VIII. POLYNOMIAL APPROXIMATION OF STRICTLY STATIONARY PROCESSES CONTINUOUS IN PROBABILITY

1. Introduction

Let $\mathcal{I}$ be the system of all finite intervals.

A white noise (with continuous parameter) is defined as a Gaussian system B_I, $I \in \mathcal{I}$ with

(1) $$E(B_I) = 0 , \qquad E(B_I \cdot B_{I'}) = |I \wedge I'|$$

($| \ |$ = Lebesgue measure). Since

(2) $$\sum_{ij} |I_i \wedge I_j| \, \xi_i \bar{\xi}_j = \int | \sum_i \xi_i e_{I_i}(t)|^2 dt \geq 0 ,$$

the existence and uniqueness (up to law-equivalence) of white noise is clear.

It follows from (1) and (2) that

(3) differential: if (I_i) are disjoint, then (B_{I_i}) are independent.

(4) stationary: for any system $\{I_1, I_2, \dots , I_n\}$, we have

$$(B_{I_1+t}, \dots , B_{I_n+t}) \underset{(L)}{\sim} (B_{I_1}, B_{I_2}, \dots , B_{I_n}) .$$

where $I + t = \{s + t; s \in I\}$ and $\underset{(L)}{\sim}$ means law-equivalence.

Using (4) we can define an automorphism group $\{\Phi_t\}$ and $\mathcal{B}(B)$ such that

(5) $$\Phi_t B_I = B_{I+t}$$

Take any polynomial $p(\xi_1, \xi_2, \dots, \xi_k)$ and any system of intervals $I_1, I_2, \dots, I_k$ and consider

(6) $$x_t = p(B_{I_1+t}, B_{I_2+t}, \dots, B_{I_n+t}) .$$

Then x_t is a strictly stationary process continuous in probability. Any process of the form (6) is called a polynomial stationary process.

Now we shall ask a

Question: Given any strictly stationary process x_t continuous in probability, can we find a sequence of polynomial stationary processes $x_t^{(n)}$ such that

(7) $$x^{(n)} \xrightarrow[(L)]{} x.$$

i.e.

(7') $\forall t_1 \cdots t_k$

$$(x_{t_1}^{(n)}, \dots, x_{t_k}^{(n)}) \xrightarrow[(L)]{} (x_{t_1}, \dots, x_{t_k})$$

($\xrightarrow[(L)]{}$ means law-convergence).

However this question is not natural, because, even if $x_t^{(n)}$, $n = 1, 2, \ldots$ are all polynomial stationary processes and if $x^{(n)} \longrightarrow x.$, $x.$ is strictly stationary but not always continuous in probability.

Therefore we shall introduce the following law-topology.

Definition 1. $U(x, \epsilon)$ is the collection of all stochastic processes y such that

$$(8) \quad m, |\theta_1|, \ldots, |\theta_m|, |t_1|, \ldots, |t_m| < \epsilon^{-1}$$

$$\Longrightarrow \left| E(\exp(i \sum_{j=1}^{m} \theta_j y_{t_j})) - E(\exp(i \sum_{j=1}^{m} \theta_j x_{t_j})) \right| < \epsilon$$

$$(9) \quad \rho_U(x,y) = \inf\{\epsilon : y \in U(x,\epsilon)\} .$$

ρ_U is a metric on the space $\tilde{\mathcal{S}}$ of all stochastic processes (two stochastic processes with the same probability law being identified).

This topology is natural in the sense that (10) the set of all strictly stationary processes continuous in probability form a closed set of $\tilde{\mathcal{S}}$, and we can prove the following theorem which gives an affirmative answer to the question stated above.

Approximation Theorem. The set $\mathcal{P}$ of all polynomial processes is a dense subset of $\mathcal{J}$.

The approximation theorem for ergodic stationary processes continuous in probability was discussed by N. Wiener in his paper "The Homogeneous Chaos". Although neither his statement nor his proof is very clear, his argument contains a very ingenious idea. Polishing Wiener's method and using Oxtoby-Ulam's idea used in measure preserving flows, Nisio gave a neat proof to the approximation theorem without ~~using~~ assuming the ergodicity. We shall present her proof here.

N. Wiener: The homogeneous chaos, Amer. J. Math. 60 (1938), 897-936.

M. Nisio : On polynomial approximation for strictly stationary processes. J. Math. Soc., Japan 12 (1960), 207-226.

In the course of the proof of the theorem we shall often use the following process-topology (not law-topology).

Definition 2. Let $\Omega(\mathcal{B}, P)$ be a probability space and let $\mathcal{J}(\Omega)$ be the space of all stochastic processes defined on $\Omega(\mathcal{B}, P)$.

$V(x,\epsilon)$ is the set of all processes y defined on the same probability space such that

(11) $$|t| < \delta^{-1} \Longrightarrow P(|y_t - x_t| > \delta) < \delta$$

and

(12) $$\rho_V(x,y) = \inf\{\delta : y \in V(x,\delta)\}$$

Corollary. If x and y are defined on the same probability space, then

(13) $V(x, \epsilon^3/5) \subset U(x, \epsilon)$,

i.e.,

(13') $\rho_U(x,y) < \sqrt[3]{5\rho_V(x,y)}$;

and so

(14) $x_n \xrightarrow[\rho_V]{} x$ implies $x_n \xrightarrow[\rho_U]{} x$.

Proof. If $\epsilon > 2$, then $U(x,\epsilon) = \tilde{S}$ and so (13) is evident. Therefore we consider the case $\epsilon \leq 2$. Assume that $y \in V(x, \epsilon^3/5)$. Then we can see

$$|m|, \ |\theta_1|, \ \dots, \ |\theta_m|, \ |t_1|, \ \dots, \ |t_m| < \epsilon^{-1}$$

$$\Longrightarrow \quad |E(\exp(i \sum_{j=1}^{m} \theta_j y_{t_j})) - E(\exp(i \sum_{j=1}^{m} \theta_j t_{t_j}))|$$

$$\leq \sum_{j=1}^{m} E|e^{i\theta_j y_{t_j}} - e^{i\theta_j x_{t_j}}|$$

$$\leq \sum_{j=1}^{m} E|\exp[i\theta_j(y_{t_j} - x_{t_i})] - 1|$$

$$\leq \sum_{j=1}^{m} \left[2P(|y_{t_j} - x_{t_i}| > \frac{\epsilon^3}{5}) + |\theta_j| \frac{\epsilon^3}{5} P(|y_{t_j} - x_{t_i}| \leq \frac{\epsilon^3}{5})\right]$$

$$= \epsilon^{-1}\left(2 \ \frac{\epsilon^3}{5} \cdot \epsilon^{-1} \ \frac{\epsilon^3}{5} \cdot 1\right)$$

(Notice $\left(\frac{\epsilon^3}{5}\right)^{-1} \geq \left(\frac{4}{5}\epsilon\right)^{-1} > \epsilon^{-1}$)

$$= \frac{2\epsilon^2}{5} + \frac{\epsilon}{5} \leq \frac{4\epsilon}{5} + \frac{\epsilon}{5} = \epsilon$$

Therefore $y \in U(x,\epsilon)$. Thus (13) is proved. (13') and (14) follow from (13) immediately.

2. Case: x_t is Gaussian

To make it easier to understand the problem, let us first consider the Gaussian case.

In this case we can approximate x_t by polynomial stationary processes of degree 1.

We can assume $E(x_t) = 0$. By the continuity in probability, we have the Khinchine (Hinčin) decomposition)

(1) $E(x_t, x_s) = \int e^{i(t-s)\lambda} dF(\lambda)$.

In case $dF(\lambda) = f(\lambda)\, d\lambda$, we can express x_t as

(2) $x_t = \int g(t-s)\, dB(s)$ for some $g \in L^2$, g real

If g is a step function, then x_t itself is a polynomial process of degree 1. If g is a general L^2-function, then take a sequence of step functions g_n such that

$$\int |g_n - g|^2\, ds \longrightarrow 0$$

Then
$$x_t^{(n)} = \int g_n(t-s)\, dB(s)$$

is a polynomial process and

$$E(|x_t^{(n)} - x_t|^2) = \int |g_n - g|^2 ds \longrightarrow 0 \qquad \text{(uniformly in } t)$$

Therefore

$$x^{(n)} \xrightarrow[\rho_V]{} x \quad \text{and so} \quad x^{(n)} \xrightarrow[\rho_U]{} x .$$

In case $dF(\lambda)$ is a general measure, take a sequence $f_n(\lambda) \geq 0$ such that $f_n(\lambda)\, d\lambda \longrightarrow dF(\lambda)$ (weak*). Then

$$r_n(t) = \int e^{i\lambda t} f_n(\lambda)\, d\lambda \longrightarrow r(t) = \int e^{i\lambda t}\, dF(\lambda)$$

uniformly in each bounded t-region. Construct a Gaussian process $x_n(t)$ with mean 0 and the covariance function $r_n(t)$ for each n. Then

$$x_n \xrightarrow[\rho_U]{} x,$$

because, as $n \to \infty$,

$$\left| E\left(\exp\left[i \sum_{j=1}^{m} x_n(t_j)\, \theta_j\right]\right) - E\left(\exp\left[i \sum_{j=1}^{m} x(t_j)\, \theta_j\right]\right) \right|$$

$$= \left| \exp\left\{-\frac{1}{2} \sum_{j,k=1}^{m} r_n(t_j - t_k)\, \theta_j \theta_k\right\} - \exp\left\{-\frac{1}{2} \sum_{j,k=1}^{m} r(t_j - t_k)\, \theta_j \theta_k\right\} \right|$$

$\longrightarrow 0$ uniformly in $|\theta_j|$, $|m|$, $|t_i| < \epsilon^{-1}$ for every $\epsilon > 0$.

This completes the proof.

3. Approximation Theorem for the Discrete Time parameter Case.

To explain the crusial point of Wiener's technique, let us consider an analogous theorem for the discrete time parameter case.

We shall first introduce some preliminary notions.

White Noise. $(B_k, k \in Z)$ is called a white noise with discrete parameter, if B_k, $k \in Z$, are normally $(N(0, 1))$ and independently distributed.

Polynomial Stationary Process. A process of the form

$$(4) \qquad y_k = p(B_{\ell_1+k}, B_{\ell_2+k}, \dots, B_{\ell_n+k})$$

(p = polynomial of n real variables)

is called a polynomial stationary process.

Law-Topology. $U(x, \varepsilon)$ and $\rho_U(x, y)$ are defined in the same way as in the continuous time parameter case.

$$x^{(n)} \underset{\rho_U}{\longrightarrow} x$$

$$\Leftrightarrow (x_{k_1}^{(n)}, \dots, x_{k_m}^{(n)}) \underset{(L)}{\longrightarrow} (x_{k_1}, \dots, x_{k_m}) \quad \text{for every } (k_i).$$

Process-Topology. $V(x, \varepsilon)$ and $\rho_V(x, y)$ are defined in the same way as in the continuous parameter case

$$\rho(x^{(n)}, x) \to 0 \iff x_k^{(n)} \to x_k \quad \text{in probability for every } k.$$

$$V(x, \frac{\varepsilon^3}{5}) \subset V(x, \varepsilon)$$

$$\rho_U(x, y) \leq \sqrt[3]{5\rho_V(x, y)}$$

Approximation Theorem. Given any strictly stationary sequence $x_k(\omega^*)$, $k \in Z$, $\omega^* \in \Omega^*(\mathcal{B}^*, P^*)$, we can find a sequence of polynomial stationary sequences $(x_p^{(n)}(\omega))$ of $B_k(\omega)$, $k \in Z$, $\omega \in \Omega(\mathcal{B}, P)$ such that

$$x^{(n)} \xrightarrow[(\rho_U)]{} x .$$

We shall first prepare two lemmas.

Lemma 1. If $x_k(\omega^*)$ is bounded, (i.e., $\exists M < \infty$ such that $|x_k(\omega^*)| < M$ for every k, ω^*) and ergodic, then there exists $\omega_0^* \in \Omega^*$ such that, for every $m, \theta_1, \ldots, \theta_m, k_1, \ldots, k_m$

$$(1) \qquad \frac{1}{n}\sum_{\ell=0}^{n-1} \exp\left\{i \sum_{j=1}^{m} \theta_j \, x_{k_j+\ell}(\omega_0^*)\right\} \longrightarrow E\left\{\exp\left(i \sum_{j=1}^{m} \theta_j \, x_{k_j}\right)\right\}$$

Proof of Lemma 1. Fix any $(m, \theta_1, \ldots, \theta_m, k_1, \ldots, k_m)$. By Birkhoff's ergodic theorem, the set $\Omega_{m,\theta_1,\ldots,\theta_m,k_1,\ldots,k_m}$ of ω^* for which (1) holds has P^*-measure 1. Therefore

$$\Omega_0^* = \bigcap_{\substack{m \\ k_i \\ \theta_i \text{ rational}}} \Omega_{m,\theta_1,\ldots,\theta_m,k_1,\ldots,k_m}$$

has P_0-measure 1. Take any point ω_0^* from Ω_0^*. Then (1) holds for any m, any θ_i rational and any k_i. In order to verify (1) for any m, any θ_i (may be irrational) and any k_i, observe

$$(2) \qquad \left|\frac{1}{n}\sum_{\ell=0}^{n-1} \exp\left(i \sum_{j=1}^{m} \theta_j \, x_{k_j+\ell}(\omega_0^*)\right) - \frac{1}{n}\sum_{\ell=0}^{n-1} \exp\left(i \sum_{i=1}^{m} \theta_j' \, x_{k_j+\ell}(\omega_0^*)\right)\right|$$

$$\leq \frac{1}{n}\sum_{\ell=0}^{n-1}\sum_{j=1}^{m} \left|\exp\left(i\,\theta_j \, x_{k_i+\ell}(\omega_0^*)\right) - \exp\left(i\,\theta_j' \, x_{k_j+\ell}(\omega_0^*)\right)\right|$$

$$\leq \frac{1}{n}\sum_{\ell=0}^{n-1}\sum_{j=1}^{m} \left|\theta_j - \theta_j'\right| \left|x_{k_j+\ell}(\omega_0^*)\right|$$

$$\leq m \cdot M \sup_j \left|\theta_j - \theta_j'\right|$$

and similarly

(3) $$\left| E\left(\exp\left\{i \sum_{j=1}^{m} \theta_j x_{k_j}\right\}\right) - E\left(\exp\left\{i \sum_{j=1}^{m} \theta'_j x_{k_j}\right\}\right)\right|$$

$$\leq m \cdot M \sup_j |\theta_j - \theta'_j| ,$$

take rational θ'_j such that $m \cdot M \sup_j |\theta_j - \theta'_j| < \varepsilon/3$ and determine $n_0 = n_0(\varepsilon)$ such that $n > n_0$ implies

$$\left|\frac{1}{n} \sum_{\ell=0}^{n-1} \exp\left(i \sum_{j=1}^{m} \theta'_j x_{k_j+\ell}(\omega^*_0)\right) - E\left(\exp\left\{ i \sum_{j=1}^{m} \theta'_j x_{k_j} \right\}\right)\right| < \frac{\varepsilon}{3} .$$

Then $n > n_0$ implies

$$\left|\frac{1}{n} \sum_{\ell=0}^{n-1} \exp\left\{i \sum_{j=1}^{m} \theta_j x_{k_j+\ell}(\omega^*_0)\right\} - E\left(\exp\left\{i \sum_{j=1}^{m} \theta_j x_{k_j}\right\}\right)\right| < \varepsilon .$$

<u>Lemma 2</u>.

$$a_n(\omega) = \min(i \geq 0 : |B_{-n+i}(\omega)| \leq 1,\ |B_{-n+i+1}(\omega)| > 1,\ \ldots,\ |B_{-n+i+n}(\omega)| > 1)$$

($= \infty$ if such i does not exist)

satisfies

(a.0) $$P(a_n < \infty) = 1 ,$$

(a.1) $$P(a_n = 0) = P(a_n = 1) = \cdots = P(a_n = n) \geq P(a_n = n+1)$$
$$\geq P(a_n = n+2) \geq \cdots \geq P(a_n = n+k) \geq \cdots ,$$

(a.2) $$P(a_n = i) = \sum_{k=1}^{\infty} c_{n+k}(i)\, [P(a_n = n+k-1) - P(a_n = n+k)]$$

where

$$c_k(i) = \frac{1}{k} , \qquad i = 0, 1, 2, \ldots , k-1$$
$$= 0, \qquad i = k, k+1, \ldots ,$$

(a.3) $$P(\Phi_\ell\, a_n = a_n - \ell \quad \text{for} \quad |\ell| \leq m) > 1 - \frac{2m}{n} .$$

<u>Proof of Lemma 2.</u>

(a.0) $$P(a_n < \infty)$$
$$= P(\exists_i \geq 0; |B_{-n+i}| \leq 1, |B_{-n+i+1}| > 1, \ldots, |B_1| > 1)$$
$$\geq P(\exists_i \geq 0; |B_{(n+1)i}| \leq 1, |B_{(n+1)i+1}| > 1, \ldots , |B_{(n+1)i+n}| > 1)$$
$$= 1 \qquad \text{(by Borel-Cantelli's lemma)}$$

(a.1). This follows from

$$P(a_n = i) = P(|B_{-n+i}| \leq 1,\ |B_{-n+i+1}| > 1,\ \dots\ ,\ |B_1| > 1)$$

$$\text{for } i = 0, 1, 2, \dots , n$$

and

$$P(a_n = n+i) = P(a_n \neq 0, 1, 2, \dots , i-1,\ |B_i| \leq 1,\ |B_{i+1}| > 1,\ \dots\ ,\ |B_{i+n}| > 1)$$

$$\text{for } i = 1, 2, \dots$$

(a.2). Since $\sum_{i \geq 0} P(a_n = i) = 1$, we get $P(a_n = k) \to 0$ as $k \to \infty$, and therefore

$$P(a_n = i) = \sum_{k > i} [P(a_n = k-1) - P(a_n = k)] = \sum_{k>i} c_k(i)\, k\, [P(a_n = k-1) - P(a_n = k)]$$

$$= \sum_{k \geq 1} c_k(i) \cdot k \cdot [P(a_n = k-1) - P(a_n = k)]$$

$$= \sum_{k \geq n+1} c_k(i) \cdot k \cdot [P(a_n = k-1) - P(a_n = k)] \quad \text{(by (a.1))}$$

(a.3). By the definition of a_n and Φ_ℓ, we have, for $\ell > 0$,

$$\Phi_\ell\, a_n = i \Longrightarrow a_n = i+\ell,\ 0,\ 1,\ \dots\ , \text{ or } \ell-1$$

$$\Phi_\ell\, a_n \neq a_{n-\ell} \Longrightarrow a_n = 0,\ 1,\ \dots\ , \text{ or } \ell-1$$

$$\Phi_\ell a_n \neq a_n - \ell \quad \text{for some} \quad \ell = 1, 2, \ldots, m \Longrightarrow a_n = 0, 1, 2, \ldots, \text{or} \quad m-1$$

$$P(\Phi_\ell a_n \neq a_n - \ell \quad \text{for some} \quad \ell = 1, 2, \ldots, m) \leq \sum_{i=0}^{m-1} P(a_n = i) \leq \frac{m}{n}$$

(by (a.1)).

Similarly

$$P(\Phi_\ell a_n \neq a_n - \ell \quad \text{for some} \quad \ell = -1, -2, \ldots, -m) \leq \frac{m}{n} ,$$

and so

$$P(\Phi_\ell an = a_n - \ell \quad \text{for every} \quad \ell = 0, \pm 1, \pm 2, \ldots, \pm m) > 1 - \frac{2m}{n} .$$

Proof of approximation theorem

Case 1. x_k is ergodic and bounded $(|x_k| \leq M)$.

Consider two processes

$$y_k^{(n)} = x_{-a_n(\omega)+k}(\omega_0^*)$$

$$z_k^{(n)} = x_{-\Phi_k a_n(\omega)}(\omega_0^*) = \Phi_k\, x_{-a_n(\omega)}(\omega_0^*)$$

with the element ω_0^* picked up in Lemma 1 and the random sequence $a_n(\omega)$ defined in Lemma 2, where Φ_k is the shift operator acting on the space of functions measurable $\mathcal{B}(B)$.

By Lemma 2 (a.3) we have

$$P(y_k^{(n)} = z_k^{(n)},\ k = 0, \pm 1, \pm 2, \dots, \pm m) \geq 1 - \frac{2m}{n}$$

and so

$$\rho_V(y^{(n)}, z^{(n)}) \longrightarrow 0 \qquad \text{as} \quad n \to \infty$$

a fortiori

(4) $$\rho_U(y^{(n)}, z^{(n)}) \longrightarrow 0 \qquad \text{as} \quad n \to \infty$$

On the other hand we have, by Lemma 2 (a.2) and Lemma 1

(a.2), $$E(\exp(i \sum_{j=1}^{m} \theta_j y_k^{(n)}(\omega)))$$

$$= \sum_{p\geq 0} (n+p) [P(a_n = n+p-1) - P(a_n = n+p)]$$

$$\times \frac{1}{n+p} \sum_{\ell=0}^{n+p} \exp[i \sum_{i=1}^{m} \theta_j x_{k_j-\ell} (\omega^*)]$$

$$\longrightarrow \quad E[\exp(i \sum_{i=1}^{m} \theta_j x_k(\omega^*))]$$

for every $(m, k_1, k_2, \ldots , k_m, \theta_1, \ldots , \theta_m)$. Furthermore this convergence is uniform in the set $S(C)$ of all $(m, k_., \theta_.)$ such that

(5) $$m, \; |k_1|, \ldots , |k_m|, \; |\theta_1|, \ldots , |\theta_m| \leq C \; ;$$

in fact the power of the set $S_\epsilon(C)$ of all $(m, k_., \theta'_. = \ell_. \delta) \in S(C)$ $(\delta = \epsilon/6MC)$ is finite and so we can find $n_0 = n_0(\epsilon)$ such that $n > n_0$ implies

$$E[\exp(i \sum_{j=1}^{m} \ell_i\delta \; y_{k_j}^{(n)}(\omega))] - E[\exp(i \sum_{j=1}^{m} \ell_i\delta \; x_{k_j}^{(n)} (\omega^*))] < \epsilon/3$$

for every $(m, k_., \ell_.\delta) \in S_\epsilon(C)$. Take any $(m, k_., \theta_.) \in S(C)$ and choose $(m, k_., \ell_.\delta)$ such that $|\ell_i\delta - \theta_i| < \delta$, $i = 1, 2, \ldots , m$. Then

$$|E[\exp(i \sum_{j=1}^{m} \theta_j\, y_{k_j}^{(n)}(\omega))] - E[\exp(i \sum_{j=1}^{m} \theta_j\, x_{k_j}(\omega^*))]| < \epsilon$$

by the same argument as in page VIII.11. Therefore

(6) $$\rho_U(y^{(n)}, x) \longrightarrow 0 \qquad \text{as} \qquad n \to \infty$$

which, combined with (4), implies

(7) $$\rho_U(z^{(n)}, x) \longrightarrow 0 \qquad \text{as} \qquad n \to \infty.$$

It is now enough to approximate the process $z_k^{(n)}$, i.e., the process of the form $\Phi_k f(\omega)$ $(|f| \leq M)$ by polynomial stationary sequences.

By Cameron-Martin's approximation theorem (expansion) we can find a sequence of random variables of the form

$$f_n(\omega) = p_n(B_{-N_n}, B_{-N_n+1}, \dots, B_{N_n})$$

$$p_n : \text{polynomial}$$

such that

$$E[|f_n(\omega) - f(\omega)|^2] \longrightarrow 0$$

Then $\Phi_k f_n(\omega)$ is a polynomial stationary sequence and

$$E(|\Phi_k f_n(\omega) - \Phi_k f(\omega)|^2) = E(|f_n(\omega) - f(\omega)|^2) \longrightarrow 0 ,$$

which implies

$$\rho_V[\Phi_\cdot f_n, \Phi_\cdot f] \longrightarrow 0$$

a fortiori

$$\rho_U(\Phi_\cdot f_n, \Phi_\cdot f) \longrightarrow 0 .$$

This completes the proof of Case 1.

Case 2. x_k is bounded ($|x_k| \leq M$).

To prove the approximation theorem in this case it is enough to approximate a bounded stationary sequence by bounded ergodic stationary sequences,

We shall use the following

Theorem: (Oxtoby-Ulam):

Let $\Omega(\mathcal{B},P)$ be a Lebesgue probability measure space (= separable atomless probability measure space), $\mathcal{G}$ the group of all automorphisms and $\mathcal{G}_\epsilon$ that of all ergodic automorphisms. Then

$$\forall T \in \mathcal{G} \quad \forall \epsilon > 0 \quad \forall E_1, E_2, \dots, E_n \in \mathcal{B}$$

$$\exists S = S(T, \epsilon, E_1, \dots, E_n) \in \mathcal{G}_\epsilon \text{ such that } P(TE_i \sim SE_i) < \epsilon$$

First notice that it follows at once from this theorem that

Corollary: Under the same assumptions as above, it holds that

(a) $\forall T \in \mathcal{G} \quad \forall \epsilon > 0, \quad \forall f_1, \dots, f_n \in L^2 = L^2(\Omega, \mathcal{B}, P)$
$\exists S = S(T, \epsilon, f_1, \dots, f_n) \in \mathcal{G}_\epsilon$ such that

$$\|Tf_k - Sf_k\| < \epsilon, \qquad k = 1, 2, \dots, n$$

(Here T means the function-transformation induced by the set-transformation T, and similarly for S.)

(b) $\forall T \in \mathcal{G}, \quad \forall \epsilon > 0, \quad \forall f \in L^2 \quad \forall$ n positive integer.
$\exists\ S = S(T, \epsilon, f, n) \in \mathcal{G}_\epsilon$ such that

$$\|S^k f - T^k f\| < \epsilon \qquad k = 0, \pm 1, \ldots, \pm n$$

Proof: (a) is easy. To prove (b), observe

$$\|S^k f - T^k f\|$$

$$\leq \|S^k f - S^{k-1} Tf\| + \|S^{k-1} Tf - S^{k-2} T^2 f\| + \cdots + \|ST^{k-1} f - T^k f\|$$

$$\leq \|Sf - Tf\| + \|S(Tf) - T(Tf)\| + \cdots + \|S(T^{k-1} f) - T(T^{k-1} f)\|$$

and

$$\|S^{-k} f - T^{-k} f\|$$

$$= \|f - S^k T^{-k} f\|$$

$$= \|S^k T^{-k} f - T^k T^{-k} f\|$$

$$\leq \|S(T^{-k} f) - T(T^{-k} f)\| + \|S(T^{-k+1} f) - T(T^{-k+1}) f\| + \cdots + \|S(T^{-1} f) - T(T^{-1} f)\|$$

and define

$$S \equiv S(T, \epsilon, f, n) = S(T, \frac{\epsilon}{n}, T^{-n} f, T^{-n+1} f, \ldots, T^n f) \text{ in (a).}$$

Let $x_k(\omega)$, $\omega \in \Omega(\mathcal{B}, P)$, be any given bounded stationary sequence. We can assume that $\mathcal{B} = \mathcal{B}(x)$. Therefore $\Omega(\mathcal{B}, P)$ is a separable probability measure space. Let T be an automorphism which carries x_k to x_{k+1}. Then

$$x_k = T^k f \qquad \text{where} \qquad f(\omega) = x_0(\omega).$$

Now we shall construct a Lebesgue probability space $\tilde{\Omega}(\tilde{\mathcal{B}},\tilde{P})$, an automorphism $\tilde{T}$ on it and a bounded function $\tilde{f}(\tilde{\omega})$ such that $\tilde{x}_k(\tilde{\omega}) \equiv \tilde{T}^k \tilde{f}(\tilde{\omega})$ is equivalent in law to $x_k(\omega)$ as processes.

Let $\mathfrak{A}$ be the class of all atoms. $\mathfrak{A}$ is a disjoint system, clearly countable and invariant under T. Therefore $\mathfrak{A}$ is of the form

$$\mathfrak{A} = \{A_1, TA_1, \ldots, T^{\ell_1 - 1}A_1, A_2 TA_2, \ldots, T^{\ell_2 - 1}A_2, \ldots\}$$

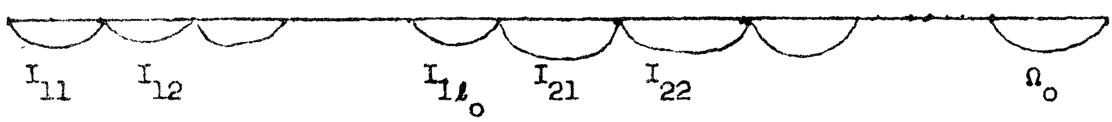
$$(*) \qquad T^{\ell_i}A_i = A_i$$

Therefore we get a disjoint decomposition of Ω:

$$\Omega = (\bigcup_{ij} T^i A_j) \cup \Omega_0 \qquad \Omega_0 : \text{atomless}$$

We shall define $\tilde{\Omega}$ as follows

$$\tilde{\Omega} = (\bigcup_{ij} I_{ij}) \vee \Omega_0 \; ; \qquad |I_{ij}| = P(A_i); \qquad I_{ij} : \text{interval}$$



$\tilde{T}$ is defined as follows: $T = \tilde{T}$ on Ω_o $(= \tilde{\Omega}_o)$.

$$\tilde{T}x = x + P(A_i) \quad (\text{mod } \ell_i P(A_i))$$

$$\tilde{T} : I_{i1} \dashrightarrow I_{i2}$$

$$I_{i2} \dashrightarrow I_{i3} \qquad \tilde{T}^{\ell_i} = \text{identity}$$

$$I_{i\ell_i-1} \dashrightarrow I_{i1}$$

$\tilde{f}$ is defined on $\tilde{\Omega}$, $(\tilde{\mathcal{B}}, \tilde{P})$ will be defined in natural way

$$\tilde{f}(\tilde{\omega}) = f(\tilde{\omega}) \qquad \text{on } \Omega_o$$

$$= f(T^j A_i) \qquad \text{on } I_{ij},\ j = 0, 1, \ldots, \ell_{i-1}$$

Then there exists a homomorphism $U : \mathcal{B} \xrightarrow{(\text{into})} \tilde{\mathcal{B}}$.

$$U : T^j A_i \dashrightarrow I_{ij} \qquad f \to \tilde{f}$$

$$\Omega_o \dashrightarrow \Omega_o \qquad T \to \tilde{T}$$

Therefore $(\tilde{T}^k \tilde{f}(\tilde{\omega}),\ k \in Z) \underset{(L)}{\sim} (T^k f(\omega),\ k \in Z)$.

Thus we can assume that $\Omega(\mathcal{B}, P)$ is a Lebesgue probability measure space and T is an automorphism on Ω and

$$x_k(\omega) = T^k f(\omega) \qquad f : \text{bounded meas.}$$

Using (b) proved above, there exists $S_n \in \mathcal{G}_\epsilon$ for every n such that

$$\|S_n^k f - T^k f\| < \frac{1}{n} \qquad \text{for} \qquad |k| \leq n$$

i.e.

$$\rho_V(S_n^{\cdot} f, T^{\cdot} f) < \frac{1}{n} \longrightarrow 0$$

and so

$$\rho_U(S_n^{\cdot} f, T^{\cdot} f) \longrightarrow 0$$

Since S_n is ergodic, $y_k \equiv S_n^k f$, $k \in Z$, is also ergodic, This completes the proof of case 2.

Case 3. General case.

This can be reduced to case 2 by truncation.