# Ergodic Theorems For Stationary and Strictly Stationary Processes

BY CRAMER & LEADBETTER

## Table of contents



## 1 Ergodic Theorems: Stationary Processes

Let $\xi(t)$ now be a complex-valued stationary process with zero mean, covariance function $r(t)$, and the spectral representation (see equation (?)). The ergodic theorems of Section ? (see equation (?)) yield directly the following results:

1. If we have, as $T \to \infty$,

$$\frac{1}{T}\int_0^T r(t)\ d t \to 0, \tag{1}$$

then the time average

$$\frac{1}{T}\int_0^T \xi(t)\ d t \tag{2}$$

tends in quadratic mean to zero.

2. If, moreover, $r(t) = O(|t|^{-\beta})$ for some $\beta > 0$, as $|t| \to \infty$, the average in equation (2) tends to zero with probability one (see also [?]).

We observe that, by a formula for characteristic functions given in equation (?), the relation in equation (1) will be satisfied if and only if the spectral distribution function $F(\lambda)$ is continuous at the origin.

However, from the spectral representation of $\xi(t)$, further information concerning the quadratic mean behaviour of the time average (2) and certain related averages can be obtained. For any $\xi(t)$ such that the representations (?) and (?) hold, the following relations, which are formally identical with the corresponding relation for characteristic functions given in equation (?), just as was the case in respect of the inversion formulas (?) and (?), hold:

$$\frac{1}{T}\int_0^T \xi(t)\, e^{-\mu i t}\, d\,t \xrightarrow{\text{q.m.}} \Delta\,\zeta(\mu) = \zeta\,(\mu+0) - \zeta\,(\mu-0), \tag{3}$$

$$\frac{1}{T}\int_0^T r(t)\, e^{-\mu i t}\, d\,t \to \Delta\,F(\mu) = \mathbb{E}\,|\Delta\,\zeta(\mu)|^2. \tag{4}$$

Here $\mu$ is any real number, and $\Delta\,\zeta(\mu)$ denotes the jump of the spectral process at the point $\mu$, which is a random variable, while $\Delta\,F(\mu)$ is the corresponding jump of $F$, which is a nonnegative constant. Thus, if, in particular, the spectral distribution is continuous at $\mu$, the time averages in the first members of equations (3) and (4) both have the quadratic mean limit zero.

The second relation (4) is equivalent to the corresponding formula of Section ?, while the first formula is readily proved by the same kind of argument as indicated in respect of equation (?).

## 2 Ergodic Theorems: Strictly Stationary Processes

In the case of a strictly stationary process, it is possible to obtain much deeper results than those given in the preceding section. We shall here only deal with the real-valued case, and recall that, in particular, a real normal and stationary process is always strictly stationary (see Section ?).

In Section ? some introductory remarks were given concerning a "flow" in the probability space, or a one-parameter group of one-one transformations of this space onto itself. We now follow up these remarks and apply them to the present case. Without going deeper into ergodic theory, we shall only prove some results required for the applications to be made later in this book.

Let $\xi(t)$ be a given real-valued and strictly stationary process. For the proof of the Birkhoff-Khintchine ergodic theorem (see p. 5), which is the main result of this section, we shall have to impose certain additional conditions on $\xi(t)$. However, until further notice we shall be concerned with a perfectly general, real-valued, and strictly stationary $\xi(t)$.

We now choose as our probability space the space $(X, \mathcal{B}, \Pi)$ considered in Section ?, in connection with the proof of the Kolmogorov theorem. $X$ is the space of all finite and real-valued functions $x(t)$ of the real variable $t$. $\mathcal{B}$ is the $\sigma$-field of Borel sets in the space $X$, that is, the smallest $\sigma$-field over the intervals of $X$ (see Section ?). Finally, $\Pi$ is the probability measure uniquely determined on all sets of $\mathcal{B}$ by the finite-dimensional distributions of the given strictly stationary process $\xi(t)$.

An elementary event, or a point $\omega$ of the probability space, is now an individual sample function, say $\omega = x(t)$. We define a *shift transformation* $U_\tau$ in the probability space taking $\omega = x(t)$ into $\omega_\tau = x\,(t+\tau)$. Similarly, $U_\tau$ takes any set $S$ of functions $x(t)$ into the set $S_\tau$ formed by the shifted functions $x\,(t+\tau)$.

It is evident that the transformations $U_\tau$, for all real $\tau$, form a group: $U_{\tau+\rho} = U_\tau U_\rho$. It also follows from the strict stationarity of the finite-dimensional distributions defining $\Pi$ that the transformations $U_\tau$ are *measure-preserving*, that is, for any Borel set $S$ of sample functions we have $\Pi(S_\tau) = \Pi(S)$ for all real $\tau$. The shift transformations $U_\tau$ thus determine a *measure-preserving flow* in the probability space $(X, \mathcal{B}, \Pi)$.

A Borel set $S$ is called an *invariant set* of the $\xi(t)$ process if, for every fixed $\tau$, the sets $S$ and $S_\tau$ differ, at most, by sets of $\Pi$-measure zero. That is, the so-called symmetric difference between $S$ and $S_\tau$ should be a set of $\Pi$-measure zero, or in other words we should have

$$S_\tau = (S + N_1) - N_2, \tag{5}$$

where $N_1$ and $N_2$ are $\Pi$-null sets, which may depend on $\tau$. The invariant sets form a $\sigma$-field contained in $\mathcal{B}$. It is readily seen that all sets of probability $\Pi = 0$ or $\Pi = 1$ are invariant.

**Definition 1.** *The strictly stationary process $\xi(t)$ will be called* ergodic *or* metrically transitive *if the $\sigma$-field of invariant sets only contains sets of probability* 0 *or* 1. *Since all such sets are in any case invariant, the property of ergodicity implies that these are the only invariant sets of the process. Any set of probability between zero and one will be effectively "mixed" with other sets under the shift transformation.*

Any random variable $\eta$ defined by means of the random variables $\xi(t)$ for any values of $t$ will be called a random variable defined on the $\xi(t)$-process. Then $\eta$ will be a $\mathcal{B}$-measurable function of the elementary event $\omega = x(t)$, and we define the shift transform of $\eta$ by writing $U_\tau \eta = \eta(\omega_\tau)$. We call $\eta$ an *invariant random variable* of the $\xi(t)$-process if, for every fixed $\tau$, the random variables $\eta$ and $U_\tau \eta$ are equivalent (see Section ?), that is, if $U_\tau \eta = \eta$ with probability one. If $\eta$ is an invariant random variable, and $A$ is any Borel set of real numbers, it is readily seen that the set of all $\omega$ in the probability space such that $\eta(\omega) \in A$ will be an invariant set. It follows that, in the particular case when the $\xi(t)$-process is ergodic, any invariant random variable must be constant with probability one.

For any random variable $\eta$ defined on the $\xi(t)$-process, the family of its shift transforms

$$\eta(\tau, \omega) = U_\tau \eta = \eta(\omega_\tau) \tag{6}$$

for all real $\tau$ is seen to determine a new strictly stationary process. To every fixed elementary event $\omega$ in the probability space $(X, \mathcal{B}, \Pi)$, that is, to every individual sample function $\omega = x(t)$ of the $\xi(t)$-process, there corresponds a uniquely determined sample function

$$y(\tau, \omega) = \eta\left[x\left(t + \tau\right)\right] \tag{7}$$

of the $\eta$ process. As usual we write briefly $\eta(\tau)$ instead of $\eta(\tau, \omega)$.

We may regard $\omega' = y(\tau)$ as an elementary event of the $\eta(\tau)$-process. The transformation which takes $\omega = x(t)$ into $\omega' = y(\tau)$ thus gives a transformation of the original probability space $(X, \mathcal{B}, \Pi)$ into a new probability space (see Section ?), say $(X', \mathcal{B}', \Pi')$. Here $X'$ is the same function space as $X$, and $\mathcal{B}'$ is a $\sigma$-field of sets in $X'$ including all Borel sets, while the probability measure $\Pi'$ corresponding to the $\eta(\tau)$ process is defined for any set $S' \in \mathcal{B}'$ by the relation (see Section ?)

$$\Pi'(S') = \Pi(S), \tag{8}$$

where $S$ is the inverse image of $S'$. From the relation (7) between $y(\tau)$ and $x(t)$, it is seen that $y\left(\tau + h\right)$ corresponds to $x\left(t + h\right)$, so that the shift transformation is the same in both spaces.

Suppose now that $S' \in \mathcal{B}'$ is an invariant set of the $\eta(\tau)$-process, so that we have for every $\tau$

$$S'_\tau = (S' + N'_1) - N'_2, \tag{9}$$

where $N'_1$ and $N'_2$ are $\Pi'$-null sets. We then have the corresponding relation between the inverse images

$$S_\tau = (S + N_1) - N_2, \tag{10}$$

and it follows from equation (8) that $N_1$ and $N_2$ are $\Pi$-null sets, so that $S$ is invariant for the $\xi(t)$-process if and only if $S'$ is an invariant set of the $\xi(t)$-process. If, in particular, the $\xi(t)$-process is ergodic, $\Pi(S)$ must be either 0 or 1, and it follows from equation (?) that the same holds true for $\Pi'(S')$, so that the $\eta(\tau)$-process is also ergodic. We have thus proved the following proposition:

**Proposition 2.** *Let $\xi(t)$ be strictly stationary and ergodic, while $\eta$ is a random variable defined on the $\xi(t)$-process. The stochastic process generated by the shift transforms $\eta(\tau) = \eta(\tau, \omega)$ defined by equation (?) for all real $\tau$ is then strictly stationary and ergodic.*

For a later purpose, we observe that, if $\eta_1(\tau), \ldots, \eta_k(\tau)$ are processes defined in the same way as $\eta(\tau)$ above, the joint distribution of the $\eta_j(\tau)$ will obviously show the same invariance under a translation in $\tau$. Some examples of ergodic and nonergodic processes will be given later in this section.

From now on, we shall suppose that our strictly stationary process $\xi(t)$ satisfies the following two additional conditions:

1. $\mathcal{E}\,|\xi(0)| < \infty$.

2. With probability one, the sample functions of the $\xi(t)$-process are Riemann integrable on every finite interval. (See the criteria given in 4.2, 4.4, and 5.4.)

We observe that, while condition 1 is an essential condition, condition 2 could be replaced by a more general condition involving measurability. Condition 2 as given above seems, however, to be sufficiently general for most applications.

For processes satisfying these conditions, we now proceed to prove the main result of this section, the famous Birkhoff-Khintchine ergodic theorem:

**Theorem 3. (Birkhoff-Khintchine Ergodic Theorem)** *If $\xi(t)$ is strictly stationary and satisfies the conditions 1 and 2, each of the time averages*

$$\frac{1}{T}\int_{-T}^{0}\xi(t)\;dt \quad \text{and} \quad \frac{1}{T}\int_{0}^{T}\xi(t)\;dt \tag{11}$$

*converges with probability one to an invariant random variable of the $\xi(t)$-process, as $T\to\infty$. If, in particular, $\xi(t)$ is ergodic, both limits are, with probability one, equal to the constant $\mathcal{E}\,\xi(0)$.*

According to this theorem we have for an ergodic process, with probability one,

$$\lim_{T\to\infty}\frac{1}{T}\int_{-T}^{0}\xi(t)\;dt=\mathcal{E}\,\xi(0), \tag{12}$$

and the corresponding relation for the time average over $(0,T)$. Here the second member is an average value of the random variable $\xi(0)$, extended over the set of all possible sample functions of the process. On the other hand, the first member is the limit of a time average, extended over the values assumed in the past by one individual sample function $\xi(t)$. Thus, the theorem asserts the equality, with probability one, of the *ensemble average* in the second member and the *time average* in the first member, the latter bearing on one single realization of the process. In this assertion lies the physical significance of the theorem.

The first part of the proof of the ergodic theorem that will now be given is a modified form of a proof due to Kolmogorov [Kolmogorov] (see also Gnedenko [Gnedenko]). It will be sufficient to give the proof for the time average over the positive part of the $t$ axis. Writing, for $n=\ldots,-1,0,1,\ldots$,

$$\xi_n=\int_{n}^{n+1}\xi(t)\;dt, \tag{13}$$

it will be seen that the sequence of the $\xi_n$ forms a strictly stationary process with discrete time parameter $n$, such that

$$\mathcal{E}\,|\xi_n|=c<\infty. \tag{14}$$

Writing further, for any $m \le n$,

$$\eta_{m,n} = \frac{\xi_m + \xi_{m+1} + \cdots + \xi_n}{n - m + 1} \tag{15}$$

we shall first show that a finite limit

$$\lim_{n \to \infty} \eta_{0,n} \tag{16}$$

exists with probability one. By Fatou's lemma we have

$$\mathcal{E} \operatorname{liminf} |\eta_{0,n}| \le \operatorname{liminf} \mathcal{E} |\eta_{0,n}| \le c, \tag{17}$$

which shows that liminf $|\eta_{0,n}|$ is finite with probability one. It follows that, with probability one, the sequence $\eta_{0,n}$ has at least one finite limit point.

Let us now assume that, with a positive probability, $\eta_{0,n}$ has not a definite limit. Some reflection will then show that it is possible to find two finite numbers $\alpha$ and $\beta$ such that $\alpha < \beta$, and that with a positive probability both the inequalities

$$\operatorname{limsup} \eta_{0,n} > \beta, \tag{18}$$
$$\operatorname{liminf} \eta_{0,n} < \alpha, \tag{19}$$

are satisfied. If $S$ is the set of all elementary events $\omega$ [that is, sample functions of the $\xi(t)$-process] such that both inequalities (18) and (19) hold, we shall thus have $\Pi(S) > 0$, and we now have to show that this assumption implies a contradiction.

Consider any fixed sample sequence $x_n$ of the discrete $\xi_n$ process and write

$$y_{m,n} = \frac{x_m + x_{m+1} + \cdots + x_n}{n - m + 1}. \tag{20}$$

We shall say that $(m, n)$ is a *characteristic interval* of length $n - m$ for the given $x_n$ sequence if we have

$$y_{m,i} \le \beta \quad \text{for } i = m, m+1, \ldots, n-1, \tag{21}$$
$$y_{m,n} > \beta \tag{22}$$

Note that in the case $n = m$ only the second condition is relevant, and $(m, n)$ is then a characteristic interval of length zero. Suppose now that $(m_1, n_1)$ and $(m_2, n_2)$ are two distinct characteristic intervals. Obviously, we cannot have $m_1 = m_2$, so that we may assume for example, $m_1 < m_2$. It will be shown that the intervals cannot overlap, in the sense that we cannot have

$$m_1 < m_2 < n_1 < n_2. \tag{23}$$

In fact, the identity

$$y_{m_1,n_1} = \frac{(m_2 - m_1)\, y_{m_1,m_2-1} + (n_1 - m_2 + 1)\, y_{m_2,n_1}}{n_1 - m_1 + 1} \tag{24}$$

shows that, in order to make the first member $>\beta$, at least one of the $y$'s appearing in the second member must be $>\beta$. But it is easily seen that this is impossible if equation (23) holds. Thus, the two intervals cannot overlap.

Suppose further that each of the two characteristic intervals just considered contains a third characteristic interval $(m,n)$, without being identical with it. For the same reason as before, we must then have $m_1 < m_2 < m$, and since the intervals cannot overlap we obtain

$$m_1 < m_2 < m \leq n \leq n_2 \leq n_1. \tag{25}$$

It follows in particular that $n_2 - m_2 < n_1 - m_1$, so that any two distinct characteristic intervals containing a third must have unequal lengths, the shorter interval being contained in the longer. Thus, all characteristic intervals containing a given characteristic interval will form a strictly increasing sequence of intervals.

Hence, if $I = (m,n)$ is a given characteristic interval, and if $N \geq n - m$ is a given integer, there will always be a uniquely determined largest characteristic interval $I_N$ of length $\leq N$ containing $I$. (It is, of course, possible that we may have $I_N = I$, if every characteristic interval strictly containing $I$ has a length exceeding $N$.)

Any elementary event $\omega$ belonging to the set $S$ will correspond to a definite sample sequence of the $\xi_n$-process. By the first inequality (18), this sample sequence will have a uniquely determined characteristic interval $(0,n)$ of length $n \geq 0$. Let $S_N$ for every $N \geq 0$ denote the subset of $S$ formed by all $\omega$ such that $0 \leq n \leq N$. Then $S_N$ is never decreasing as $N$ increases, and $S_N \to S$ as $N \to \infty$. Consequently, $\Pi(S_N) \to \Pi(S)$, so that by hypothesis we must have $\Pi(S_N) > 0$ for all sufficiently large $N$. In the sequel, we only consider values of $N$ such that $\Pi(S_N) > 0$.

For any $\omega \in S_N$, it now follows from the above that there will be a uniquely determined largest characteristic interval $I_N$ of length $\leq N$ containing the characteristic interval $(0,n)$, and possibly coinciding with it. Let $R_{pq}$ be the subset of $S_N$ such that

$$I_N = (-p, -p+q). \tag{26}$$

Then we must have $0 \leq p \leq q$, $0 \leq q \leq N$, and two different $R_{pq}$ will always be disjoint, so that

$$\Pi(S_N) = \sum_{p=0}^{N} \sum_{q=0}^{N} \Pi(R_{pq}). \tag{27}$$

From the strict stationarity of the $\xi_n$-sequence we obtain

$$\Pi(R_{pq}) = \Pi(R_{0q}), \tag{28}$$

and, if $\Pi(R_{0q}) > 0$,

$$\mathcal{E}(\xi_0 | R_{pq}) = \mathcal{E}(\xi_p | R_{0q}). \tag{29}$$

Observing that, for every $\omega \in R_{0\,q}$, we must have $\eta_{0\,q} > \beta$, we now obtain, excluding from summation every term with $\Pi(R_{pq}) = 0$,

$$\begin{aligned}
\Pi(S_N)\ \mathcal{E}(\xi_0|\, S_N) &= \sum_q \sum_p \Pi(R_{pq})\ \mathcal{E}(\xi_0|\, R_{pq}) && (30)\\
&= \sum_q \Pi(R_{0\,q}) \sum_p \mathcal{E}(\xi_p|\, R_{0\,q}) && (31)\\
&= \sum_q \Pi(R_{0\,q})\,(q+1)\,\mathcal{E}(\eta_{0\,q}|\, R_{0\,q}) && (32)\\
&> \sum_q \Pi(R_{0\,q})\,(q+1)\,\beta && (33)\\
&= \beta \sum_q \sum_p \Pi(R_{pq}) && (34)\\
&= \beta\,\Pi(S_N). && (35)
\end{aligned}$$

Thus, for all sufficiently large $N$,

$$\mathcal{E}(\xi_0|\, S_N) > \beta, \tag{36}$$

and consequently, as $N \to \infty$,

$$\mathcal{E}(\xi_0|\, S) \geq \beta. \tag{37}$$

However, arguing in the same way from the second inequality (19), and changing the sign of all relevant inequalities, one obtains

$$\mathcal{E}(\xi_0|\, S) \leq \alpha < \beta. \tag{38}$$

Thus, the assumption $\Pi(S) > 0$ implies a contradiction, and it follows that a finite limit (16) exists with probability one.

**Lemma 4.** *The sequence*

$$\eta_{0,n-1} = \frac{1}{n} \int_0^n \xi(t)\ d\,t \tag{39}$$

*tends with probability one to a certain random variable, say $\eta$, as $n \to \infty$ through integral values.*

For $n \leq T < n+1$,

$$\frac{1}{T} \int_0^T \xi(t)\ d\,t = \frac{n}{T} \cdot \frac{1}{n} \int_0^n \xi(t)\ d\,t + \frac{1}{T} \int_n^T \xi(t)\ d\,t. \tag{40}$$

The first term in the second member obviously tends to $\eta$ with probability one. The last term in the second member is bounded in absolute value by

$$\frac{1}{n} \int_n^{n+1} |\xi(t)|\ d\,t. \tag{41}$$

But $|\xi(t)|$ is strictly stationary and satisfies the conditions 1 and 2. Replacing in the preceding proof $\xi(t)$ by $|\xi(t)|$, it follows that the limit analogous to (16) exists, and consequently (41) tends to zero with probability one. Thus, the limiting relation

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \xi(t)\; dt = \eta \tag{42}$$

holds with probability one.

Obviously, for every real $\tau$ with probability one,

$$\lim_{T\to\infty} \frac{1}{T} \int_\tau^{\tau+T} \xi(t)\; dt = \lim_{T\to\infty} \frac{1}{T} \int_0^T \xi(t)\; dt, \tag{43}$$

so that the limit $\eta$ is an invariant random variable of the $\xi(t)$-process. In particular, if $\xi(t)$ is ergodic, $\eta$ must be constant with probability one, say $\eta = c$.

It remains to show that $c = \mathcal{E}\,\xi(0)$. To prove this, one may assume $\mathcal{E}\,\xi(0) = 0$. Writing

$$\eta_T = \frac{1}{T} \int_0^T \xi(t)\; dt, \tag{44}$$

one has $\mathcal{E}\,\eta_T = 0$ for all $T$, and $\eta_T \to c$ with probability one. For any given $\delta > 0$, by condition 1 and the strict stationarity of $\xi(t)$, there exists $\epsilon > 0$ such that

$$\int_S |\xi(t)|\; d\Pi < \delta \tag{45}$$

for every $t$, as soon as $\Pi(S) < \epsilon$ (see Loève [Loeve], p. 124). It follows that then also

$$\int_S |\eta_T|\; d\Pi < \delta. \tag{46}$$

Obviously, one may here take $\epsilon < \delta$. Now assume, for example, $c > 0$. As convergence with probability one implies convergence in probability (see Section ?), one has

$$\Pi\,(\eta_T > c/2) \geq 1 - \epsilon \tag{47}$$

for all sufficiently large $T$. Denoting by $S$ the set of all elementary events $\omega$ such that $\eta_T > c/2$, then $\Pi(S) \geq 1 - \epsilon > 1 - \delta$, and $\Pi(S^*) < \epsilon$. Consequently,

$$0 = \mathcal{E}\,\eta_T = \int_S \eta_T\; d\Pi + \int_{S^*} \eta_T\; d\Pi \tag{48}$$

$$> \Pi(S)\frac{c}{2} - \delta \tag{49}$$

$$> (1-\delta)\,\frac{c}{2} - \delta. \tag{50}$$

For sufficiently small $\delta$, this implies a contradiction, so that $c=0$. This completes the proof of the Birkhoff-Khintchine ergodic theorem.

## 3 Examples: Ergodic and Nonergodic Processes

Consider the class of normal stationary processes. Let $\xi(t)$ be real-valued, normal, and stationary, with zero mean and such that, with probability one, its sample functions are continuous over any finite interval. (Sufficient conditions for sample function continuity in the present case will be given in Sections ? and ?.) The conditions 1 and 2 of the Birkhoff-Khintchine theorem are then certainly satisfied. For such a $\xi(t)$, and for the corresponding covariance function $r(t)$, the spectral representations (?) and (?) hold. A proposition due to Maruyama [Maruyama] and Grenander [Grenander] asserts that, under these conditions, $\xi(t)$ *will be ergodic if and only if the spectral distribution function* $G(\lambda)$ *of (?) is everywhere continuous.*

To prove the necessity of this condition, suppose that $\xi(t)$ is ergodic and assume, as one may, $\mathcal{E}\,\xi^2(t)=1$. The spectral d.f. $F(\lambda)$ associated with $G(\lambda)$ according to (?) must then be everywhere continuous.

By the lemma proved above, the process $\xi^2(t)$ is strictly stationary and ergodic, with mean 1. The Birkhoff-Khintchine theorem then asserts that the random variable

$$X(T)=\frac{1}{T}\int_0^T \xi^2(t)\;dt-1 \tag{51}$$

tends to zero with probability one, as $T\to\infty$. By the properties of the multidimensional normal distribution, $X(T)$ has a fourth order moment which remains bounded as $T\to\infty$. It follows that $X(T)$ tends to zero also in quadratic mean, so that $\mathcal{E}\,X^2(T)\to 0$ as $T\to\infty$. One has

$$\mathcal{E}\,X^2(T) =\frac{1}{T^2}\,\mathcal{E}\left[\int_0^T\int_0^T \xi^2(t)\,\xi^2(u)\;dt\;du\right]-1 \tag{52}$$

$$=\frac{2}{T^2}\int_0^T\int_0^t r^2(u)\;du\;dt \tag{53}$$

$$=\frac{4}{T^2}\int_0^T t\;dt\int_0^t r^2(u)\;du. \tag{54}$$

However, by equation (?), the average value of $r^2(u)$ over $(0,t)$ tends to the sum of the squares of the saltuses of $F(\lambda)$ in all its points of discontinuity, as $t\to\infty$. In order that $\mathcal{E}\,X^2(T)$ should tend to zero as $T\to\infty$, it is thus necessary that $F(\lambda)$ should be everywhere continuous.

For sufficiency, follow the lines of Grenander's proof, under the more restrictive condition that $G(\lambda)$ is even *absolutely* continuous. It then follows from (?) by the Riemann-Lebesgue theorem that $r(t)\to 0$ as $|t|\to\infty$.

Let $S$ be an invariant set of the $\xi(t)$-process. The shift transformation $U_\tau$ then takes $S$ into a set $S_\tau$, which differs from $S$ only by sets of $\Pi$-measure zero, so that

$$\Pi\,(S\,S_\tau) = \Pi(S). \tag{55}$$

For any given $\epsilon > 0$, approximate $S$ by a finite interval sum $I$ in the function space $X$ (see Section ?) so that $I = (S + S') - S''$, where $S'$ and $S''$ have $\Pi$-measure $<\epsilon$. Hence,

$$|\Pi(I) - \Pi(S)| < \epsilon, \tag{56}$$
$$|\Pi\,(I\,I_\tau) - \Pi\,(S\,S_\tau)| < 2\,\epsilon, \tag{57}$$

where $I_\tau = U_\tau I$.

Now, $I$ is the set of all sample functions of the $\xi(t)$-process which satisfy certain inequalities of the form (3.3.1), giving bounds for the values of $\xi(t)$ in a finite number of points, say $t_1, \ldots, t_n$. Then $I_\tau$ will be the set of all $\xi(t)$ satisfying the corresponding inequalities with the arguments $t_1 - \tau, \ldots, t_n - \tau$. The random variables $\xi(t_i)$ and $\xi\,(t_j - \tau)$ are all normal, and the covariance functions

$$\mathcal{E}\,[\xi(t_i)\,\xi\,(t_j - \tau)] = r\,(t_i - t_j + \tau) \tag{58}$$

all tend to zero as $|\tau| \to \infty$. Thus, the two groups of normal variables $\xi(t_i)$ and $\xi\,(t_j - \tau)$ tend to be independent, so that

$$\Pi\,(I\,I_\tau) - \Pi(I)\,\Pi(I_\tau) \to 0. \tag{59}$$

Owing to strict stationarity, $\Pi(I_\tau) = \Pi(I)$, and thus

$$\Pi\,(I\,I_\tau) \to \Pi^2(I). \tag{60}$$

Since $\epsilon$ is arbitrary, it finally follows from (55) and (57) that

$$\Pi^*(S) = \Pi(S), \tag{61}$$

which shows that $\Pi(S) = 1$ or $\Pi(S) = 0$. Therefore, every invariant set of the $\xi(t)$ process has probability 0 or 1, and so the process is ergodic.

Thus, every normal $\xi(t)$ with continuous sample functions, obtained from the spectral representation (?) with a continuous spectral d.f. $G(\lambda)$ gives an instance of an ergodic process. Any process $\eta(t)$ generated by the shift transformation from some random variable defined on the $\xi(t)$ process is then, by the first proposition proved above, also strictly stationary and ergodic. As examples we may mention the processes

$$\eta_1(t) = p[\xi\,(t_1 + t), \ldots, \xi\,(t_n + t)], \tag{62}$$
$$\eta_2(t) = \int_t^{t+1} \xi(u)\;d\,u, \tag{63}$$

where $p$ is a polynomial in the arguments indicated. A further example, which will be discussed in detail in Chapters 10 and 11, is the case when $\eta(t)$ is related to the number of crossings between the $\xi(t)$ curve and some given level $u$ in a convenient time interval, say $(t, t+1)$.

# 4 A Simple Example of a Nonergodic Process

In order to give a simple example of a nonergodic process, consider a normal process $\xi(t)$ given by the spectral representation (?) in the particular case when $G(\lambda)$ is constant except for a single jump of magnitude 1 at $\lambda = \lambda_0$. Then (?) gives

$$\xi(t) = u \cos \lambda_0 t + v \sin \lambda_0 t, \tag{64}$$

where the random variables $u$ and $v$ are independent and normal $(0,1)$. Replacing $u$ and $v$ by new variables $\rho$ and $\theta$ by means of the relations $u = \rho \cos \theta$, $v = \rho \sin \theta$, we have

$$\xi(t) = \rho \cos(\lambda_0 t - \theta), \tag{65}$$

while the joint probability density is transformed as follows:

$$\frac{1}{2\pi} e^{-(u^2+v^2)/2} \, du \, dv = \rho \, e^{-\rho^2/2} \, d\rho \cdot \frac{d\theta}{2\pi}, \tag{66}$$

so that $\rho$ and $\theta$ are independent, $\rho > 0$ having the density $\rho \, e^{-\rho^2/2}$, while $\theta$ is uniformly distributed over $(0, 2\pi)$. Thus, we are led back to the case considered at the end of Section ? of a simple harmonic oscillation with constant angular frequency $\lambda_0$, random amplitude $\rho$, and random phase angle $\theta$.

That this is a nonergodic process is seen, for example, from the fact that the set of all sample functions with $a < \rho < b$ is invariant under the shift transformation. The "mixing" properties of this process are not strong enough to ensure ergodicity. Nevertheless, if $\lambda_0 \neq 0$, the time averages considered in the Birkhoff-Khintchine theorem are easily seen to converge with probability one to the limit

$$\mathcal{E} \xi(0) = 0. \tag{67}$$

On the other hand, when $\lambda_0 = 0$, we have $\xi(t) = \rho \cos \theta = u$, so that $\xi(t)$ is independent of $t$, and is identical with its own time average, which is normal $(0,1)$ and thus not equivalent to zero.

# Bibliography


**[Kolmogorov]** Kolmogorov, A.N., *Foundations of the Theory of Probability*.

**[Gnedenko]** Gnedenko, B.V., *The Theory of Probability*.

**[Loeve]** Loève, M., *Probability Theory*.

**[Maruyama]** Maruyama, G., "Continuous Parameter Stationary Processes," Memoirs of the Faculty of Science, Kyushu University, Series A, 1949.

**[Grenander]** Grenander, U., *Stochastic Processes and Statistical Inference*.