{\infty} c_j^2 < \infty$.

Even if the coefficients in a generalized autoregressive scheme are known and a sample series $x(t)=x_t$ $(t=1, 2, \cdots, t_0)$ is given, the sample values of $y(t)$ are indeterminate except the case of $k=0$ that is an autoregression process (a.p.). If we denote the solution of a finite difference equation in y_t,

$$\sum_{j=0}^{k} b_j y_{t-j} = \sum_{j=0}^{h} a_j x_{t-j} \qquad (t=h+1, h+2, \cdots, t_0) \qquad (2.2,22)$$

under the initial condition $y_h = y_{h-1} = \cdots = y_{h-k+1} = 0$, by y_t^0 $(t=h+1, \cdots, t_0)$, then the general solution of (2.2.22) is given by

$$y(t) = y_t^0 + \sum_{j=1}^{k} A_j(t) y(h+1-j) \qquad (t \geqq h+1) , \qquad (2.2.23)$$

where $A_j(t)$ is the solution of

$$\sum_{\nu=0}^{k} b_\nu A_j(t-\nu) = 0 \quad (b_0=1) \quad (t=h+1, h+2, \cdots, t_0) , \quad j=1, 2, \cdots, k \qquad (2.2.24)$$

and where the values of $A_j(t)$ $(t=h+1, h+2, \cdots, h+k)$, which give the boundary condition for the equation (2.2.24), are directly determined as the coefficient of $y(h+1-j)$ $(j=1, 2, \cdots, k)$ in the expressions of $y(t)$ $(t=h+1, h+2, \cdots, h+k)$ respectively, namely

$$y(h+1) = y^0_{h+1} - b_1 y(h) - \cdots - b_k y(h-k+1)$$
$$y(h+2) = y^0_{h+2} + (b_1^2 - b_2)\, y(h) + \cdots + (b_{k-1}^2 - b_k)\, y(h-k+2) + b_1 b_k y(h-k+1)$$

....

hence

$$A_1(h+1) = -b_1, \quad \cdots\cdots\cdots, \quad A_k(h+1) = -b_k$$
$$A_1(h+2) = b_1^2 - b_2, \quad \cdots\cdots\cdots, \quad A_k(h+2) = b_1 b_k$$

....

and, if the g.a.p. is regular, $A_j(t) \to 0$ $(t \to \infty)$.

Thus the linear prediction of $x(t_0+s)$ $(s>0)$ is given by the expectation and variance of the conditional random variable $x^*(t_0+s)$, given $x_t(t=1, 2, \cdots, t_0)$, that is*

$$E\{x^*(t_0+s)\} = \sum_{j=0}^{t_0-h-1} c_{s+j}\, y^0_{t_0-j}, \tag{2.2.25}$$

$$V\{x^*(t_0+s)\} = \left[\sum_{j=0}^{s-1} c_j^2 + \sum_{i=1}^{k}\left\{\sum_{j=0}^{t_0-h-1} c_{s+j} A_i(t_0-j)\right\}^2\right] V(y) \tag{2.2.26}$$

and we have

$$\left\{\sum_{j=0}^{t_0-h-1} c_{s+j} A_i(t_0-j)\right\}^2 \leqq \left\{\sum_{j=0}^{t_0-h-1} c_{s+j}^2\right\}\left\{\sum_{j=0}^{t_0-h-1} A_i^2(t_0-j)\right\} \to 0 \quad (s\to\infty).$$

Theorem 4. *Let $x_\nu(t)$ $(\nu=1, 2, \cdots, m)$ be m independent autoregression processes of order h_ν $(\nu=1, 2, \cdots, m)$ respectively. Then*

$$x(t) = \sum_{\nu=1}^{m} x_\nu(t) \tag{2.2.27}$$

is a generalized autoregression process of orders (h, k), where

$$h = \sum_{\nu=1}^{m} h_\nu, \qquad k = h - \min_\nu (h_\nu). \tag{2.2.28}$$

Lemma 3. *Let*

$$f(\lambda) \equiv \varphi(z) \equiv c_n(z^n + z^{-n}) + c_{n-1}(z^{n-1} + z^{-(z-1)}) + \cdots + c_1(z+z^{-1}) + c_0,$$

where $z \equiv e^{i\lambda}$, the c's are real constants and $c_0 c_n \neq 0$. If $f(\lambda) \geqq 0$ $(-\pi \leqq \lambda \leqq \pi)$ and if $f(\lambda)$ is an even function of λ. Then the $f(\lambda)$ can be written in the form

---

* It will be desirable to adopt the prediction scheme minimizing (2.2.26) with respect to initial values of y_t, however the influence of the initial values may be considered to be negligible, because of ergodicity, when $t_0 - h$ is not so small.

$$f(\lambda) = |a_0 + a_1 z + \cdots + a_n z_n|^2 ,$$

where the a's are real constants.

Proof. We can write

$$f(\lambda) \equiv c_n e^{-in\lambda} \prod_{j=1}^{n} (e^{i\lambda} - z_j) \left(e^{i\lambda} - \frac{1}{z_j}\right),$$

where $0 < |z_j| \leq 1$ $(j = 1, 2, \cdots, n)$. Since $\overline{f(\lambda)} = f(\lambda)$

$$f(\lambda) \equiv c_n e^{-in\lambda} \prod_{k=1}^{n} (e^{i\lambda} - \bar{z}_k) \left(e^{i\lambda} - \frac{1}{\bar{z}_k}\right)$$

$$0 < |\bar{z}_k| \leq 1 \qquad (k = 1, 2, \cdots, n) .$$

Hence each z_j is equal to a $\bar{z}_k$, and

$$\left|e^{i\lambda} - \frac{1}{\bar{z}_k}\right| = \frac{1}{|z_k|} |\bar{z}_k e^{i\lambda} - 1| = \frac{1}{|z_k|} |e^{i\lambda} - z_k| .$$

Therefore

$$f(\lambda) = |f(\lambda)| = \frac{|c_n|}{\prod_{j=1}^{n} |z_j|} \prod_{j=1}^{n} |e^{i\lambda} - z_j|^2 = K|1 + a_1' e^{i\lambda} + \cdots + a_n' e^{in\lambda}|^2 ,$$

where $K > 0$, since $f(\lambda) > 0$. Thus we can write, finally

$$f(\lambda) = |a_0 + a_1 e^{i\lambda} + \cdots + a_n e^{in\lambda}|^2 ,$$

where the a's are real, since $f(\lambda)$ is an even function of λ.

Proof of Theorem 4. Let $\Gamma(\tau)$ and $\Gamma_\nu(\tau)$ denote the autocovariance function of $x(t)$ and of $x_\nu(t)$ respectively and let the corresponding spectral functions be $F(\lambda)$ and $F_\nu(\lambda)$. Then

$$\Gamma(\tau) = \sum_\nu \Gamma_\nu(\lambda) ,$$

$$\Gamma(\tau) = \int_{-\pi}^{\pi} e^{i\tau\lambda} dF(\lambda) , \qquad \Gamma_\nu(\tau) = \int_{-\pi}^{\pi} e^{i\tau\lambda} F_\nu'(\lambda)\, d\lambda ,$$

where

$$F_\nu'(\lambda) = K_\nu / |z^{h_\nu} + a_1^{(\nu)} z^{h_\nu - 1} + \cdots + a_{h_\nu}^{(\nu)}|^2 , \qquad z \equiv e^{i\lambda} ,$$

$$K_\nu > 0 , \quad a_{h_\nu} \neq 0 .$$

Thus

$$F'(\lambda) = \sum_\nu F_\nu'(\lambda) = \sum_\nu \left\{ K_\nu / |z^{h_\nu} + a_1^{(\nu)} z^{h_\nu - 1} + \cdots + a_{h_\nu}^{(\nu)}|^2 \right\}$$

$$= \left\{ \sum_\nu \prod_\mu^{(\nu)} K_\mu \left| \sum_{j=0}^{h_\mu} a_j^{(\mu)} z^{h_\mu - j} \right|^2 \right\} \Big/ \left\{ \prod_\nu \left| \sum_{j=0}^{h_\nu} a_j^{(\nu)} z^{h_\nu - j} \right|^2 \right\} ,$$

where $\prod_\mu^{(\nu)}$ denotes the product except the νth factor. The denominator clearly has the form

$$\left|\sum_{j=0}^{h} a_j z^{h-j}\right|,$$

where $h=\sum_\nu h_\nu$ and $a_0=1$, $a_h=\prod_\nu a_{h_\nu}^{(\nu)}\neq 0$. While the numerator can be written in the form

$$\varphi(z)=g_k(z^k+z^{-k})+g_{k-1}(z^{k-1}+z^{-(k-1)})+\cdots+g_1(z+z^{-1})+g_0$$

with real coefficients, where

$$k=\max_\nu(h_1+\cdots+h_{\nu-1}+h_{\nu+1}+\cdots+h_m)=\sum_\nu h_\nu-\min_\nu(h_\nu),$$

$$g_k=K_{\nu_0}\prod_\nu^{(\nu_0)} a_{h_\nu}^{(\nu)}\neq 0 \qquad (h_{\nu_0}=\min_\nu(h_\nu)).$$

Therefore, according to Lemma 4, $\varphi(z)$ can be written in the form

$$\varphi(z)\equiv|b_0+b_1z+\cdots+b_kz_k|^2, \qquad b_0b_k\neq 0.$$

Thus the proposition was proved.

If, conversely, a g.a.p. (2.2.12) with $k<h\leqq 2k$ can be decomposed into the sum of two independent autoregression processes,

$$\left.\begin{aligned} &x(t)=x_1(t)+x_2(t),\\ &\sum_{j=0}^{h_\nu} a_j^{(\nu)}x_\nu(t-j)=y_\nu(t) \qquad (\nu=1,2),\\ &h_1=k, \quad h_2=h-k, \end{aligned}\right\} \tag{2.2.29}$$

the coefficients $a_j^{(\nu)}$ $(j=1,2,\cdots,h;\ \nu=1,2)$ must satisfy the following relation.

$$\sum_{l=1}^{j} a_{j-l}^{(1)}a_l^{(2)}=a_j \qquad (j=1,2,\cdots,h). \tag{2.2.30}$$

As we have alrady seen, regular g.a.p., consequently either of a.p. with the condition (2.2.4) and moving averages, is a special case of the *lenear process* or *general moving averages* which is defined by

$$x(t)=\sum_{j=-\infty}^{\infty} c_jy(t-j) \qquad (\sum_{j=-\infty}^{\infty} c_j^2<\infty). \tag{2.2.31}$$

Finite parameter schemes we have refered to in this section are straitforwardly extended to the case of vector process (Ogawara [2]). About the component proceses of an n dimensional a.p. the following theorem holds.

Theorem 5. *A component process of an n dimensional a.p. of order h_1 is a g.a.p. of orders (h, k), where*

$$h=nh_1, \quad k=(n-1)h_1, \tag{2.2.32}$$

except the possible degenerated case.

Conversely, if $x(t)$ is a one dimensional g.a.p of orders (h, k) such that $h/(h-k)$ is a positive integer, then the $x(t)$ is a componet process of an $n=h/(h-k)$ dimensional a.p. of order $h_1=h-k$.

A particular case, where $k=h-1$, of this converse corresponds to Theorem 3.8 of Doob [25].

Proof. Let $x(t)$ be an n dimensional a.p. of order h_1, $\mathrm{Ex}(t)=0$, that is

$$x(t)+a_1x(t-1)+\cdots+a_{h_1}x(t-h_1)=y(t)\quad \text{(a.s.)},\qquad (2.2.33)$$

where

$$x(t)=\begin{pmatrix}x_1(t)\\ \vdots\\ x_n(t)\end{pmatrix},\qquad y(t)=\begin{pmatrix}y_1(t)\\ \vdots\\ y_n(t)\end{pmatrix},\qquad a_\nu=\begin{pmatrix}a_{11}^{(\nu)} & \cdots & a_{1n}^{(\nu)}\\ \vdots & \ddots & \vdots\\ a_{n1}^{(\nu)} & \cdots & a_{nn}^{(\nu)}\end{pmatrix}.$$

In the $(k+1)n$ equations

$$\sum_{\nu=0}^{h_1}\sum_{j=1}^{n} a_{ij}^{(\nu)}x_j(t-u-\nu)=y_i(t-u)\ \text{(a.s.)}\ (a_{ij}^{(0)}=\delta_{ij})\qquad \begin{matrix} i=1,2,\cdots,n;\\ u=0,1,\cdots,k,\end{matrix}\qquad (2.2.34)$$

the number of random variables

$$x_j(t-u-\nu)\qquad (j=2,3,\cdots,n;\ \nu=1,2,\cdots,h_1;\ u=0,1,\cdots,k)\qquad (2.2.35)$$

is $(k+h_1+1)(n-1)$. Therefore, if we take the value of k such that

$$(k+1)n-(k+h_1+1)(n-1)=1\quad \text{or}\quad k=(n-1)h_1,$$

we can eliminate (2.2.35) in (2.2.34), getting an equation of the form

$$\sum_{\nu=0}^{h} c_\nu x_1(t-\nu)=\sum_{\nu=0}^{k} r_\nu\xi_\nu(t)\quad \text{(a.s.)},\qquad (2.2.36)$$

where $h=nh_1$, $k=(n-1)h_1$ and where $\xi_\nu(t)$ is the linear combination of $y_1(t-\nu),\cdots,y_n(t-\nu)$ and is non-correlated with $\xi_{\nu'}(t)$ $(\nu'\neq\nu)$. Owing to (2.2.9), $\zeta(t)=\sum_{\nu=0}^{k} r_\nu\xi_\nu(t)$ is a moving averages of order k and is written in the form

$$\zeta(t)=z(t)+b_1z(t-1)+\cdots+b_kz(t-k),\qquad (2.2.37)$$

where $z(t)$ is a non-autocorrelated stationary process.

To prove the converse, we only set $h_1=h-k$ and $n=h/(h-k)$ in (2.2.33). Then we can get the relations between the $a_{ij}^{(\nu)}$ $(i,j=1,2,\cdots,n;$ $\nu=1,2,\cdots,h)$ and the constants $c_1,c_2,\cdots,c_h$; $b_1,b_2,\cdots,b_k$ which define the process $x(t)(=x_1(t)$ (a.s.)) by the expressions (2.2.36 and 37). Since $n^2h_1=h^2/(h-k)>h+k(k\geqq 1)$, the n dimensional process is not uniquely determined.

§2.3 Finite parameter schemes with continuous time parameter. Throughout this section we consider a real (wide sense) stationary continuous parameter process $x(t)(-\infty<t<\infty)$ with $Ex(t)=0$, $Vx(t)=1$ and $Ex(t)x(t+\tau)=\rho(\tau)$.

Lemma 4. *If the first 2n derivatives* $\rho^{(\nu)}(\tau)$ $(\nu=1, 2, \cdots, 2n)$ *exist for* $-\infty<\tau<\infty$, then

$$\rho^{(2\nu-1)}(0)=0\,, \quad \rho^{(2\nu)}(0)=(-1)^{\nu}\sigma_\nu^2(\sigma_\nu^2>0) \quad (\nu=1, 2, \cdots, n) \qquad (2.3.1)$$

and the first n derived processes $x^{(\nu)}(t)$ $(-\infty<t<\infty)$ $(\nu=1, 2, \cdots, n)$ *exist in quadratic mean convergence, they are stationary and, for* $\nu, \nu'\leqq n$,

$$E\ x^{(\nu)}(t)=0\,, \quad V x^{(\nu)}(t)=\sigma_\nu^2\,, \qquad (2.3.2)$$

$$\left.\begin{aligned} &E\{x^{(\nu)}(t)\,x^{(\nu')}(t-\tau)\}=(-1)^{\nu'}\rho^{(\nu+\nu')}(\tau) \qquad (\tau>0)\\ &E\{x^{(\nu)}(t)\,x^{(\nu')}(t)\}=\begin{cases}(-1)^{\frac{\nu+\nu'}{2}}\rho^{(\nu+\nu')}(0) & \text{if } \nu+\nu'=\text{even}\\ 0 & \text{if } \nu+\nu'=\text{odd}\end{cases}\end{aligned}\right\}. \qquad (2.3.3)$$

Conversely, if the first n derived processes of $x(t)$ *exist (in q.m.), then* $\rho^{(\nu)}(\tau)$ $(\nu=1, 2, \cdots, 2n)$ *exist and* (2.3.1), (2.3.2) *and* (2.3.3) *hold.*

Proof. This lemma comes immediately from the well known case of $n=1$, and we give here a simple and direct proof of the converse only.

Since 'E' and 'lim in q.m.' can be interchanged, differentiating ν times with respect to τ both sides of

$$E\{x(t+\tau)\,x(t)\}=\rho(\tau) \qquad (\tau>0)\,,$$

we get

$$E\{x^{(\nu)}(t+\tau)\,x(t)\}=\rho^{(\nu)}(\tau)\,.$$

Since the $x^{(\nu)}(t)$ is also stationary,

$$E\{x^{(\nu)}(t)\,x(t-\tau)\}=\rho^{(\nu)}(\tau)\,.$$

Differentiating ν' times with respect to τ, we have

$$(-1)^{\nu'}E\{x^{(\nu)}(t)\,x^{(\nu')}(t-\tau)\}=\rho^{(\nu+\nu')}(\tau)\,,$$

which gives the first line of (2.3.3). Next, from the fact that

$$(-1)^{\nu}\rho^{(\nu+\nu)}(-\tau)=(-1)^{\nu'}\rho^{(\nu+\nu')}(\tau)$$

and the continuity of $\rho^{(\nu+\nu')}(\tau)$ at $\tau=0$, we get the second line of (2.3.3).

Now, as a finite parameter scheme with continuous time parameter, let us first consider the stationary (wide sense) process $x(t)$ which has the followig mutually equivalent properties.

1) The $x(t)$ has an absolutely continuous spectral distribution having the densities

$$F'(\lambda)=1\Big/\Big|\sum_{j=0}^{h} a_j(i\lambda)^j\Big|^2 \quad (-\infty<\lambda<\infty)\,, \quad a_0a_h\neq0\,, \qquad (2.3.4)$$

where the a's are real constants such that all roots of the denominator are complex and lie in the upper half-plane or all are in the lower half-plane (we suppose the former in the following) (2. 3. 5)

2) The autocorrelation function is expressed in the form

$$\rho(\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\tau\lambda}\frac{d\lambda}{\left|\sum_{j=0}^{h} a_j(i\lambda)^j\right|^2}, \qquad a_0 a_h \neq 0, \quad (2.3.6)$$

with the condition (2. 3. 5).

3) The derivatives $\rho^{(\nu)}(\tau)$ $(\nu=1, 2, \cdots)$ exist for $\tau\neq 0$ and the $\rho(\tau)$ satisfies the differential equations

$$\sum_{j=0}^{h} a_j\rho^{(j)}(\tau)=0 \qquad (\tau>0), \qquad (2.3.7)$$

$$\sum_{j=0}^{h}(-1)^j a_j\rho^{(j)}(\tau)=0 \qquad (\tau<0), \qquad (2.3.8)$$

with the boundary conditions at $\tau=0$,

$$\left.\begin{aligned}&\rho^{(2\nu-1)}(0+)=\rho^{(2\nu-1)}(0-)=0\\ &\rho^{(2\nu)}(0+)=\rho^{(2\nu)}(0-)=(-1)^\nu\sigma_\nu^2 \quad (\sigma_\nu^2>0)\end{aligned}\right\}\ \nu=1, 2, \cdots, h-1. \qquad (2.3.9)$$
$$\rho^{(2h-1)}(0+)-\rho^{(h-1)}(0-)=(-1)^h/a_h^2 .$$

In the following we assume moreover that the process $x(t)$ is normal. Let $z(t)$ be a Brownian motion on $(-\infty, \infty)$,

$$E(z(t_2)-z(t_1))=0, \quad E(|z(t_2)-z(t_1)|^2)=2\pi(t_2-t_1) \qquad (t_1<t_2) .$$

Then the spectral representation of the $x(t)$ is written in the form

$$x(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{it\lambda}\frac{dz^*(\lambda)}{\sum_j a_j(i\lambda)^j}, \qquad E\{|z^*(\lambda)|^2\}=d\lambda , \qquad (2.3.10)$$

where the $z^*(\lambda)$ is the Fourier transform of $z(t)$, and the $x(t)$ has the first $h-1$ derivatives (in q.m.), by (2. 3. 9) and Lemma 4, and satisfies the equation

$$\sum_{j=0}^{h-1} a_j\int_{-\infty}^{\infty} f^*(t)\,x^{(j)}(t)\,dt+a_h\int_{-\infty}^{\infty} f^*(t)\,dx^{(h-1)}(t)=\int_{-\infty}^{\infty} f^*(t)\,dz(t) \quad \text{(a.s.)} \qquad (2.3.11)$$

for an arbitrary function f^* which is continuous with a continuous derivative in some finite closed interval and vanishes outside the interval. For brevity we write

$$a_0x(t)+a_1x'(t)+\cdots+a_hx^{(h)}(t)=z'(t) \quad \text{(a.s.)} \qquad (2.3.12)$$

instead of (2. 3. 11). In (2. 3. 12), $x(t)$ is independent of $z'(t)$; strictly speaking,

$$E\{(z(t_2)-z(t_1))x(t)\}=0, \quad t \leq t_1 < t_2. \tag{2.3.13}$$

Now, since

$$\Psi(\lambda)=1/\left\{\sum_{j=1}^{h} a_j(i\lambda)^j\right\} \tag{2.3.14}$$

is regular and has no zero on the real line and in the lower half-plane,

$$\psi(t)=\frac{1}{2\pi}\ \underset{A\to\infty}{\text{l.i.m.}}\int_{-A}^{A} e^{i\lambda t}\Psi(\lambda)\,d\lambda=0 \qquad (t<0) \tag{2.3.15}$$

and

$$x(t)=\int_{-\infty}^{t}\psi(t-\tau)\,dz(t) \tag{2.3.16}$$

has the spectral densities (2.3.4) and the autocorrelation function (2.3.6). Thus the $x(t)$ given by (2.3.16) is a solution of the stochastic differential equation (2.3.12), $\psi(t)$ being the solution of the differential equation

$$a_0\psi(t)+a_1\psi'(t)+\cdots+a_h\psi^{(h)}(t)=0 \quad (t>0) \tag{2.3.17}$$

with the boundary conditions

$$\psi^{(j)}(0+)=0 \quad (j=0,1,\cdots,h-2), \quad \psi^{(h-1)}(0+)=1/a_h. \tag{2.3.18}$$

On the other hand, the stationary process

$$x(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{it\lambda}\frac{dz^*(\lambda)}{\sum_j a_j(-i\lambda)^j} \tag{2.3.19}$$

has as well the same spectral densities (2.3.4) and the same autocorrelation function (2.3.6) and satisfies the *backward* stochastic differential equation

$$a_0x(t)-a_1x'(t)+\cdots+(-1)^h a_h x^{(h)}(t)=z'(t) \quad \text{(a.s.)} \quad \text{(formally)} \tag{2.3.20}$$

and, by residue calculation and lemma 4,

$$E\{(z(t_2)-z(t_1))x(t)\}=0, \quad t_1<t_2\leq t. \tag{2.3.21}$$

Corresponding to (2.3.15), since $\overline{\Psi}(\lambda)=1/\{\sum_j a_j(-i\lambda)^j\}$ is regular and has no zero on the real line and in the upper half-plane, we have

$$\psi_1(t)=\frac{1}{2\pi}\ \underset{A\to\infty}{\text{l.i.m.}}\int_{-A}^{A} e^{i\lambda t}\overline{\Psi}(\lambda)\,d\lambda=0 \qquad (t>0). \tag{2.3.22}$$

Accordingly,

$$x(t)=\int_{t}^{\infty}\psi_1(t-\tau)\,dz(\tau) \tag{2.3.23}$$

is a solution of (2. 3. 20), and clearly $\psi_1(t)=\psi(-t)$. We may call (2. 3. 16) the *forward process* and (2. 3. 23) the *backward process.*

The solution of the forward equation (2. 3. 12), under the initial condition

$$x^{(j)}(t_1)=x_{t_1}^{(j)} \qquad (j=0,1,\cdots,h-1)\,, \tag{2.3.24}$$

is given by

$$x(t)=\sum_{j=1}^{h}\Big(\sum_{k} c_{jk}e^{u_k(t-t_1)}\Big)x_{t_1}^{(j-1)}+\int_{t_1}^{t}\psi(t-\tau)\,dz(\tau)\,,\qquad t>t_1\,, \tag{2.3.25}$$

where the u_k is the root of the equation

$$\sum_{j=0}^{h} a_j u^j=0$$

and its real part is negative owing to (2. 3. 5), and if its multiplicity is m, c_{jk} is a polynomial in t of order $m-1$ with the coefficients depending on t_1 and u_k $(k=1,2,\cdots)$. On account of (2. 3. 13), the solution $x(t)$ $(t>t_1)$ is independent of $x(s)$ $(s<t_1)$.

Quite similarly, the solution of the backward equation, under the initial condition

$$x^{(j)}(t_2)=x_{t2}^{(j)} \qquad (j=0,1,\cdots,h-1)\,, \tag{2.3.26}$$

is given by

$$x(t)=\sum_{j=1}^{h}\Big(\sum_{k} c_{jk}^{*}e^{u_k(t_2-t)}\Big)x_{t_2}^{(j-1)}+\int_{t}^{t_2}\psi_1(t-\tau)\,dz(\tau)\,,\qquad t<t_2\,, \tag{2.3.27}$$

where c_{jk}^{*} is obtained by replacing t_2 and $-u_k$ in c_{jk} instead of t_1 aud u_k respectively, and where $\psi_1(t-\tau)$ can be replaced by $\psi(\tau-t)$. From (2.3.21) and (2. 3. 27) we observe thst, under the condition (2. 3. 26), $x(t)$ $(t<t_2)$ and $x(s)$ $(s>t_2)$ are independent. Thus we get the following theorems.

Theorem 6. *Let* $x(t)$ $(-\infty<t<\infty)$ *be a normal stationary* (*wide sense*) *process having the spectral densities* (2. 3. 4) *or the autocorrelation function* (2. 3. 6). *If we assign the values*

$$x^{(j)}(t_0)=x_{t_0}^{(j)} \quad (j=0,1,\cdots,h-1)$$

the conditional random variables $x(s)$ *and* $x(t)$ $(s<t_0<t)$ *are independent, and the conditional distribution of the* $x(t)$ *is normal with the following mean value and variance,*

$$E\{x(t)|x_{t_0}^{(j)};\ j=0,1,\cdots,h-1\}=\mu+c_0(x_{t_0}-\mu)+\sum_{j=1}^{h-1}c_jx_{t_0}^{(j)}\,, \tag{3.2.28}$$

$$V\{x(t)|x_{t_0}^{(j)};\ j=0,1,\cdots,h-1\}=\sigma^2[1-\sum_{j=0}^{h-1}c_j(-1)^j\rho^{(j)}(t-t_0)]\,, \tag{3.2.29}$$

where $\mu=Ex(t)$, $\sigma^2=Vx(t)$ *and*

$$c_j = \frac{D_j}{D}$$

$$D_j = \begin{vmatrix} 1 & \rho'(0) & \cdots \rho^{(j-1)}(0) & \rho(t-t_0) & \rho^{(j+1)}(0) \cdots \rho^{(h-1)}(0) \\ \rho'(0) & \rho''(0) & \cdots \rho^{(j)}(0) & \rho'(t-t_0) & \rho^{(j+2)}(0) \cdots \rho^{(h)}(0) \\ \vdots & \vdots & \vdots & \vdots & \vdots \quad\quad \vdots \\ \rho^{(h-1)}(0) & \rho^{(h)}(0) & \cdots \rho^{(h+j-2)}(0) & \rho^{(h-1)}(t-t_0) & \rho^{(h+j)}(0) \cdots \rho^{(2h-2)}(0) \end{vmatrix}$$

$$D = \begin{vmatrix} 1 & \cdots\cdots & \rho^{(h-1)}(0) \\ \vdots & \ddots & \vdots \\ \rho^{(h-1)}(0) & \cdots\cdots & \rho^{(2h-2)}(0) \end{vmatrix} \qquad (j=0, 1, \cdots, h-1),$$

where $\rho^{(j)}(0)=0$ *for* j *odd, because of Lemma 4. Similar formulas hold for the conditional distribution of the* $x(s)$.

Theorem 7. *Let* $x(t)$ *be a stationary process supposed in the preceding theorem. If we assign the values*

$$C: \begin{cases} x^{(j)}(t_1) = x_{t_1}^{(j)}, \\ x^{(j)}(t_2) = x_{t_2}^{(j)} \end{cases}$$

$$(t_1 < t_2;\ j=0, 1, \cdots, h-1)$$

conditional random variables $x(s_1)$, $x(t)$ *and* $x(s_2)$ $(s_1 < t_1 < t < t_2 < s_2)$ *are mutually independent, and the conditional distribution of the* $x(t)$ *is normal with the following conditional mean value and conditional variance,*

$$E\{x(t)\,|\,C\} = \mu + c_1(x_{t_1}-\mu) + \sum_{j=2}^{h} c_j x_{t_1}^{(j-1)} + c_{h+1}(x_{t_2}-\mu) + \sum_{j=2}^{h} c_{h+j} x_{t_2}^{(j-1)}, \tag{2.3.30}$$

$$V\{x(t)\,|\,C\} = \sigma^2\Delta/\Delta_{11} = \sigma^2\Big[1 - \sum_{j=1}^{h} c_j(-1)^j \rho^{(j-1)}(t-t_1)\Big] - \sum_{j=1}^{h} c_{h+j}\rho^{(j-1)}(t_2-t), \tag{2.3.31}$$

where

$$\Delta=\left|\begin{array}{cccc}
1 & \rho(t-t_1) & -\rho'(t-t_1) & \cdots \\
\rho(t-t_1) & 1 & (-1)^{\frac{1}{2}}\rho'(0) & \cdots \\
-\rho'(t-t_1) & (-1)^{\frac{1}{2}}\rho'(0) & (-1)^{\frac{2}{2}}\rho''(0) & \cdots \\
\rho''(t-t_1) & (-1)^{\frac{2}{2}}\rho''(0) & (-1)^{\frac{3}{2}}\rho'''(0) & \cdots \\
\vdots & \vdots & \vdots & \\
(-1)^{h-1}\overset{(h-1)}{\rho}(t-t_1) & (-1)^{\frac{h-1}{2}}\overset{(h-1)}{\rho}(0) & -\rho^{\frac{h}{2}}\overset{(h)}{\rho}(0) & \cdots \\
\rho(t_2-t) & \rho(t_2-t_1) & -\rho'(t_2-t_1) & \cdots \\
\rho(t_2-t) & \rho'(t_2-t_1) & -\rho''(t_2-t_1) & \cdots \\
\vdots & \vdots & \vdots & \\
\overset{(h-1)}{\rho}(t_2-t) & \overset{(h-1)}{\rho}(t_2-t_1) & -\overset{(h)}{\rho}(t_2-t_1) & \cdots
\end{array}\right.$$

$$\left.\begin{array}{cccc}
\cdots(-1)^{h-1}\overset{(h-1)}{\rho}(t-t_1) & \rho(t_2-t) & \cdots & \overset{(h-1)}{\rho}(t_2-t) \\
\cdots(-1)^{\frac{h-1}{2}}\overset{(h-1)}{\rho}(0) & \rho(t_2-t_1) & \cdots & \overset{(h-1)}{\rho}(t_2-t_1) \\
\cdots(-1)^{\frac{h}{2}}\overset{(h)}{\rho}(0) & -\rho'(t_2-t_1) & \cdots & -\overset{(h)}{\rho}(t_2-t_1) \\
\cdots(-1)^{\frac{h+1}{2}}\overset{(h+1)}{\rho}(0) & \rho''(t_2-t_1) & \cdots & \overset{(h+1)}{\rho}(t_2-t_1) \\
\vdots & \vdots & & \vdots \\
\cdots(-1)^{h-1}\overset{(2h-2)}{\rho}(0) & (-1)^{h-1}\overset{(h-1)}{\rho}(t_2-t_1) & \cdots & (-1)^{h-1}\overset{(2h-2)}{\rho}(t_2-t_1) \\
\cdots(-1)^{h-1}\overset{(h-1)}{\rho}(t_2-t_1) & 1 & \cdots & (-1)^{\frac{h-1}{2}}\overset{(h-1)}{\rho}(0) \\
\cdots(-1)^{h-1}\overset{(2h-2)}{\rho}(t_2-t_1) & (-1)^{\frac{1}{2}}\rho'(0) & \cdots & (-1)^{\frac{h-1}{2}}\overset{(h)}{\rho}(0) \\
\vdots & \vdots & & \vdots \\
\cdots(-1)^{h-1}\overset{(2h-2)}{\rho}(t_2-t_1) & (-1)^{\frac{h-1}{2}}\overset{(h-1)}{\rho}(0) & \cdots & (-1)^{h-1}\overset{(2h-2)}{\rho}(0)
\end{array}\right|, \tag{2.3.32}$$

$\rho^{(j)}(0)=0$ *for* j *odd and where*

$$c_j=-\Delta_{1,j+1}/\Delta_{11} \qquad (j=1,2,\cdots,2h),$$

Δ_{pq} *being the cofactor of the* (p,q) *element in* Δ.

Proof. A set of random variables

$$x(t),\ x(t_1),\ x'(t_1),\ \cdots,\ x^{(h-1)}(t_1),\ x(t_2),\ x'(t_2),\ \cdots,\ x^{(h-1)}(t_2)$$

submits to a $2h+1$ dimensional normal distribution. Therefore, by Lemma 4, we get the second part of the theorem.

The autocorrelation determinant Δ' of $2h+2$ random variables

$$x(t),\ x(t_1),\ x'(t_1),\ \cdots,\ x^{(h-1)}(t_1),\ x(t_2),\ \cdots,\ x^{(h-1)}(t_2),\ x(s)$$

is given by adding to Δ the $(2h+2)$th row and the $(2h+2)$th column with the same composition

$$\rho(s-t),\ \rho(s-t_1),\ -\rho'(s-t_1),\ \cdots,\ (-1)^{h-1}\rho^{(h-1)}(s-t_1),\ \rho(s-t_2),\ \cdots,\ (-1)^{h-1}\rho^{(h-1)}(s-t_2),\ 1$$

and, by means of (2.3.7) or (2.3.8), we can easily show that the cofactor of the $(1, 2h+2)$ element in Δ' is equal to zero for either case $s<t_1<t<t_2$, $t_1<t<t_2<s$ or $t<t_1<t_2<s$. Thus the first part of the theorem is proved.

Corollary. *Under the assumptions and the condition C of the theorem 7, arbitrary linear functional $L_{(t_1,t_2)}(x(\cdot))$ of the process $x(t)$ $(t_1<t<t_2)$ is independent of $x(s)$ $(s<t_1$ or $t_2<s)$ and consequently independent of its arbitrary functional $L_{(-\infty,t_1)\cup(t_2,\infty)}(x(\cdot))$, and $E\{L_{(t_1,t_2)}(x(\cdot))\}$ is a linear combination of $x_{t_1}^{(j-1)}$, $x_{t_2}^{(j-1)}$ $(j=1, 2, \cdots, h)$, while $V\{L_{(t_1,t_2)}(x(\cdot))\}$ is independent of them.*

As a more general finite parameter shceme we have stationary processes with absolutely continuous spectral distributions having rational spectral densities

$$F'(\lambda)=\left|\sum_{j=0}^{k} b_j(i\lambda)^j\right|^2\Big/\left|\sum_{j=0}^{h} a_j(i\lambda)^j\right|^2, \qquad a_0a_hb_0b_k\neq 0, \qquad k<h, \tag{2.3.33}$$

where the a's and the b's are real, and where we can suppose, for convenience' sake, that all roots of the denominator and the imaginary roots of the numerator are in the upper half-plane. The autocorrelation function $\rho(\tau)$ of a process of this type satisfies the differential equations (2.3.7) and (2.3.8), but the $\rho(\tau)$ has only the first $2(h-k-1)$ derivatives at $\tau=0$. Accordingly the $x(t)$ has only the first $h-k-1$ derivatives (in q.m.). It is therefore the matter of course that the autocorrelation function $\rho(\tau)$ is not uniquely determined by the differential equation (2.3.7) or (2.3.8), but it depends on the coefficients b's, while in the case of scheme (2.3.4) the $\rho(\tau)$ was concretely determined by the boundary conditions (2.3.9).

Theorem 8. *Let $y(t)$ $(-\infty<t<\infty)$ be a real stationary process with spectral densities* (2.3.4). *Then the stationary process $x(t)$ with the spectral densities* (2.3.33) *is written in the form*

$$x(t)=b_0y(t)+b_1y'(t)+\cdots+b_ky^{(k)}(t) \quad \text{(a.s.)} \quad (k<h). \tag{2.3.34}$$

Proof. Since the $y(t)$ satisfies the differential equation

$$a_0y(t)+a_1y'(t)+\cdots+a_hy^{(h)}(t)=z'(t) \quad \text{(formally)}, \tag{2.3.35}$$

where $z(t)$ is a process with orthogonal increments, by means of the spectral representations of $x(t)$ and $y(t)$, the proof is immediate.

It is well known that the most general form of a stationary (wide sense) process with absolutely continuous spectral distribution is the *moving averages*

$$x(t)=\int_{-\infty}^{\infty} f(s)\,d\xi(t-s)\,, \tag{2.3.36}$$

where $f(s)$ is a Lebesque measurable function, $\int_{-\infty}^{\infty}|f(s)|^2 ds<\infty$, and the $\xi(t)$ is a process of orthogonal increments, $E\{|d\xi(t)|^2\}=dt$. Finite parameter scheme of this type, which we shall treat in the next chapter, may be specified as follows. We suppose that the $f(s)$ is a real function whose functioal type is known but involves finite number of parameters. Moreover we assume that it is continuous except an enumerable set of points, for a fixed t_0,

$$f(s)=0 \quad \text{for} \quad s>t_0>0 \quad \text{and for} \quad s<0$$

and it has continuous and non-zero points in $(0, \varepsilon)$ and in $(t_0-\varepsilon, t_0)$, ε being an arbitrary positive number less than t_0. Then the autocovariance function is given by

$$\begin{aligned} \Gamma(\tau) &= \int_{-\infty}^{\infty} f(u)f(u+\tau)\,du \qquad |\tau|<t_0, \\ &= 0 \qquad\qquad\qquad\qquad\qquad |\tau|\geqq t_0. \end{aligned} \tag{2.3,37}$$

§ 2.4 Relation between continuous parameter process and discrete parameter process belongs to it.

Definition. *Let $x(t)\,(-\infty<t<\infty)$ be a stochastic process with continuous time parameter. Then a discrete parameter process $x(s+n\Delta t)$ $(\Delta t>0, n=0, \pm 1, \pm 2, \cdots)$ may be called a process belongs to the $x(t)$ for each s real.*

Lemma 5. *Let $F(\lambda)\,(-\infty<\lambda<\infty)$ be the spectral distribution function of a real stationary process $x(t)$ and let $G_{\Delta t}(\mu)\,(-\pi/\Delta t\leqq\mu\leqq\pi/\Delta t)$ be the spectral distribution function of $x(n\Delta t)$ belongs to the $x(t)$. Then*

$$G_{\Delta t}(\mu)=\sum_{n=-\infty}^{\infty}\Big[F\Big(\frac{2n\pi}{\Delta t}+\mu\Big)-F\Big(\frac{(2n-1)\pi}{\Delta t}\Big)\Big] \tag{2.4.1}$$

and if the spectral density function $F'(\lambda)$ exists and is continuous, then $G'_{\Delta t}(\mu)$ exists and

$$G'_{\Delta t}(\mu)=\sum_{n=-\infty}^{\infty} F'\Big(\frac{2n\pi}{\Delta t}+\mu\Big)\,, \tag{2.4.2}$$

provided that the right hand side converges.

Proof. Let $\rho(\tau)$ be the autocorrelation function of $x(t)$. Then

$$\rho(\nu\Delta t) = \int_{-\infty}^{\infty} \cos\nu\Delta t\lambda dF(\lambda)$$
$$= \sum_{n=-\infty}^{\infty} \int_{(2n-1)\pi/\Delta t}^{(2n+1)\pi/\Delta t} \cos\nu\Delta t\lambda dF(\lambda)$$
$$= \sum_{n=-\infty}^{\infty} \int_{-\pi/\Delta t}^{\pi/\Delta t} \cos\nu\mu dF\left(\frac{2n\pi}{\Delta t} + \mu\right).$$

As this series is absolutely convergent, the proposition is proved.

Theorem 8. *Let $x(t)$ $(-\infty < t < \infty)$ be a continuous parameter stationary process having spectral density function*

$$F'(\lambda) = \left|\sum_{j=0}^{k} b_j(i\lambda)^j\right|^2 \Big/ \left|\sum_{j=0}^{h} a_j(i\lambda)^j\right|^2, \quad k < h, \quad a_0 b_0 a_h b_k \neq 0, \qquad (2.4.3)$$

with the conditions about the coefficients a's and b's for (2.3.33). *Then, every discrete parameter process $x(n\Delta t)$ $(n = 0, \pm 1, \pm 2, \cdots)$ belongs to the $x(t)$ is a generalized autoregression process of orders $(h, h-1)$, independently of k, except possible degenerated case.*

Proof. A perspective proof is as follows. The autocorrelation function of the $x(t)$ process has the form

$$\rho(\tau) = \sum_j c_j e^{i\tau z_j} \qquad \tau \geqq 0 \qquad (\rho(-\tau) = \rho(\tau)),$$

where c_j is a polynomial in τ with coefficients $c_j^{(\nu)}$ $(\nu = 1, 2, \cdots, h_j;\ \sum_j h_j = h)$ and the z_j's are zeros of $\sum a_j z^j = 0$, with the multiplicity h_j's respectively. Hence,

$$\rho((n-j)\Delta t) = \sum_j c_j e^{i(n-j)\Delta t z_j}, \quad j = 0, 1, \cdots, h;\ n \geqq h. \qquad (2.4.4)$$

Eliminating h coefficients $c_j^{(\nu)}$'s from $h+1$ equations (2.4.4), we get

$$\sum_{j=0}^{h} a^*(\Delta t)\rho((n-\nu)\Delta t) = 0 \quad \text{for} \quad n \geqq h$$
$$\neq 0 \quad \text{for} \quad n = h-1, \qquad (2.4.5)$$

where $a_j^*(\Delta t)$ is a function of $\exp(z_j\Delta t)$ $(j = 1, 2, \cdots)$. (2.4.5) is nothing but the characteristic property for $x(n\Delta t)$ to be a g.a.p. of orders $(h, h-1)$.

An alternative proof of the theorem is to show that the spectral densities of the $x(n\Delta t)$ process has the form

$$G'_{\Delta t}(\mu) = c\frac{\left|\sum_{j=0}^{h-1} b_j^* e^{i(h-1-j)\mu\Delta t}\right|^2}{\left|\sum_{j=0}^{h} a_j^* e^{i(h-j)\mu\Delta t}\right|^2}, \quad -\frac{\pi}{\Delta t} \leqq \mu \leqq \frac{\pi}{\Delta t}, \ c > 0. \qquad (2.4.6)$$

To this we need the following lemma, which is derived from the formula in the theory of infinite series,

$$\frac{1+e^{-z}}{2z(1-e^{-z})}=\sum_{n=-\infty}^{\infty}\frac{1}{z^2+(2n\pi)^2}\qquad (z\neq 0)\ .$$

Lemma 6.

$$\sum_{n=-\infty}^{\infty}\frac{1}{\zeta^2+(2n\pi+\mu)^2}=\frac{1-e^{-2\zeta}}{2\zeta(1-e^{-(\zeta+i\mu)})(1-e^{-(\zeta-i\mu)})}\,,$$

where ζ and μ are arbitrary constants.

By Lemmas 5 and 6, after some calculations, we get

$$G'_{\Delta t}(\mu)=\sum_{j=0}^{h-1}P_j(e^{ij\mu\Delta t}+e^{-ij\mu\Delta t})\Big/\Big|\sum_{j=0}^{h}a_j^*\,e^{i(h-j)\mu\Delta t}\Big|^2\,,$$

where the constants a_j^*'s are functions of $\exp(-z_j\Delta t)$ ($j=1,2,\cdots$; z_j's being constants depend on a_j's), the P_j's depend on constants z_j's and w_j's (which depend on b_j's) and both are independent of μ. Since $G'_{\Delta t}(\mu)\geqq 0(-\pi/\Delta t\leqq\mu\leqq\pi/\Delta t)$, the $G'_{\Delta t}(\mu)$ can be written in the form (2.4.6), owing to Lemma 3.

Conversely, we observe the following

Theorem 9. *Let $x(n\tau)$ ($\tau>0$; $n=0,\pm 1,\pm 2,\cdots$) be a discrete parameter process belongs to a real stationary continuous parameter process $x(t)$ ($-\infty<t<\infty$) with continuous spectral density function $F'(\lambda)$. If then for any $\tau>0$*

$$y(n\tau)=\sum_{j=0}^{h}\alpha_j(\tau)\,x((n-j)\tau)\,,\qquad \alpha_0(\tau)\equiv 1\,,\qquad n=0,\pm 1,\pm 2,\cdots \tag{2.4.7}$$

is a moving averages of order $m\geqq 0$, where $\alpha_j(\tau)$ ($j=1,2,\cdots,h$) are real valued continuous functions of τ only and $\alpha_h(\tau)\neq 0$ for any $\tau>0$. Then $m=h-1$ and $F'(\lambda)$ has the form (2.4.3).

Proof. According to Ghurye's theorem (Ghurye[28]), under the assumptions of our theorem, the autocorrelation coefficient $\rho(\tau)$ of $x(t)$ has the form

$$\rho(\tau)=\sum_j\Big(\sum_{s=0}^{h_j-1}c_{js}\tau^s\Big)e^{-\lambda_j\tau}\,,\qquad \tau>0\,,\qquad \sum_j h_j=h\,, \tag{2.4.8}$$

where c_{js}'s are constants independent of τ and the real parts of the λ_j are positive. Therefore the $\rho(\tau)$ satisfies a differential equation

$$\sum_{j=0}^{h}a_j\rho^{(j)}(\tau)=0\qquad (\tau>0)\ .$$

Thus, owing to Doob's theorem (Doob [25]), $F'(\lambda)$ has the form (2.4.3) and, by Theorem 8, $m=h-1$.

Chapter III. Statistical Inference of Time Series

§ **3.1 Autoregression process.** When the autocorrelation coefficients ρ_k $(k=1, 2, \cdots)$ of a stationary time series x_t with integral time parameter are known, the statistical inference of the mean value and the variance is easy. For example, for a normal simple Markov process with autocorrelation coefficients $\rho_k = \rho^k$ $(\rho < 1)$, the maximum likelihood estimates for the mean value m and the variance σ^2 are given by

$$\hat{m} = \Big[x_1 + x_N + (1-\rho)\sum_{t=2}^{N-1} x_t\Big] \Big[2 + (1-\rho)(N-2)\Big]$$

$$\hat{\sigma}^2 = \frac{1}{N}\Big[(x_1 - \hat{m})^2 + \frac{1}{1-\rho^2}\sum_{t=2}^{N}\Big\{(x_t - \hat{m}) - \rho(x_{t-1} - \hat{m})\Big\}^2\Big].$$

The $\hat{m}$ is the unbiassed efficient statistic for m, while, for $\bar{x} = \sum_{t=1}^{N} x_t/N$, we have

$$V(\bar{x}) = \frac{\sigma^2}{N}\Big[\frac{1+\rho}{1-\rho} - \frac{2\rho(1-\rho^N)}{N(1-\rho)^2}\Big] > V(\hat{m}),$$

provided that $\rho \neq 0$.

In general case where all of the population parameters are unknown, our methods of statistical inference are based on the conditional independence and the normal regression theory. The conditional independence is introduced in two ways according to two types of sample schemes given in § 1.1.

Sample scheme I. Let x_t $(t=1, 2, \cdots, N)$ be a sample of type I of a normal stationary multiple Markov process of order h, $x(t)$ $(t \in T)$, where $T = \{\cdots, -1, 0, 1, \cdots\}$, and

$$x(t) - m + a_1(x(t-1) - m) + \cdots + a_h(x(t-h) - m) = y(t), \quad a_h \neq 0. \tag{3.1.1}$$

If the variables $x(k(h+1) - p)$, $x(k(h+1) + p)$ $(p=1, 2, \cdots, h;\ k=1, 2, \cdots, n)$ are fixed at their observed values, the random variables $x(k(h+1))$ $(k=1, 2, \cdots, n)$ are independent each other by Theorem 2 and their conditional probability densities are given by

$$\begin{aligned} &f(x_{k(h+1)} \mid x_{k(h+1)-p}, x_{k(h+1)+p};\ p=1, 2, \cdots, h) \\ &= \frac{1}{\sqrt{2\pi}\,\sigma_0} \exp\Big[-\frac{1}{2\sigma_0^2}\Big\{x_{k(h+1)} - \sum_{p=0}^{h} b_p x'_{pk}\Big\}^2\Big] \quad (k=1, 2, \cdots, n), \end{aligned} \tag{3.1.2}$$

where $x'_{pk} = (x_{k(h+1)-p} + x_{k(h+1)+p})/2$ $(p=1, 2, \cdots, h)$, $x'_{0k} = 1$, and where

$$b_0 = m\Big(1 - 2\sum_{p=1}^{h} c_p\Big), \quad b_p = 2c_p \quad (p=1, 2, \cdots, h), \tag{3.1.3}$$

$$\begin{bmatrix} c_h \\ c_{h-1} \\ \vdots \\ c_1 \\ c_1 \\ \vdots \\ c_h \end{bmatrix} = \begin{bmatrix} 1 & \cdots \rho_{h-1} & \rho_{h+1} & \cdots \rho_{2h} \\ \rho_1 & \cdots \rho_{h-2} & \rho_h & \cdots \rho_{2h-1} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{h-1} & \cdots 1 & \rho_2 & \cdots \rho_{h+1} \\ \rho_{h+1} & \cdots \rho_2 & 1 & \cdots \rho_{h-1} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{2h} & \cdots \rho_{h+1} & \rho_{h-1} & \cdots 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_h \\ \rho_{h-1} \\ \vdots \\ \rho_1 \\ \rho_1 \\ \vdots \\ \rho_h \end{bmatrix} \tag{3.1.4}$$

and

$$\sigma_0^2 = \frac{1 + a_1\rho_1 + \cdots + a_h\rho_h}{1 + a_1^2 + \cdots + a_h^2}\sigma^2 , \tag{3.1.5}$$

where

$$\rho_k + a_1\rho_{k-1} + \cdots + a_h\rho_{k-h} = 0 \qquad (k = 1, 2, \cdots) . \tag{3.1.6}$$

Let $\hat{b}_k$ $(k = 0, 1, \cdots, h)$ and $\hat{\sigma}_0^2$ be the conditional maximum likelihood estimates of b_k $(k = 0, 1, \cdots, h)$ and σ_0^2 respectively. Then, the conditional maximum likelihood estimates $\hat{m}$ and $\hat{c}_p$ $(p = 1, 2, \cdots, h)$ are obtained from (3.1.3), and the estimates of the autocorrelation coefficients are found, if they exist, from the equations

$$\hat{c}_h\hat{\rho}_{h+j} + \hat{c}_{h-1}\hat{\rho}_{h+j-1} + \cdots + \hat{c}_1\hat{\rho}_{j+1} - \hat{\rho}_j + \hat{c}_1\hat{\rho}_{j-1} + \cdots + \hat{c}_h\hat{\rho}_{j-h} = 0 \qquad (j = 1, 2, \cdots, h) \tag{3.1.7}$$

and

$$\hat{\rho}_k + \hat{a}_1\hat{\rho}_{k-1} + \cdots + \hat{a}_h\hat{\rho}_{k-h} = 0 \qquad (k = 1, 2, \cdots, 2h) , \tag{3.1.8}$$

under the condition (2.2.4), where $\hat{\rho}_1, \hat{\rho}_2, \cdots, \hat{\rho}_{2h}$ and $\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_h$ are unknowns. Furthermore, as an estimate of the spectral densities $f(\lambda)$ we have

$$\hat{f}(\lambda) = K\Big/\left|1 + \hat{a}_1 e^{i\lambda} + \cdots + \hat{a}_h e^{ih\lambda}\right|^2 , \tag{3.1.9}$$

where $K = [2\pi \sum_{j=0}^{\infty} \alpha_j^2]^{-1}$, the α_j being the solution of the equations

$$\sum_{j=0}^{k} \hat{a}_{k-j}\alpha_j = 0 \quad (k = 1, 2, \cdots) , \qquad \hat{a}_0 = \alpha_0 = 1 , \qquad \alpha_j = 0 \quad (j < 0) .$$

The consistency of the conditional maximum likelihood estimates $\hat{b}_k$, $\hat{\sigma}_0$, consequently the consistency of the derived estimates for the other parameters and the spectral densities can be easily proved, where the x'_{pk} together with $x_{k(h+1)}$ are random variables.

As we see in the following theorems, the efficiency of such conditional estimations is inferior to that of ordinary maximum likelihood estimates, though it is generally difficult to obatin explicit expressions for the later. Above mentioned conditional estimates would be rather useful for testing statistical hypotheses which will be given later on.

Theorem 10. *Let $f(x_1, \cdots, x_n, x_1', \cdots, x_m'; \theta, \theta', \cdots)$ be the probability density function of $m+n$ random variables with population parameters $\theta, \theta', \cdots$, and let $g(x_1', \cdots, x_m'; \theta, \theta', \cdots)$ be the probability density function of $x'=(x_1', \cdots, x_m')$. If $\hat{\theta}$ is the efficient estimate of θ and if $\hat{\theta}_c$ is a conditional estimate of θ given x', then*

$$V(\hat{\theta}) \leqq \int V_c(\hat{\theta}_c)\, g\, dx',$$

where V_c denotes the conditional variance given x', and the equality holds if and only if i) *g is independent of θ,* ii) *$\hat{\theta}_c$ is the conditional efficient estimate and* iii) *$V_c(\hat{\theta}_c)$ is independent of x'.*

Proof. If we write the conditional probability density function of $x=(x_1, \cdots, x_n)$ given x' as $h(x; \theta, \theta', \cdots | x')$, then $f=gh$ and, since

$$\int \frac{\partial \log f}{\partial \theta} f\, dx\, dx' = \int \frac{\partial \log h}{\partial \theta} h\, dx = 0,$$

we have

$$E\Big(\frac{\partial \log f}{\partial \theta}\Big)^2 = E\Big(\frac{\partial \log g}{\partial \theta}\Big)^2 + \int E_c\Big(\frac{\partial \log h}{\partial \theta}\Big)^2 g\, dx',$$

where E_c stands for the conditional expectation given x'. Therefore

$$\begin{aligned} V(\hat{\theta}) &= 1\Big/ E\Big(\frac{\partial \log f}{\partial \theta}\Big)^2 \\ &\leqq 1\Big/\Big[E\Big(\frac{\partial \log g}{\partial \theta}\Big)^2 + \int \frac{1}{V_c(\hat{\theta}_c)} g\, dx'\Big] \\ &\leqq 1\Big/\Big[\int \frac{1}{V_c(\hat{\theta}_c)} g\, dx'\Big] \leqq \int V_c(\hat{\theta}_c)\, g\, dx'. \end{aligned}$$

Similarly we have the following

Theorem 11. *Let $(\hat{\theta}, \hat{\theta}')$ be the joint efficient estimate of (θ, θ') based on the sample (x, x'), where $x=(x_1, \cdots, x_n)$ and $x'=(x_1', \cdots, x_m')$, and let $(\hat{\theta}_c, \hat{\theta}_c')$ be a conditional joint estimate of (θ, θ') by (x, x') when the value of x' is given. Then the mean concentration ellipse* of $(\hat{\theta}_c, \hat{\theta}_c')$ with respect to x' includes the concentration ellipse of $(\hat{\theta}, \hat{\theta}')$ in it. If $(\hat{\theta}, \hat{\theta}')$ is asymptotically efficient the relation mentioned above also asymptotically holds.*

Now, let $H(\theta_1, \theta_2, \cdots, \theta_l)$ denote a hypothesis specifing the values of parameters $\theta_1, \theta_2, \cdots, \theta_l$ and let $H_1 \sim H_2$ denote the equivalency of two hypotheses H_1 and H_2. Then, by (2.2.5), (3.1.3) and (3.1.4),

$$\begin{aligned} H(\rho_1, \rho_2, \cdots) &\sim H(a_1, a_2, \cdots, a_h) \\ &\sim H(c_1, c_2, \cdots, c_h) \sim H(b_1, b_2, \cdots, b_h) \end{aligned} \tag{3.1.10}$$

* H. Cramèr [29].

and

$$H(m, \rho_1, \rho_2, \cdots) \sim H(m, c_1, \cdots, c_h) \sim H(b_0, b_1, \cdots, b_h)\,. \quad (3.1.11)$$

Thus these hypotheses can be tested by the normal regression theory (Ogawara [4]) :

Theorem 12.* *For the test of hypothetical correlogram $H(\rho_1, \rho_2, \cdots)$ of an autoregression process of order h we may use the following test function which is distributed according to the F distribution with h and $n-h-1$ degrees of freedom.*

$$F^h_{n-h-1} = \frac{\sum_{p,q=1}^{h} a_{pq}(\bar{b}_p - b_p)(\bar{b}_q - b_q)}{\sum_{j=1}^{n}\left(z_j - \sum_{p=1}^{h} \bar{b}_p z'_{pj}\right)^2} \cdot \frac{n-h-1}{h}\,, \quad (3.1.12)$$

where

$$z_j = x_{j(h+1)} - \sum_{j=1}^{n} x_{j(h+1)}/n \qquad (j = 1, 2, \cdots, n)$$

are random variables and

$$z'_{pj} = x'_{pj} - \sum_{j=1}^{n} x'_{pj}/n\,, \qquad x'_{pj} = (x_{j(h+1)-p} + x_{j(h+1)+p})/2 \quad \begin{pmatrix} p = 1, 2, \cdots, h \\ j = 1, 2, \cdots, n \end{pmatrix}$$

are fixed variates,

$$a_{pq} = a_{qp} = \sum_{j=1}^{n} z'_{pj} z'_{qj}$$

and $\bar{b}_q$ is the solution of the simultaneous equations

$$\sum_{q=1}^{h} a_{pq}\bar{b}_q = a_{p0} \qquad (p = 1, 2, \cdots, h)$$

where

$$a_{p0} = \sum_{j=1}^{n} z'_{pj} z_j\,.$$

If we write $a_{00} = \sum_{j=1}^{n} z_j^2$, the denominator of (3.1.12) is written as

$$a_{00} - \sum_{p=1}^{h} a_{p0}\bar{b}_p\,.$$

Further, we can estimate the order of the Markov process. Since the above stated theory holds whenever the essential order h_0 of the process is not larger than h, if the hypothesis $b_{h_1+1} = b_{h_1+2} = \cdots = b_h = 0$ is not rejected, then we may suppose that $h_0 \leq h_1$ on the assigned level of significance.

More generally, we can test the correlogram through the test of regression coefficients of a linear combination of $x_{j(h+s)+i-1}$ $(i = 1, 2, \cdots, \nu\,;\ \nu \leq s)$, such as $\bar{x}_j = \sum_{i=1}^{\nu} x_{j(h+s)+i-1}/\nu$, on $x_{j(h+s)-k}$ and $x_{j(h+s)+s+k-1}$ $(k = 1, 2, \cdots, h\,;\ j = 1, 2, \cdots, n)$.

* This theorem cannot be extended to the generalized autoregression process.

By way of an illustration, let us consider the case of stationary normal simple Markov process for which $h=1$ and $\rho(\tau)=\rho^{\tau}$ ($|\rho|<1$). If we set $s=\nu\geqq 1$, the conditional distribution of

$$\bar{x}_j=\sum_{i=1}^{\nu} x_{j(\nu+1)+i-1}/\nu \qquad (j=1, 2, \cdots, n),$$

given $x_{j(\nu+1)-1}$ and $x_{j(\nu+1)+\nu}$, is normal with the conditional mean value

$$E\{\bar{x}_j \mid x_{j(\nu+1)-1},\ x_{j(\nu+1)+\nu}\}=a+bx_j' \qquad (j=1, 2, \cdots, n),$$

where

$$\begin{aligned} a &= (1-b)\,m \\ b &= \frac{2\rho(1-\rho^{\nu})}{\nu(1-\rho)(1+\rho^{\nu+1})} \\ x_j' &= (x_{j(\nu+1)-1}+x_{j(\nu+1)+\nu})/2 \end{aligned} \tag{3.1.13}$$

and the conditional variance

$$\begin{aligned} \sigma_{\nu}^2 &= V\{\bar{x}_j \mid x_{j(\nu+1)-1},\ x_{j(\nu+1)+\nu}\} \\ &= \frac{\sigma^2}{1-\rho^{2(\nu+1)}}\Big[\frac{1+\rho^{2\nu+2}}{\nu}+\frac{2(\nu-1)(\rho-\rho^{2\nu+2})-2(\rho^2-\rho^{2\nu+1})}{\nu^2(1-\rho)} \\ &\qquad -\frac{2(1+\rho)(\rho^2-\rho^{\nu})(1-\rho^{\nu-1})}{\nu^2(1-\rho)^2}\Big], \end{aligned}$$

and the conditional random variables x_j ($j=1, 2, \cdots, n$) are mutually independent.

Since $|\rho|<1$, from (3.1.13), we have

$$\begin{aligned} &\text{if } \nu \text{ is even} \qquad -1/(\nu-1)<b<1, \\ &\text{if } \nu \text{ is odd} \qquad -1/\nu<b<1. \end{aligned}$$

Let us denote this region of b by D_{ν}. The conditional maximum likelihood estimate of a, b and σ_{ν}^2 is given by

$$\begin{aligned} \hat{a} &= \bar{x}^*-\hat{b}\,\bar{x}', \\ \hat{b} &\begin{cases} =\bar{b} & \text{if } \bar{b}\in D_{\nu} \\ \text{does not exist} & \text{if } \bar{b}\notin D_{\nu} \end{cases} \\ \hat{\sigma}_{\nu}^2 &= \sum_{j=1}^{n}(\bar{x}_j-\hat{a}-\hat{b}x_j')^2/n \end{aligned}$$

respectively, where

$$\bar{x}^*=\sum_{j=1}^{n}\bar{x}_j/n, \qquad \bar{x}'=\sum_{j=1}^{n}x_j'/n,$$

$$\bar{b}=\sum_{j=1}^{n}(x_j'-\bar{x}')(\bar{x}_j-\bar{x}^*)\Big/\sum_{j=1}^{n}(x_j'-\bar{x}')^2,$$

and

$$F^{1}_{n-2} = \frac{(\hat{b}-b)^2 \sum_{j=1}^{n} (x_j' - \bar{x}')^2 \cdot (n-2)}{\sum_{j=1}^{n} (\bar{x}_j - \bar{a} - \bar{b}x_j')^2} \tag{3.1.14}$$

has the F distribution with the pair of degrees of freedom $(1, n-2)$, where $\bar{a} = \bar{x}^* - \bar{b}a'$ and x_j' are fixed variates, which enables us to test the hypothesis $H(b)$ or $H(\rho)$.

It will be desirable to take the value of ν which minimizes the length of confidence interval for ρ, Generally speaking, comparatively large ν is preferable if ρ is near to one, relatively small ν is better if ρ is positive and small and we should take $\nu = 1$ if $\rho < 0$. (Ogawara [7])

Let two time series $x_t^{(1)}$ and $x_t^{(2)}$ has the mean value m_1 and m_2, the variance σ_1^2 and σ_2^2 and the autocorrelation coefficients $\rho_1(\tau) = \rho_1^{|\tau|}$ and $\rho_2(\tau) = \rho_2^{|\tau|}$ respectively. If we define $\bar{x}_j^{(1)} = \sum_{k=1}^{\nu_1} x_{j(\nu_1+1)+k-1}^{(1)}/\nu_1$ $(j = 1, 2, \cdots, n_1)$, $\bar{x}_j^{(2)} = \sum_{k=1}^{\nu_2} x_{j(\nu_2+1)+k-1}^{(2)}/\nu_2$ $(j = 1, 2, \cdots, n_2)$ and other quantities in the same manner as before,

$$F^{n_1-2}_{n_2-2} = \frac{\sum_{j=1}^{n_1} (\bar{x}_j^{(1)} - \bar{a}^{(1)} - \bar{b}^{(1)} x_j^{(1)\prime})^2 \cdot (n_2-2)}{\sum_{j=1}^{n_2} (\bar{x}_j^{(2)} - \bar{a}^{(2)} - \bar{b}^{(2)} x_j^{(2)\prime})^2 \cdot (n_1-2)} \cdot \frac{\sigma_2^2}{\sigma_1^2} \cdot \frac{\varphi_2(\rho_2)}{\varphi_1(\rho_1)} \tag{3.1.15}$$

is distributed as the F distribution with degrees of freedom (n_1-2, n_2-2), where

$$\varphi_i(\rho_i) = \frac{1}{1-\rho_i^{2\nu_i+2}} \Big[\frac{1+\rho_i^{2\nu_i+2}}{\nu_i} + \frac{2(\nu_i-1)(\rho_i - \rho_i^{2\nu_i+2}) - 2(\rho_i^2 - \rho_i^{2\nu_i+1})}{\nu_i^2(1-\rho_i)} - \frac{2(1+\rho_i)(\rho_i^2 - \rho_i^{\nu_i})(1-\rho_i^{\nu_i-1})}{\nu_i^2(1-\rho_i)^2} \Big] \quad (i = 1, 2).$$

Thus, if σ_2^2/σ_1^2 is known the hypothesis H_1: $\varphi_2(\rho_2)/\varphi_1(\rho_1) = \theta$ $(-\infty < m_1, m_2 < \infty)$ is tested by (3.1.15), and if $\varphi_2(\rho_2)/\varphi_1(\rho_1)$ is known the hypothesis H_2: $\sigma_2^2/\sigma_1^2 = \theta$ $(-\infty < m_1, m_2 < \infty)$ is tested by (3.1.15). For the test of H_1, ν_1 and ν_2 should be so selected that φ_2/φ_1 is even sensitive for a small difference between ρ_1 and ρ_2.

Sample scheme II. First we observe the following

Theorem 13. *An autoregression process $x(t)$ $(t = 0, \pm 1, \cdots)$ of oder h, with zero mean, satisfies almost surely the stochastic finite difference equation*

$$x(t+s) + b_1 x(t-1) + \cdots + b_h x(t-h) = z(t) \qquad (s \geqq 0, b_h \neq 0), \tag{3.1.16}$$

where the $z(t)$ is a moving averages of order s and is orthogonal to $x(t-\tau)$ $(\tau = 1, 2, \cdots)$, and the autocorrelation coefficients satisfy the finite difference equation

$$\rho(\tau+s)+b_1\rho(\tau-1)+\cdots+b_h\rho(\tau-h)=0 \qquad (\tau=1,2,\cdots)\,. \tag{3.1.17}$$

The coefficients in (3.1,16) *and* (3.1.17) *have the following relation to the coefficients in* (3.1.1) *and to the autocorrelation coefficients.*

$$\textit{When } s=0 \qquad b_k=a_k \ \ (k=1,2,\cdots,h) \tag{3.1.18}$$

$$\textit{When } s\geqq 1 \qquad b_k=\begin{vmatrix} a_1 & a_2 & \cdots & a_s & a_{s+k} \\ 1 & a_1 & \cdots & a_{s-1} & a_{s-1+k} \\ 0 & 1 & \cdots & a_{s-2} & a_{s-2+k} \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & a_1 & a_{1+k} \\ 0 & 0 & \cdots & 1 & a_k \end{vmatrix} \qquad \begin{pmatrix} k=1,2,\cdots,h \\ a_j=0 \ \ for \ \ j>h \end{pmatrix} \tag{3.1.19}$$

or

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_h \end{bmatrix} = -\begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{h-1} \\ \rho_1 & 1 & \cdots & \rho_{h-2} \\ \vdots & \vdots & & \vdots \\ \rho_{h-1} & \rho_{h-2} & \cdots & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_{s+1} \\ \rho_{s+2} \\ \vdots \\ \rho_{s+h} \end{bmatrix}, \tag{3.1.20}$$

where $\rho_\tau \equiv \rho(\tau)$.

Proof. Eliminating $\rho(\tau+s-\tau')$ $(\tau'=1,2,\cdots,s)$ from the $s+1$ equations

$$\rho(\tau+s-\tau')+a_1\rho(\tau+s-\tau'-1)+\cdots+a_h\rho(\tau+s-\tau'-h)=0$$
$$(\tau'=0,1,\cdots,s)$$

we get (3.1.17) from which we can immediately deduce (3.1.16).

Now, let $x(t)$ be a stationary normal mutiple Markov process of order h satisfing (3.1.1). For a given time series x_t $(t=1,2,\cdots,N)$, we suppose that

$$\{x_{(j-1)(h+s+1)+k}\,;\ k=1,2,\cdots,h+s+1\} \qquad (s\geqq 0)$$

is a sample of a partial process $x(t,\omega_j)$, $t\in T_j\equiv\{(j-1)(h+s+1)+k\,;$ $k=1,2,\cdots,h+s+1\}$, $\omega_j\in R^{T_j}(P_{T_j})$ $(j=1,2,\cdots,n)$. Such n samples are not independent, but, by Theorem 2, if we fix the values $x_{(j-1)(h+s+1)+k}$ $(k=1,2,$ $\cdots,h\,;\ j=1,2,\cdots,n)$, $x_{j(h+s+1)}$ $(j=1,2,\cdots,n)$ are conditionally independent and the conditional probability densities of $x_{j(h+s+1)}$ is given by

$$f(x_{j(h+s+1)}\mid x_{j(h+s+1)-s-k}\,;\ k=1,2,\cdots,h)$$
$$=\frac{1}{\sqrt{2\pi}\,\sigma_1}\exp\left[-\frac{1}{2\sigma_1^2}\left\{x_{j(h+s+1)}+b_0+\sum_{k=1}^{h}b_k x_{j(h+s+1)-s-k}\right\}^2\right], \tag{3.1.21}$$

where $b_0=-m(1+b_1+\cdots+b_h)$, the coefficients b_k are given by (3.1.20) and

$$\sigma_1^2=\sigma^2(1+b_1\rho_{s+1}+b_2\rho_{s+2}+\cdots+b_h\rho_{s+h})\,.$$

The conditional estimation of parameters in this case is quite similar to the case of Sample scheme I and if we set

$$z_j = x_{j(h+s+1)} - \sum_{i=1}^{n} x_{i(h+s+1)}/n ,$$
$$z'_{pj} = x_{j(h+s+1)-s-p} - \sum_{i=1}^{n} x_{i(h+s+1)-s-p}/n$$

the test function (3.1.12) is also used for testing correlogram. It will be a matter of course that we have different tests according to the values of non-negative integer s.

Our methods mentioned above are easily extended to multidimensional autoregression processes. Let $x(t)$ $(-\infty < t < \infty)$ be a stochastic process with continuous time parameter and suppose that the observation is made at a series of time points t_i $(i = 0, \pm 1, \pm 2, \cdots)$ which constitute a stochastic process. Such process can be treated as a two dimensional process (x_i, y_i) with integral time parameter i, whose components are $x_i = x(t_i)$ and $y_i = t_i - t_{i-1}$ and we sometimes meet with practical problems of this form which is supposed to be a two dimensional multiple Markov process.

§ 3.2 Powers of conditional tests. Let $T(x, x'; \theta)$ be a conditional test function for a set of parameters $\theta = (\theta_1, \theta_2, \cdots, \theta_l)$, where $x = (x_1, \cdots, x_n)$ is a set of random variables and $x' = (x'_1, \cdots, x'_m)$ is a set of condition variables. Let the conditional distribution densities* of x be $h(x|x'; \theta)$, let the joint distribution densities of (x, x') be $f(x, x'; \theta)$ and let the density function of x' only be $g(x'; \theta)$. Then a critical region of x, $w(x'; \alpha)$, for the conditional test of assigned θ value, with a significance level α, is given by

$$\Pr\{T(x, x'; \theta) \in S|x'; \theta\} = \int_{w(x', \alpha)} h(x|x'; \theta)\, dx = \alpha , \qquad (3.2.1)$$

where S is a critical region of T. Therefore

$$\int \left\{ \int_{w(x', \alpha)} h(x|x'; \theta)\, dx \right\} g(x'; \theta)\, dx' = \alpha$$

or

$$\int_W f(x, x'; \theta)\, dx dx' = \alpha , \qquad (3.2.2)$$

where W is the set of (x, x') such that $x \in w(x'; \alpha)$ holds.

On the other hand, the conditional power function for the region $w(x'; \alpha)$ is given by

$$P(\theta_1; x') = 1 - \beta(\theta_1; x') = \int_{w(x'; \alpha)} h(x|x'; \theta_1)\, dx , \qquad (3.2.3)$$

where θ_1 is a counter hypothesis, and the *mean power function* is given by

$$P(\theta_1) = 1 - \beta(\theta_1) = \int_W f(x, x'; \theta_1)\, dx dx' . \qquad (3.2.4)$$

---

* Discrete case and more general cases are quite similarly discussed.

We consider in detail the case of testing autocorrelation for a stationary normal simple Markov process. Let H_0: $\rho=\rho_0$ (or $b=b_0$) be the null hypothesis and H_1: $\rho=\rho_1$ (or $b=b_1$) be the alternative hypothesis.

Firstly, for the sample scheme I, if we take $\nu=1$ in (3.1.13) ~ (3.1.14), the conditional second type error is given by

$$\beta\equiv\beta(x')=\sum_{i=0}^{\infty}\frac{\lambda^i e^{-\lambda}}{i\,!}I_{c(\alpha)}\Big(\frac{f_1}{2}+i,\ \frac{f_2}{2}\Big), \tag{3.2.5}$$

where $I_c(p,q)=B_c(p,q)/B(p,q)$, $B_c(p,q)$ stands for the incomplete beta function, $f_1=1$, $f_2=n-2$, $I_{c(\alpha)}(f_1/2, f_2/2)=\alpha$, and where

$$\begin{aligned}\lambda\ &=(b_1-b_0)^2\sum_{j=1}^{n}(x_j'-x')^2/2\sigma_1^2\,,\\ b_i\ &=2\rho_i/(1+\rho_i^2)\qquad(i=0,1)\,,\\ x_j'&=(x_{2j-1}+x_{2j+1})/2\,,\qquad \bar{x}'=\sum_{j=1}^{n}x_j'/n\,,\\ \sigma_1^2&=\sigma^2(1-\rho_i^2)/(1+\rho_i^2)\,.\end{aligned} \tag{3.2.6}$$

If we set

$$A_i=\underset{x'}{E}(\lambda^i e^{-\lambda})\,,\qquad B_i=\frac{1}{i!}I_{c(\alpha)}\Big(\frac{1}{2}+i,\ \frac{n-2}{2}\Big), \tag{3.2.7}$$

$$\underset{x'}{E}(\beta)=\sum_{i=0}^{\infty}\mathrm{A}_iB_i \tag{3.2.8}$$

and, if we moreover put

$$\begin{aligned}C_i&=\underset{x'}{E}(\lambda^i e^{-2\lambda})\,,\\ \underset{x'}{E}(\beta^2)&=\sum_{k=0}^{\infty}\Big(\sum_{i+j=k}B_iB_j\Big)C_k\end{aligned} \tag{3.2.9}$$

and the variance of $\beta(x')$ is given by

$$V(\beta)=\underset{x'}{E}(\beta^2)-\{\underset{x'}{E}(\beta)\}^2. \tag{3.2.10}$$

Now, the variance matrix of x' is given by

$$V=\sigma^2\begin{bmatrix}1 & (1+\rho^2)/2 & (1+\rho^2)\rho^2/2\cdots & (1+\rho^2)\rho^{2n-4}/2\\ (1+\rho^2)/2 & 1 & (1+\rho^2)/2 & \cdots(1+\rho^2)\rho^{2n-6}/2\\ \vdots & \vdots & \vdots & \vdots\\ (1+\rho^2)\rho^{2n-4}/2\cdots & \cdots & \cdots & 1\end{bmatrix}$$

and the cumulants $\kappa_r(q)$ of the quadratic form $q=\sum_{j=1}^{n}(x_j'-\bar{x}')^2$ is equal to the trace of the matrix $(VQ)^r$, Q being the matrix of the quadratic form q. Thus, if we denote the characteristic roots of $VQ\equiv(a_{ij})$ by λ_j $(j=1,2,\cdots,n)$, we get

$$\kappa_r(q)=2^{r-1}(r-1)\sum_j\lambda_j^r\,,$$

where $\sum_j \lambda_j = \sum_i a_{ii}$, $\sum_j \lambda_j^2 = \sum_i \sum_j a_{ij} a_{ji}$ and so on. By these formulas we have

$$\kappa_1 \equiv E(\lambda) = \frac{(b_1 - b_0)^2 (1+\rho^2)^2}{4(1-\rho^2)} \left[n - \frac{2}{1-\rho^2} + \frac{(1+\rho^2)(1-\rho^{2n})}{n(1-\rho^2)^2} \right] \quad (3.2.11)$$

$$\begin{aligned} \kappa_2 \equiv V(\lambda) = \frac{(b_1 - b_0)^4 (1+\rho^2)^4}{8(1-\rho^2)} \Big[& \frac{n(3-\rho^2)}{2(1-\rho^2)} \\ & - \frac{1}{2(1-\rho^2)^2} \{ 9 + 2(1+\rho^2)^2 \rho^{2(n-1)} - \rho^{4n} \} \\ & + \frac{1+\rho^2}{n(1-\rho^2)^3} \{ 4 - 2\rho^2 + \rho^4 - (1+\rho^2)^2 \rho^{2(n-1)} + \rho^{4n} \} \\ & + \frac{(1+\rho^2)^2}{n^2 (1-\rho^2)^4} \{ 1 - \rho^{2n} \}^2 \Big], \end{aligned} \quad (3.2.12)$$

where, under the alternative hypothesis, $\rho = \rho_1$.

It may be supposed that the λ has approximately Γ distribution

$$\frac{p^\nu}{\Gamma(\nu)} \lambda^{\nu-1} e^{-p\lambda}.$$

Then

$$p = \kappa_1 / \kappa_2, \qquad \nu = \kappa_1^2 / \kappa_2,$$

and

$$\left. \begin{aligned} A_i &= \frac{\Gamma(\nu+i)}{\Gamma(\nu)} \frac{p^\nu}{(p+1)^{\nu+i}} \\ C_i &= \frac{\Gamma(\nu+i)}{\Gamma(\nu)} \frac{p^\nu}{(p+2)^{\nu+i}} \end{aligned} \right\} \quad i = 0, 1, 2, \cdots \quad (3.2.13)$$

Quite similarly, for the second sample scheme we have (3.2.5), where

$$\begin{aligned} \lambda &= \frac{(\rho_1 - \rho_0)^2 \sum_{j=1}^{n} (x_j' - \bar{x}')^2}{2\sigma^2 (1-\rho_1^2)}, \\ x_j' &= x_{2j-1}, \quad \bar{x}' = \sum_{j=1}^{n} x_j'/n \end{aligned} \quad (3.2.14)$$

and we get

$$\kappa_1 \equiv E(\lambda) = \frac{(\rho - \rho_0)^2}{2(1-\rho^2)} \left[n - \frac{1+\rho^2}{1-\rho^2} + \frac{2\rho^2 (1-\rho^{2n})}{n(1-\rho^2)^2} \right], \quad (3.2.15)$$

$$\begin{aligned} \kappa_2 \equiv V(\lambda) = \frac{(\rho - \rho_0)^4}{2(1-\rho^2)^2} \Big[& \frac{n(1+\rho^4)}{1-\rho^4} \\ & - \frac{1}{(1-\rho^2)^2 (1+\rho^2)^2} \{ (1+\rho^2)^4 + 2\rho^4 + 4(1+\rho^2)^2 \rho^{2n+2} - 2\rho^{4n+4} \} \end{aligned}$$

$$+\frac{4\rho^2}{n(1-\rho^2)^3(1+\rho^2)}\{1+\rho^2+\rho^4-(1+\rho^2)^2\rho^{2n}+\rho^{4n+4}\}$$

$$+\frac{4\rho^4}{n^2(1-\rho^2)^4}\{1-\rho^{2n}\}^2\Big], \qquad (3.2.16)$$

where the ρ should be replaced by ρ_1 under the alternative hypothesis. Assuming Γ distribution of λ as before, we can also calculate the approximate values of the mean probability and the variance of the second type error in this case.

Thus calculated approximate values of $E(\beta)$ and the standard deviation $\sqrt{V(\beta)}$ (in parenthesis) are shown in Table 1.* The difference between both sample schemes is not so remarkable.

Table 1. Approximate mean values and standard deviations (%) of conditional probabilities of the error of the second type. Significance level: $\alpha=5\%$, $f=n-2$.

The case of the first sample shceme.

H_0 \ H_1	f	$\rho_1=0$	$\rho_1=0.2$	$\rho_1=0.5$	$\rho_1=0.8$
$\rho_0=0$	10		87 (4.3)	43 (22.5)	10 (17.1)
	20		77 (6.6)	13 (12.6)	1 (3.0)
	30		67 (8.4)	3 (5.0)	0.02 (0.4)
	60		43 (9.3)	0.04 (0.2)	$<1\times10^{-5}$
	120		21 (6.8)	$<1\times10^{-4}$	$<1\times10^{-11}$
$\rho_0=0.5$	10	64 (14.4)	85 (5.3)		86 (6.7)
	20	36 (15.2)	73 (7.9)		72 (11.4)
	30	18 (11.2)	62 (9.9)		59 (14.1)
	60	2 (1.8)	34 (9.5)		30 (13.3)
	120	0.06 (0.1)	14 (5.7)		10 (7.0)
$\rho_0=0.8$	10	51 (17.6)	64 (15.8)	75 (10.5)	
	20	21 (13.9)	35 (16.5)	53 (13.5)	
	30	7 (7.2)	16 (14.5)	36 (12.9)	
	60	0.2 (0.4)	2 (2.3)	9 (5.9)	
	120	$<1\times10^{-2}$ (0.006)	0.05 (0.1)	1 (1.1)	

The case of the second sample scheme.

H_0 \ H_1	f	$\rho_1=0$	$\rho_1=0.2$	$\rho_1=0.5$	$\rho_1=0.8$
$\rho_0=0$	10		91 (1.2)	55 (31.0)	18 (18.0)
	20		86 (3.1)	33 (13.3)	2 (4.6)
	30		80 (4.2)	16 (9.2)	0.1 (0.8)
	60		66 (5.0)	1.3 (1.4)	$<1\times10^{-4}$
	120		47 (5.6)	0.03 (0.05)	$<1\times10^{-9}$
$\rho_0=0.5$	10	68 (10.8)	85 (5.8)		75 (11.5)
	20	43 (12.5)	74 (6.5)		50 (16.2)
	30	25 (10.2)	62 (8.0)		31 (15.2)
	60	4 (2.9)	36 (7.3)		6 (5.9)
	120	0.2 (0.3)	15 (4.8)		0.5 (0.9)
$\rho_0=0.8$	10	37 (16.8)	56 (14.2)	82 (5.9)	
	20	9 (8.4)	27 (12.4)	68 (8.2)	
	30	2 (2.8)	11 (7.7)	55 (10.0)	
	60	0.01 (0.06)	0.6 (0.8)	26 (7.9)	
	120	$<1\times10^{-5}$	$<1\times10^{-2}$ (0.03)	8 (3.7)	

* The numerical calculation is due to Miss H. Yamazaki and Miss O. Gotô.

§ 3.3 Continuous parameter process with rational spectral densities. In this paragraph we consider the normal stationary process $x(t)$ $(-\infty< t<\infty)$ which has the spectral densities

$$F'(\lambda)=c\left|\sum_{j=0}^{h} a_j(i\lambda)^j\right|^2, \quad a_0=1,\ a_h\neq 0 \qquad (3.3.1)$$

and satisfies the stochastic differential equation

$$\sum_{j=0}^{h} a_j x^{(j)}(t)=z'(t) \qquad \text{(formally)}, \qquad (3.3.2)$$

where the $z(t)$ $(-\infty<t<\infty)$ is a Brownian motion. Owing to (2.3.7) $\sim$ (2.3.9), statistical hypothsesis $H(\rho(\tau))$ concerning the correlogram is equivalent to the hypothesis $H(a_1, \cdots, a_h)$. Let x_t $(0\leqq t\leqq T)$ be a sample function, let $n\Delta t=T$, $t_k=k\Delta t$ $(k=0, 1, \cdots, n)$ and let

$$L_k\equiv L_{(t_{k-1},t_k)}(x(\cdot)) \qquad (k=1, 2, \cdots, n) \qquad (3.3.3)$$

be a linear functional of $x(t)\,(t_{k-1}<t<t_k)$ having the same type for every k. Then we have two methods for testing correlogram, corresponding to two types of sample schemes.

Scheme I. A set of random variables L_k $(k=1, 2, \cdots, n)$ and $x^{(j-1)}(t_k)$ $(j=1, 2, \cdots, h;\ k=0, 1, \cdots, n)$ has an $n+h(n+1)$ dimensional normal distribution and, according to Theorem 7 and Corollary to it, if we assign the values of

$$C:\quad x^{(j-1)}(t_k)=x_{t_k}^{(j-1)}, \quad j=1, 2, \cdots, h,\ k=0, 1, \cdots, n,$$

conditional random variables L_k $(k=1, 2, \cdots, n)$ are mutually independent and each L_k depends only on $x_{t_{k-1}}^{(j-1)}$ and $x_{t_k}^{(j-1)}$ $(j=1, 2, \cdots, h)$ and the regression coefficients are given by

$$b_j=L_{(t_{k-1},t_k)}(c_j(\cdot)) \qquad (j=1, 2, \cdots, 2h),$$

where the $c_j(t)$ $(t_{k-1}<t<t_k;\ j=1, 2, \cdots, 2h)$ are the regression coefficients of $x(t)$ on $x_{t_{k-1}}^{(j-1)}$ $(j=1, 2, \cdots, h)$ and $x_{t_k}^{(j-1)}$ $(j=1, 2, \cdots, h)$ which have been introduced in (2.3.30). The $2h$ coefficients b_j are functions of $\rho(\tau)$ and consequently the functions of $a_1, a_2, \cdots, a_h$. Thus there must be h restrictions among b_j $(j=1, 2, \cdots, 2h)$, and we should so choose the functional L that the restrictions are linear. Then

$$H(\rho(\tau))\sim H(b_1, b_2, \cdots, b_h). \qquad (3.3.4)$$

For instance. if we adopt

$$L_k=x((t_{k-1}+t_k)/2) \qquad (3.3.5)$$

or

$$L_k=\int_{t_{k-1}}^{t_k} x(t)\,dt, \qquad (3\text{-}3.6)$$

then we have

$$b_j = (-1)^{j-1} b_{h+j} \quad (j=1, 2, \cdots, h). \qquad (3.3.7)$$

In the general case where (3.3.7) holds, if we set

$$x_{pj}' = (x_{t_{j-1}}^{(p-1)} + (-1)^{p-1} x_{t_j}^{(p-1)})/2 \quad (p=1, 2, \cdots, h; j=1, 2, \cdots, n) \qquad (3.3.8)$$

and

$$z_j = L_j - \sum_{j=1}^{n} L_j/n \qquad (3.3.9)$$

then the test function (3.1.12) is also applied.

Scheme II. We may suppose that the $x(t)$ is a forward directed process and that $x_t (t_{k-1} \leqq t < t_k)$ $(k=1, 2, \cdots, n)$ are sample functions of the corresponding partial processes respectively. If then the conditions

$$C_k: \quad x^{(j-1)}(t_{k-1}) = x_{t_{k-1}}^{(j-1)} \quad (j=1, 2, \cdots, h) \quad k=1, 2, \cdots, n \qquad (3.3.10)$$

are given, the conditioned partial processes $x(t)$ $(t_{k-1} < t < t_k)$ $(k=1, 2, \cdots, n)$ are independent and depend only on $x_{t_{k-1}}^{(j-1)}$ $(j=1, 2, \cdots, h)$ respectively, although the conditions C_k are not independently given. Thus quite similar method to that given above is applied for the test of correlogram by using Theorem 6.

It should be noticed that when a sample function x_t $(0 \leqq t \leqq T)$ is given we can take $n = T/\Delta t$ indefinitely large, however the larger n does not always give us the better information about the autocorrelation function. Let us observe this fact for the case of normal simple Markov process for which the autocorrelation function is given by $\rho(\tau) = \exp(-\beta|\tau|)$ $(\beta > 0)$.

According to the scheme II, for instance, the regression coeficient of (3.3.6) on $x_{t_{k-1}}$ is given by

$$b = (1 - \exp(-\beta \Delta t))/\beta \qquad (3.3.11)$$

and the confidence interval length of β is approximately proportional to

$$\left| \frac{1}{\sqrt{n}} \frac{d\beta}{db} \right| = \beta^2 \Big/ \left[\sqrt{n} - \left(\sqrt{n} + \frac{\beta T}{\sqrt{n}} \right) e^{-\beta T/n} \right] \qquad (3.3.12)$$

$$\doteqdot \begin{cases} 2n^{3/2}/T^2 & \text{if } \beta T/n \ll 1 \\ \beta^2/\sqrt{n} & \text{if } \beta T/n \gg 1. \end{cases}$$

Minimizing the right hand side of (3.3.12) with respect to n, we observe that the approximately optimum value of n may be the positive integer larger than 1 and nearest to $1/u$, where u is the finite positive root of the equation

$$1 + \beta T u + 2\beta^2 T^2 u^2 = e^{\beta T u}. \qquad (3.3.13)$$

Also for the scheme I we have similar result (Ogawara [7]).

§ 3.4 Moving averages.* Let

$$x(t)=\sum_{j=0}^{k} b_j y(t-j) \qquad (b_0=1\,;\ t=0,\ \pm 1,\ \cdots) \tag{3.4.1}$$

be a normal moving averages of order k and let x_t $(t=1, 2, \cdots, N)$ be a sample sequence drawn from it. Let $1 \leqq t_1 < t_2 < \cdots < t_n \leqq N$, $t_{i+1}-t_j \geqq k+1$ $(j=1, 2, \cdots, n-1)$, then $x(t_j)$ $(j=1, 2, \cdots, n)$ are mutually independent. Thus we can apply ordinary methods for drawing statistical inference about the mean value, and the variance from the conditionally random sample x_{t_j} $(j=1, 2, \cdots, n)$. If, moreover, $x(t)$ is a regular moving averages,

$$H(\rho_1, \rho_2, \cdots, \rho_k) \sim H(b_1, b_2, \cdots, b_k) \qquad (\rho_\tau \equiv \rho(\tau))\ ,$$

and then if we denote the regression coefficients of $z_j \equiv x(j(2k+1)+k+1)$ on $x'_{pj} \equiv x(j(2k+1)+p)$ by $c_p (p=1, 2, \cdots, k)$,

$$H(\rho_1, \rho_2, \cdots, \rho_k) \sim H(c_1, c_2, \cdots, c_k)$$

and we can use the conditional F-test for these statistical hypotheses quite similarly to Theorem 12. We can also find the consistent estimate of the spectral densities through that of $b_1, b_2, \cdots, b_k$.

Next, let

$$x(t)=\int_{-\infty}^{\infty} f(s)\, d\xi\,(t-s) \tag{3.4.2}$$

be a continuous parameter moving averages (c.p.m,a.) defined by (2.3.36), where $\xi(t)$ is a Brownian motion on $(-\infty, \infty)$,

$$f(s) \equiv 0 \quad \text{for} \quad s > t_0 > 0 \quad \text{and for} \quad s < 0$$

and $f(s)$ has continuous and non-zero point in $(0, \xi)$ and in $(t_0-\varepsilon, t_0)$, for an arbitrary posive number ε smaller than t_0.

Now let $(l+1)\,\Delta t = t_0$, l being an arbitrary positive integer, and let

$$b_j y(t+j\Delta t)=\int_0^{\Delta t} f(j\Delta t+s)\, d\xi\,(t-j\Delta t-s) \qquad (j=0, 1, \cdots, l)\ , \tag{3.4.3}$$

where b_j are constants such that the $y(t+j\Delta t)$ $(j=0, 1, \cdots, l)$ have the same variance

$$\sigma_y^2=\int_0^{\Delta t} |f(s)|^2 ds > 0\,, \tag{3.4.4}$$

accordingly

$$b_j^2=\int_0^{\Delta t} |f(j\Delta t+s)|^2 ds \big/ \sigma_y^2 \qquad (j=0, 1, \cdots, l\,;\ b_0^2=1)\ . \tag{3.4.5}$$

Then, for an arbitrary constant t',

---

* This paragraph is the outline of 'Ogawara[22]' including some points improved.

$$x(t'+i\Delta t)=\sum_{j=0}^{l} b_j y(t'+i\Delta t-j\Delta t) \qquad (i=0, \pm 1, \cdots) \qquad (3.4.6)$$

is a discrete parameter moving averages (d.p.m.a.) of order l at most, and, from (2.2.8),

$$b_{l-j}+b_1 b_{l-j+1}+\cdots+b_j b_l=\int_{-\infty}^{\infty} f(u)f(u+(l-j)\Delta t)\,du\Big/\sigma_y^2 \qquad (3.4.7)$$
$$(j=0, 1, \cdots, l-1),$$

where we may assume that $b_0=1$. The process (3.4.6) may be called the d.p.m.a. of formal order l *belongs to* the c.p.m.a. (3.4.2), and if (3.4.6) is regular (3.4.2) may be said to be regular at the order l. We thus have the following theorem.

Theorem 14. *To a c.p.m.a.* (3.4.2) *belongs a d.p.m.a. of an arbitrary (formal) order and if it is regular at the order l the d.p.m.a. of order l belongs to it is unique.*

Proof. The modulus of b_j is given by the square root of (3.4.5) and the sign is determined by (3.4.7) successively.

Assume that (3.4.2) is regular at the order k and that the $f(s)$ in (3.4.2) is of known functional form with k unknown parameters $\theta_1, \theta_2, \cdots, \theta_k$ such that $H(\theta_1, \theta_2, \cdots, \theta_k) \sim H(b_1, b_2, \cdots, b_k)$. Then the conditional test of the hypothesis $H(\theta)$ $(\theta=(\theta_1, \theta_2, \cdots, \theta_k))$ is reduced to the case of d.p.m.a.

As an alternative method, let $S=\{s_p\}$ be a set of numbers such that $0<s_1<s_2<\cdots<s_k<t_0$ and set

$$\left.\begin{aligned} z_j &\equiv x(ju)=\int_{-\infty}^{\infty} f(s\,;\,\theta)\,d\xi(ju-s) \qquad (u=t_0+s_k) \\ x_{p'j} &\equiv x(ju-s_p)=\int_{-\infty}^{\infty} f(s\,;\,\theta)\,d\xi(ju-s_p-s) \qquad (p=1, 2, \cdots, k) \\ & \qquad (j=1, 2, \cdots, n) \end{aligned}\right\} \qquad (3.4.8)$$

then the regression coefficients of z_j, $c_1, c_2, \cdots, c_k$, on $x_{p'j}$ $(p=1, 2, \cdots, k)$ are the functions of

$$\int_{-\infty}^{\infty} f(s\,;\,\theta)f(s_p+s\,;\,\theta)\,ds \quad \text{and} \quad \int_{-\infty}^{\infty} f(s\,;\,\theta)f(s_p-s_q+s\,;\,\theta)\,ds \qquad (3.4.9)$$
$$(p, q=1, 2, \cdots, k),$$

consequently they are the functions of $\theta=(\theta_1, \theta_2, \cdots, \theta_k)$. Thus the conditional test for the hypothesis $H(\theta)$ is reduced to that of $H(c)$ $(c=(c_1, c_2, \cdots, c_k))$, provided that $H(\theta)\sim H(c)$, where the z_j are random variables and the $x_{p'j}$ are fixed variates and, if the sample function x_t is given on $0\leqq t \leqq T$ $(T>kt_0)$, the size n of the sample $(z_1, z_2, \cdots, z_n)$ is such that

$$(n-1)u+s_k \leqq T < nu. \qquad (3.4.10)$$

Hence, in any case,

$$T/2t_0 < n < (T/t_0) + 1 . \qquad (3.4.11)$$

Now, a problem rises here is what set S gives us the most efficient test. Let us consider a simple case where $k=1$ and

$$\begin{aligned} f(s\,;\theta) &= e^{-\theta s} && \text{for } 0 \leqq s \leqq t \\ &= 0 && \text{for } s<0 \text{ and for } s>t . \end{aligned} \qquad (3.4.12)$$

In this case, the regression coefficient of $x(ju)$ on $x(ju-s_1)$ $(u=t_0+s_1)$ is given by

$$c(s_1;\theta) = \rho(s_1) = e^{-\theta s_1}(1-e^{-2\theta(t_0-s_1)})/(1-e^{-2\theta t_0}) . \qquad (3.4.13)$$

This is a monotone decreasing function of θ and $c(s_1;\theta)\to 0(s_1 \downarrow 0)$, $c(s_1;\theta)\to 1(s_1 \uparrow t_0)$ uniformly for $0<\theta<\theta_1$, θ_1 being an arbitrary positive constant. Since n is limited by (3.4.11), the value of s_1 which maximizes $|\partial c/\partial\theta|$ may be approximately optimum.

§ **3.5 Time series with trend.** In this paragraph, we consider the trend of mean value exclusively.

Definition. *Let* $x(t)$ $(t=0, \pm 1, \cdots)$ *be a real stochastic process such that* $x(t)-m(t)$ *is (wide sense) stationary, where* $m(t)=E(x(t))$. *Then*

$$\begin{aligned} &x(t+s)+b_1x(t-1)+\cdots+b_hx(t-h) \\ &\quad -\beta_1v_1(t)-\beta_2v_2(t)-\cdots-\beta_rv_r(t)=z(t) \qquad (s\geqq 0) \end{aligned} \qquad (3.5.1)$$

is said to be a linear model if 1) *the* $v_j(t)$ *are linearly independent known functions,* 2) *the* b_i *and the* β_j *are independent parameters and* 3) *the* $z(t)$ *is a process of moving averages of finite order,* $E(z(t))=0$.

Our methods of conditional inference can be applied to some of such processes that follow a linear model. In the following we suppose that the mean value function of a discrete parameter process $x(t)$ is of the form

$$m(t)=\sum_{j=1}^{r} \alpha_j u_j(t) , \qquad (3.5.2)$$

where the $u_j(t)$ are linearly independent known functions and the α_j are unknown parameters.

Theorem 15. *Suppose that* $y(t)=x(t)-m(t)$ *is an autoregression process of order* h, *where the* $m(t)$ *is given by* (3.5.2). *In order that the* $x(t)$ *follows a linear model of the form* (3.5.1), *it is necessary and sufficient that the* $m(t)$ *function has the following form.*

$$\begin{aligned} m(t) &= \sum_{j=1}^{r_1} \alpha_j t^{j-1} + \sum_{j=1}^{r_2}\sum_{i=1}^{m_j} \alpha_i^{(j)} t^{i-1} p_j^t \\ &\quad + \sum_{j=1}^{r_3}\left\{\sum_{i=1}^{n_j} \alpha_{1i}^{(j)} t^{i-1} q_j^t \cos\lambda_j t + \sum_{i=1}^{n_j} \alpha_{2i}^{(j)} t^{i-1} q_j^t \sin\lambda_j t\right\}, \end{aligned} \qquad (3.5.3)$$

where α_j, $\alpha_i^{(j)}$, $\alpha_{1i}^{(j)}$ *and* $\alpha_{2i}^{(j)}$ *are parameters*, $p_j(p_j-1)\neq 0$, $p_i\neq p_j$ $(i\neq j)$, $q_j\neq 0$, $0<\lambda_j<2\pi$, $\lambda_j\neq\pi$, $\lambda_j\neq\lambda_{j'}$ $(j\neq j')$ *and*

$$r_1+\sum_{j=1}^{r_2} m_j+2\sum_{j=1}^{r_3} n_j=r \qquad (r_1\geqq 0,\ r_2\geqq 0,\ r_3\geqq 0)\ .$$

Proof. According to Theorem 13, the $y(t)$ process satisfies almost surely a stochastic finite difference equation

$$y(t+s)+b_1 y(t-1)+\cdots+b_h y(t-h)=z(t) \qquad (s\geqq 0)\,, \tag{3.5.4}$$

where the $z(t)$ is a moving averages of order s.

i) *Sufficiency*. Substituting $y(t)=x(t)-m(t)$ in (3.5.4), we get the expression of the form (3.5.1), in which corresponding to the term

$$\sum_{i=1}^{n}\alpha_{1i}t^{i-1}q^t\cos\lambda t+\sum_{i=1}^{n}\alpha_{2i}t^{i-1}q^t\sin\lambda t$$

the coefficients of $t^{i-1}q^t\cos\lambda t$ and $t^{i-1}q^t\sin\lambda t$ are given by

$$\begin{aligned}\beta_{1i}=q^s\Big\{&\cos\lambda s\sum_{\nu=i}^{n}\binom{\nu-1}{i-1}s^{\nu-i}\alpha_{1\nu}+\sin\lambda s\sum_{\nu=i}^{n}\binom{\nu-1}{i-1}s^{\nu-i}\alpha_{2\nu}\Big\}\\ &+\sum_{k=1}^{h}b_k q^{-k}\Big\{\cos\lambda k\sum_{\nu=i}^{n}\binom{\nu-1}{i-1}(-k)^{\nu-i}\alpha_{1\nu}\\ &-\sin\lambda k\sum_{\nu=i}^{n}\binom{\nu-1}{i-1}(-k)^{\nu-i}\alpha_{2\nu}\Big\}\end{aligned} \tag{3.5.5}$$

$$\begin{aligned}\beta_{2i}=q^s\Big\{&-\sin\lambda s\sum_{\nu=i}^{n}\binom{\nu-1}{i-1}s^{\nu-i}\alpha_{1\nu}+\cos\lambda s\sum_{\nu=i}^{n}\binom{\nu-1}{i-1}s^{\nu-i}\alpha_{2\nu}\Big\}\\ &+\sum_{k=1}^{h}b_k q^{-k}\Big\{\sin\lambda k\sum_{\nu=i}^{n}\binom{\nu-1}{i-1}(-k)^{\nu-i}\alpha_{1\nu}\\ &+\cos\lambda k\sum_{\nu=i}^{n}\binom{\nu-1}{i-1}(-k)^{\nu-i}\alpha_{2\nu}\Big\}\end{aligned} \tag{3.5.6}$$

respectively, therefore β_{1i}, β_{2i} $(i=1,2,\cdots,n)$ are independent parameters, the coefficent of $t^{i-1}p^t$ that comes from the term

$$\sum_{i=1}^{n}\alpha_i t^{i-1}p^t$$

is given by (3.5.5) by setting $\lambda=0$, $q=p$ and the coefficient of t^{i-1} is also given by (3.5.5) by setting $\lambda=0$, $q=1$. Thus the coefficients β_j in the form (3.5.1) are functionally independent each other and are also independent of b_j and the functions $v_j(t)$ are of the form t^{i-1} or $t^{i-1}p_j^t$ or pairs of $t^{i-1}q_j^t\cos\lambda_j t$ and $t^{i-1}q_j\sin\lambda_j t$ $(i=1,2,\cdots)$, and they are linearly independent.

ii) *Necessity*. From (3.5.1), (3.5.2) and (3.5.4) we get

$$\sum_{j=1}^{r}\alpha_j u_j(t+s)+\sum_{i=1}^{h}\sum_{j=1}^{r}b_i\alpha_j u_j(t-i)=\sum_{k=1}^{r}\beta_k v_k(t)\ . \tag{3.5.7}$$

Differentiating the both sides of (3. 5. 7) partially by b_i and then by α_j we have

$$u_j(t-i)=\sum_{k=1}^{r} c_{ijk}v_k(t) \qquad (i=1, 2, \cdots, h\, ;\ j=1, 2, \cdots, r)\,, \qquad (3.5.8)$$

where the $c_{ijk}=\partial^2\beta_k/\partial b_i\partial\alpha_j$ are constants. Because of the linear independence of $u_j(t)$ $(j=1, 2, \cdots, r)$, the rank of the $hr\times r$ matrix (c_{ijk}) is r. Eliminating r functions $v_k(t)$ from the hr equations (3. 5. 8), we get $(h-1)r$ simultaneous homogeneous finite difference equations of r functions $u_j(t)$. So long as the solutions of these equations exist and (3. 5. 1) holds, from the independence of parameters $\beta_1, \beta_2, \cdots, \beta_r$, we can deduce that (3. 5. 2) should be of the form (3. 5. 3).

Similarly we observe the following

Theorem 16. *Let $y(t)=x(t)-m(t)$ be a moving averages of order h, $m(t)=\sum_{j=1}^{r}\alpha_j u_j(t)$. A necessary and sufficient condition that the $x(t)$ process complies with a linear model of the form* (3. 5. 1) *is that the $m(t)$ is of the form* (3. 5. 3).

In this case the order of the moving averages $z(t)$ is $s+2h+1$.

When $y(t)=x(t)-m(t)$ is an autoregression process,

$$y(t)+a_1y(t-1)+\cdots+a_hy(t-h)=z(t)\,, \quad a_h\neq 0\,, \qquad (3.5.9)$$

or a moving averages,

$$y(t)=z(t)+a_1z(t-1)+\cdots+a_hz(t-h)\,, \quad a_h\neq 0\,, \qquad (3.5.10)$$

and the $x(t)$ complies with a linear model (3. 5. 1), even if the values of $x(jl-\tau)$ $(\tau=1, 2, \cdots, h;\ j=0, 1, \cdots)$ are given, the $z(jl)$ $(j=0, 1, \cdots)$ is a non-autocorrelated process, where $l=h+s+1$ in the case of (3. 5. 9) or $l=2h+s+2$ in the case of (3. 5. 10).

Therefore, for a normal autoregression process with trend (3. 5. 3), when a time series x_t $(t=1, 2, \cdots, N)$ is given, supposing $x_{j(h+s+1)}$ $(j=1, 2, \cdots, n)$ random variables and $x_{(h+s+1)-s-k}$ $(k=1, 2, \cdots, h;\ j=1, 2, \cdots, n)$ together with $v_j(j(h+s+1))$ $(j=1, 2, \cdots, n)$ fixed variates, we can find the conditional maximum likelihood estimates of $a_1, a_2, \cdots, a_h$ and $\alpha_1, \alpha_2, \cdots, \alpha_r$ (if exist) through those of $b_1, b_2, \cdots, b_h$ and $\beta_1, \beta_2, \cdots, \beta_r$ and we can apply the conditional test given in the preceding paragraphs to statistical hypotheses concerning the parameters a_i and α_j or the correlogram of $y(t)$ process.

Quite similar methods are useful for the conditional inference of a normal moving averages with trend of the form (3. 5. 3).

Furthermore, our considerations in this paragraph will be extended to multidimensional processes and to continuous parameter processes.

§ 3. 6 Discontinuous Markov process. Let $x(t)$ $(t=0, \pm 1, \cdots)$ be a strictly stationary simple Markov process which can only assume values in a finit set $\{0, 1, \cdots, m\}$, let $p_j = \Pr\{x(t) = j\}$ $(j = 0, 1, \cdots, m)$ be the absolute probabilities and let $p_{ij} = \Pr\{x(t) = j \mid x(t-1) = i\}$ $(i, j = 0, 1, \cdots, m)$ denote the transition probabilities, where $p_i > 0$ $(i = 0, 1, \cdots, m)$.

As before we have two kinds of sample schemes for an observed sequence x_t $(t = 1, 2, \cdots, 2n+1)$. For the conditional estimation or the conditional test of absolute probabilities, we may use the first one, where the x_{2k} $(k = 1, 2, \cdots, n)$ are random variables and the x_{2k-1} $(k = 1, 2, \cdots, n+1)$ are fixed variates. Let $n(j)$ be the number of j in the set of outcomes x_{2k} $(k = 1, 2, \cdots, n)$ and let

$$I_{2k}(j) = \begin{cases} 1 & \text{if } x_{2k} = j \\ 0 & \text{if } x_{2k} \neq j, \end{cases}$$

then the $I_{2k}(j)$ are conditionally independent random variables given $x' = (x_1, x_3, \cdots, x_{2n+1})$,

$$n(j) = \sum_{k=1}^{n} I_{2k}(j)$$

and

$$\underset{x}{E}\{n(j)/n \mid x'\} = \frac{1}{n} \sum_{k=1}^{n} E\{I_{2k}(j) \mid x_{2k-1}, x_{2k+1}\},$$

$$\underset{x'}{E}\ \underset{x}{E}\{n(j)/n \mid x'\} = p_j,$$

where $x = (x_2, x_4, \cdots, x_{2n})$. Let $(p_1(j), p_2(j))$ be the conditional confidence interval for p_j given x', with confidence coefficient $1 - 2\alpha$,

$$\Pr\left\{\sum_{k=1}^{n} I_{2k}(j) \geqq n(j) \mid p_1(j), x'\right\}$$

$$= \Pr\left\{\sum_{k=1}^{n} I_{2k}(j) \leqq n(j) \mid p_2(j), x'\right\} = \alpha,$$

then we have, for fixed $n(j)$,

$$\underset{x'}{E} \Pr\left\{\sum_{k=1}^{n} I_{2k}(j) \geqq n(j) \mid p_1(j), x'\right\} = \alpha$$

and similar equation about the upper limit. The confidence limits $p_1(j)$ and $p_2(j)$ will serve for setting critical region for the conditional test of hypotheses concerning p_j.

Next, for the conditional statistical inference of transition probabilities, we may use the second sample scheme. Let n_i be the number of i in the outcomes x_{2k-1} $(k = 1, 2, \cdots, n)$, $x_{2k(\nu)-1} = i$ $(\nu = 1, 2, \cdots, n_i)$ and let $n_i(j)$ be the number of pair (i, j) in the pairs $(x_{2k(\nu)-1}, x_{2k(\nu)})$ $(\nu = 1, 2, \cdots, n_i)$, $n_i(j) = \sum_{\nu=1}^{n_i} I_{2k(\nu)}$, then $n_i(j)/n_i$ is a conditional estimate for p_{ij} in the following meaning.

$$\underset{x}{E}\left\{\sum_{\nu=1}^{n_i} I_{2k(\nu)}/n_i \mid x'\right\} = p_{ij}$$

and, since $I_{2k(\nu)}$ are mutually indepenendt when x' is given,

$$V_{x}\left\{\sum_{\nu=1}^{ni} I_{2k(\nu)}/n_i \ x'\right\} = p_{ij}(1-p_{ij})/n_i .$$

According to the strong law of large numbers for Markov chain (Feller [30]), if there is only one ergodic part,

$$\lim_{n\to\infty} \frac{n_i}{n} = p_i \qquad \text{(almost surely)} .$$

Therefore as $n \to \infty$

$$\sum_{\nu} I_{2k(\nu)}/n_i \to p_{ij} \qquad \text{(in probability)} .$$

The confidence limits for p_{ij} by means of n_i and $n_i(j)$ depend on the fixed variate x', but they depend only on the frequencies $\sum_k I_{2k}$ and $\sum_\nu I_{2k(\nu)}$.

The extension of the idea mentioned above to the stationary multiple Markov process will be immediate and omitted here.

References

[1] M. OGAWARA; *On stationary normal multiple Markov process,* Bull. Math. Stat., **2** (1948), 42–46. (in Japanese)
[2] M. OGAWARA; *Statistical analysis of vector time series,* Bull. Math. Stat., **3** (1949), 46–50. (in Japanese)
[3] M. OGAWARA; *On stochastic interpolation of omitted observation,* Papers in Meteor. and Geophys. (Meteorological Research Institute), **1** (1950), 50–57.
[4] M. OGAWARA; *A note on the test of serial correlation coefficients,* Annals of Math. Stat., **22** (1951), 115–118.
[5] M. OGAWARA; *On the method of long-range-forecasting,* Jour. Meteor. Res., Central Meteor. Observatory, **4** (1952), 444–458. (in Japanese)
[6] M. OGAWARA; *Multivariate regression theory and its application to the problem of climatic change,* (with H. Yamazaki), Jour. Meteor. Soc. Japan, **30** (1952), 158–165.
[7] M. OGAWARA; *On the theory of statistical inference of time series, Progress of probability theory and statistics,* (edited by T. Kitagawa), Iwanami, (1953), 131–165. (in Japanese)
[8] M. OGAWARA; *On the prediction by small sample,* Jour. Meteor. Res. **5** (1954), 752–763. (in Japanese)
[9] M. OGAWARA; *On the stochastic seasonal prediction,* (with collaborators), Jour. Meteor. Soc. Japan, **32** (1954), 253–280. (in Japanese)
[10] M. OGAWARA; *Mathemetical theory of prediction,* Kagaku, **24** (1954), 489–495. (in Japanese)
[11] M. OGAWARA; *A general stochastic prediction formula,* Papers in Meteor. and Geophys., **5** (1955), 193–202.
[12] M. OGAWARA; *Efficiency of a stochastic prediction,* Papers in Meteor. and Geophys., **5** (1955), 203–211.
[13] M. OGAWARA; *A prediction for the next maximum of sun spot numbers,* Papers in Meteor. and Geophys., **5** (1955), 212–216.
[14] M. OGAWARA; *A stochastic numerical prediction for the* 500 *mb height along the latitude* 45° *N,* Jour. Meteor. Res., **7** (1955), 191–200. (in Japanese)
[15] M. OGAWARA; *Probability of the coming felt earthquake to Tokyo,* Kenshin-Jihô, **20** (1955), 81–92. (in Japanese)

[16] M. OGAWARA; *Statistical infcrence and prediction of a stochastic process*, Tôkei-Kagaku-Kenkyu, **1** (1956), 10–20. (in Japanese)
[17] M. OGAWARA; *Weather prediction and OR*, Applied Stat. in Meteor., **7** (1956), 2–5. (in Japanese)
[18] M. OGAWARA; *Lectures on time series analysis*, Applied Stat. in Meteor., **6** (1955), 58–61, **6** (1956), 88–93. (in Japanese)
[19] M. OGAWARA; *On duration curve*, Report of Economic Division, Central Research Institute of Electric Power Industry (1956), 1–11. (in Japanese)
[20] M. OGAWARA; *Statistical method of prediction for latent electric power*, Report of Economic Division, Central Research Institute of Electric Power Industry, (1956), 1–9. (in Japanese)
[21] M. OGAWARA; *Statistical method of prediction for latent electric power*, II, Report of Eccnomic division, Central Research Institute of Electric Power Industry, (1957), 1–11. (in Japanese)
[22] M. OGAWARA; *An exact test for moving averages*, Bull. Math. Stat., **7** (1957), 77–83.
[23] M. OGAWARA; *On fiducial prediction*, Essays and Studies, Tokyo Woman's Christian College, 8 (1957), 17–25.
[24] T. KITAGAWA; *Sampling from processes depending upon a continuous parameter*, Mem. Fac. Sci., Kyushu Univ., Ser. A, **5** (1950), 181–188.
[25] J. L. DOOB; *The elementary Gaussian process*, Ann. Math. Stat., **15** (1944), 229–282.
[26] J. L. DOOD; *Stochastic processes*, Wiley, (1953)
[27] H. WOLD; *A study in the analysis of stationary time series*, Almqvist & Wiksell, (2nd ed., 1954)
[28] S. G. GHURYE; *Random functions satisfing certain linear relations*, II, Ann. Math. Stat., **26** (1955), 105–111.
[29] H. CRAMÉR; *Mathematical Methods of Statistics*, Princeton Univ. Press, (1946)
[30] W. Feller; *An introduction to probability theory and its applications*, I, Wiley (1950)