ASYMPTOTIC PROPERTIES OF DERIVATIVES OF STATIONARY MEASURES

SHU-TEH C. MOY

1. Introduction. Let X be a non-empty set and $\mathscr{S}$ be a σ-algebra of subsets of X. Consider the infinite product space $\Omega = \prod_{n=-\infty}^{\infty} X_n$ where $X_n = X$ for $n = 0, \pm 1, \pm 2, \cdots$ and the infinite product σ-algebra $\mathscr{F} = \prod_{n=-\infty}^{\infty} \mathscr{S}_n$ where $\mathscr{S}_n = \mathscr{S}$ for $n = 0, \pm 1, \pm 2, \cdots$. Elements of Ω are bilateral infinite sequences $\{\cdots, x_{-1}, x_0, x_1, \cdots\}$ with $x_n \in X$. Let us denote the elements of Ω by w. If $w = \{\cdots, x_{-1}, x_0, x_1, \cdots\}$ x_n is called the nth coordinate of w and shall be considered as a function on Ω to X. Let T be the shift transformation on Ω to Ω: the nth coordinate of Tw is equal to the $n+1$th coordinate of w. For any function g on Ω, Tg is the function defined by $Tg(w) = g(Tw)$ so that $Tx_n = x_{n+1}$ for any integer n. We shall consider two probability measures μ, ν defined on $\mathscr{F}$. For $n = 1, 2, \cdots$ let $\Omega_n = \prod_{i=1}^{n} X_i$ where $X_i = X, i = 1, 2 \cdots, n$ and $\mathscr{F}_n = \prod_{i=1}^{n} \mathscr{S}_i$ where $\mathscr{S}_i = \mathscr{S}, i = 1, 2, \cdots, n$. Then $\Omega_1 = X$ and $\mathscr{F}_1 = \mathscr{S}$. Let $\mathscr{F}_{m\,n}, m \leqq n, n = 0, \pm 1, \pm 2, \cdots$, be the σ-algebra of subsets of Ω consisting of sets of the form

$$[w = \{\cdots, x_{-1}, x_0, x_1 \cdots\} \colon (x_m, x_{m+1}, \cdots, x_n) \in E]$$

Where $E \in \mathscr{F}_{n-m+1}$. Then $\mathscr{F}_{m\,n} \subset \mathscr{F}_{m\,n+1} \subset \mathscr{F}$. Let $\mu_{m\,n}, \nu_{m\,n}$ be the contractions of μ, ν, respectively to $\mathscr{F}_{m\,n}$. If $\nu_{m\,n}$ is absolutely continuous with respect to $\mu_{m\,n}$, the derivative of $\nu_{m\,n}$ with respect to $\mu_{m\,n}$ is a function of $x_m, \cdots, x_n$ and shall be designated by $f_{m\,n}(x_m, \cdots, x_n)$. Since $f_{m\,n}(x_m, \cdots, x_n)$ is positive with ν-probability one $1/f_{m\,n}(x_m, \cdots x_n)$ is well defined with ν-probability one. We shall let the function $1/f_{m\,n}(x_m, \cdots, x_n)$ take on the value 0 when $f_{m\,n}(x_m, \cdots, x_n) \leqq 0$. Thus $1/f_{m\,n}(x_m, \cdots, x_n)$ is well defined everywhere. In fact $1/f_{m\,n}(x_m, \cdots, x_n)$ is the derivative of $\nu_{m\,n}$-continuous part of $\mu_{m\,n}$ with respect to $\nu_{m\,n}$. According to the celebrated theorem of E. S. Anderson and B. Jessen [1] and J. L. Doob ([2]), pp. 343) $1/f_{m\,n}(x_m, \cdots, x_n)$ converges with ν-probability one as $n \to \infty$. If we assume that μ, ν are stationary, i.e., μ, ν are T invariant, more precise results may be expected. A fundamental theorem of Information Theory, first proved by C. Shannon for stationary Markovian measures [5] and later generalized to any stationary measure by B. McMillan [4], may be considered as a theorem of this sort. In their theorem X is assumed to be a finite set. In this paper we shall first treat Markovian stationary measures μ, ν with X being

Received September 2, 1959, and in revised form January 18, 1960. The author is very much indebted to the referee for his suggestions which have been of substantial benefit to this paper.

any set, finite or infinite, and $\mathscr{S}$ any σ-algebra of subsets of X. It will be proved that $n^{-1}\log f_{m\,n}(x_m, \cdots, x_n)$ converges as $n \to \infty$ with ν-probability one and also in $L_1(\nu)$ under some integrability conditions. The case that ν is only stationary is also treated. Similar convergence theorem is proved under the assumption that X is countable.

2. Asymptotic properties of derivatives of a Markovian measure with stationary transition probabilities with respect to another such measure.

Let X, $\mathscr{S}$, Ω, $\mathscr{F}$, Ω_n, $\mathscr{F}_n$, $\mathscr{F}_{m\,n}$, $\mu_{m\,n}$, $\nu_{m\,n} f_{m\,n}(x_m, \cdots, x_n)$ be as in § 1. x_n, $n = 0, \pm 1, \pm 2, \cdots$, are considered as functions or random variables on Ω to X. Notations for conditional prababilities and conditional expectations relative to one or several random variables will be as in [2], chapter 1, § 7. Since we have two probability measures we shall use subscripts μ, ν to indicate conditional probabilities and conditional expectations taken under measures μ, ν respectively. In this section μ, ν *are assumed to be Markovian* i.e., for any $A \in \mathscr{S}$, $m < n$, $n = 0 \pm 1, \pm 2, \cdots$,

(1) $P_\mu[x_n \in A \mid x_m, \cdots, x_{n-1}] = P_\mu[x_n \in A \mid x_{n-1}]$ with μ-probability one and

(2) $P_\nu[x_n \in A \mid x_m, \cdots, x_{n-1}] = P_\nu[x_n \in A \mid x_{n-1}]$ with ν-probability one. For any set $E \subset \Omega$ let I_E be the real valued function on Ω defined by

$$I_E(w) = 1 \text{ if } w \in E$$
$$= 0 \text{ if } w \notin E\,.$$

LEMMA 1. *If $\nu_{n-1\,n}$ is absolutely continuous with respect to $\mu_{n-1\,n}$ then for any $A \in \mathscr{S}$*

$$(3)\qquad P_\nu[x_n \in A \mid x_{n-1}] f_{n-1\,n-1}(x_{n-1}) = E_\mu[I_{(x_n \in A)} f_{n-1\,n}(x_{n-1}, x_n) \mid x_{n-1}] \text{ with } \mu\text{-probability one.}$$

Proof. For any $A, B \in \mathscr{S}$

$$\nu[x_n \in A, x_{n-1} \in B] = \int_{[x_{n-1} \in B]} P_\nu[x_n \in A \mid x_{n-1}]d\nu = \int_{[x_{n-1} \in B]} P_\nu[x_n \in A \mid x_{n-1}] f_{n-1\,n-1}(x_{n-1}) d\mu\,.$$

On the other hand

$$\nu[x_n \in A, x_{n-1} \in B] = \int_{[x_{n-1} \in B]} I_{x_n \in A} f_{n-1\,n}(x_{n-1}, x_n) \mid x_{n-1}) d\mu$$

$$= \int_{[x_{n-1}\in B]} E_\mu[I_{x_n\in A} f_{n-1\,n}(x_{n-1}, x_n) \mid x_{n-1}]d\mu .$$

Hence for any $B \in \mathscr{S}$

$$\int_{[x_{n-1}\in B]} P_\nu[x_n \in A \mid x_{n-1}] f_{n-1\,n-1}(x_{n-1})d\mu$$
$$= \int_{[x_{n-1}\in B]} E_\mu[I_{x_n\in A} f_{n-1\,n}(x_{n-1}, x_n) \mid x_{n-1}]d\mu ,$$

therefore (3) is true with μ-probability one. Dividing both sides of (3) by $f_{n-1\,n-1}(x_{n-1})$ we then have

$$P_\nu[x_n \in A \mid x_{n-1}] = \frac{E_\mu[I_{x_n\in A} f_{n-1\,n}(x_{n-1}, x_n) \mid x_{n-1}]}{f_{n-1\,n-1}(x_{n-1})} . \tag{4}$$

With μ-probability one on the set $[f_{n-1\,n-1}(x_{n-1}) > 0]$. Since $\nu[f_{n-1\,n-1}(x_{n-1}) > 0] = 1$, (4) is true with ν-probability one.

THEOREM 1. *If $\nu_{n-1\,n}$ is absolutely continuous with respect to $\mu_{n-1\,n}$ for $n = 0, \pm 1, \pm 2, \cdots$ then $\nu_{m\,n}$ is absolutely continuous with respect to $\mu_{m\,n}$ for $n = 0, \pm 1, \pm 2, \cdots$ and $m \leqq n$ with*

$$f_{m\,n}(x_m, \cdots, x_n) = f_{m\,m+1}(x_m, x_{m+1}) \frac{f_{m+1\,m+2}(x_{m+1}, x_{m+2})}{f_{m+1\,m+1}(x_{m+1})} \cdots \frac{f_{n-1\,n}(x_{n-1}, x_n)}{f_{n-1\,n-1}(x_{n-1})} \tag{5}$$

with μ-probability one.

Proof. We shall prove the theorem for the case that $m = 1, n = 2, 3, \cdots$. The proof for the general case that m is any integer is similar. Since $\nu_{1\,2}$ is absolutely continuous with respect to $\mu_{1\,2}$ by hypothesis, (5) is trivially true for $m = 1, n = 2$. Suppose $\nu_{1\,k}$ $(k \geqq 2)$ is absolutely continuous with respect to $\mu_{1\,k}$ and $f_{1\,k}(x_1, \cdots, x_k)$ is given by (5) with μ-probability one. For any $A \in \mathscr{S}, B \in \mathscr{F}_k$

$$\nu[x_{k+1} \in A, (x_1, \cdots, x_k) \in B]$$
$$= \int_{[(x_1,\cdots,x_k)\in B]} P_\nu[x_{k+1} \in A \mid x_1, \cdots, x_k]d\nu .$$

Since ν is Markovian and by (4)

$$\nu[x_{k+1} \in A, (x_1, \cdots, x_k) \in B]$$
$$= \int_{[(x_1,\cdots,x_k)\in B]} P_\nu[x_{k+1} \in A \mid x_k]d\nu$$

$$= \int_{[(x_1,\cdots,x_k)\in B]} \frac{E_\mu[I_{x_{k+1}\in A} f_{k\,k+1}(x_k, x_{k+1}) \mid x_k]}{f_{k\,k}(x_k)}\, d\nu$$

$$= \int_{[(x_1,\cdots,x_k)\in B]} \frac{E_\mu[I_{x_{k+1}\in A} f_{k\,k+1}(x_k, x_{k+1}) \mid x_k]}{f_{k\,k}(x_k)} f_{1\,k}(x_1, \cdots, x_k) d\mu\ .$$

Since μ is Markovian

$$E_\mu[I_{x_{k+1}\in A} f_{k\,k+1}(x_k, x_{k+1}) \mid x_k]$$
$$= E_\mu[I_{x_{k+1}\in A} f_{k\,k+1}(x_k, x_{k+1}) \mid x_1, \cdots, x_k]$$

with μ-probability one. Hence

$$\nu[x_{k+1} \in A, (x_1, \cdots, x_k) \in B]$$
$$= \int_{(x_1,\cdots,x_k)\in B} E_\mu\left[I_{x_{k+1}\in A} \frac{f_{k\,k+1}(x_k, x_{k+1})}{f_{k\,k}(x_k)} f_{1\,k}(x_1, \cdots, x_k) \mid x_1, \cdots, x_k \right] d\mu$$
$$= \int_{(x_1,\cdots,x_k)\in B} I_{x_{n+1}\in A} f_{1\,k}(x_1, \cdots, x_k) \frac{f_{k\,k+1}(x_k, x_{k+1})}{f_{k\,k}(x_k)}\, d\mu\ .$$

Hence

$$\nu[x_{k+1} \in A, (x_1, \cdots, x_k) \in B]$$
$$= \int_{[x_{k+1}\in A, (x_1,\cdots,x_k)\in B]} f_{1\,k}(x_1, \cdots, x_k) \frac{f_{k\,k+1}(x_k, x_{k+1})}{f_{k\,k}(x_k)}\, d\mu$$

for any $A \in \mathscr{S}, B \in \mathscr{F}_k$. Hence for any $E \in \mathscr{F}_{1\,k+1}$

$$\nu(E) = \int_E f_{1\,k}(x_1, \cdots, x_k) \frac{f_{k\,k+1}(x_k, x_{k+1})}{f_{k\,k}(x_k)}\, d\mu\ ,$$

Therefore $\nu_{1\,k+1}$ is absolutely continuous with respect to $\mu_{1\,k+1}$ and

$$(6) \qquad f_{1\,k+1}(x_1, \cdots, x_{k+1}) = f_{1\,k}(x_1, \cdots, x_k) \frac{f_{k\,k+1}(x_k, x_{k+1})}{f_{k\,k}(x_k)}$$

with μ-probability one. (6) together with the supposition that (5) holds true for $m = 1, n = k$ implies that (5) holds true for $m = 1, n = k + 1$. Thus the theorem for the case that $m = 1$ is proved.

Any Markovian probability measure on $\mathscr{F}$ is said to have *stationary transition probabilities* if E being a set of probability one implies that TE, $T^{-1}E$ are also of probability one and for any $A \in \mathscr{S}$ and any n

$$P[x_{n+1} \in A \mid x_n] = TP[x_n \in A \mid x_{n-1}]$$

with probability one. Thus for a Markovian probability measure with stationary transition probabilities we have for any pair of integers m, n and any $A \in \mathscr{S}$

(7) $P[x_n \in A \mid x_{n-1}] = T^{n-m}P[x_m \in A \mid x_{m-1}]$ with probability one and

(8) $E[g(x_{n-1}, x_n) \mid x_{n-1}] = T^{n-m}E[g(x_{m-1}, x_m) \mid x_{m-1}]$ wtih probability one for any real valued $\mathscr{F}_2$-measurable function g on Ω_2.

THEOREM 2. *Let both μ, ν have stationary transition probabilities. If $\nu_{n\,n}$ is absolutely continuous with respect to $\mu_{n\,n}$ for $n = 0, \pm 1, \pm 2, \cdots$ and $\nu_{1\,2}$ is absolutely continuous with respect to $\mu_{1\,2}$ then $\nu_{m\,n}$ is absolutely continuous with respect to $\mu_{m\,n}$ for $m \leqq n, n = 0, \pm 1, \pm 2, \cdots$ and*

$$f_{m\,n}(x_m, \cdots, x_n) = f_{m\,m}(x_m)\frac{f_{1\,2}(x_m, x_{m+1})}{f_{1\,1}(x_m)} \cdots \cdots \frac{f_{1\,2}(x_{n-1}, x_n)}{f_{1\,1}(x_{n-1})} \tag{9}$$

with μ-probability one.

Proof. By Lemma 1, for any $A \in \mathscr{S}$

$$P_\nu[x_2 \in A \mid x_1] = \frac{E_\mu[I_{x_2 \in A} f_{1\,2}(x_1, x_2) \mid x_1]}{f_{1\,1}(x_1)} \tag{10}$$

with ν-probability one. For any $A, B \in \mathscr{S}$

$$\begin{aligned}
&\nu[x_n \in A, x_{n-1} \in B] \\
&\quad = \int_{[x_{n-1} \in B]} P_\nu[x_n \in A \mid x_{n-1}] d\nu \\
&\quad = \int_{[x_{n-1} \in B]} T^{n-2} P_\nu[x_2 \in A \mid x_1] d\nu \\
&\quad = \int_{[x_{n-1} \in B]} \{T^{n-2} P_\nu[x_2 \in A \mid x_1]\} f_{n-1\,n-1}(x_{n-1}) d\mu\,.
\end{aligned}$$

Hence by (10) and (8)

$$\begin{aligned}
&\nu[x_n \in A, x_{n-1} \in B] \\
&\quad = \int_{[x_{n-1} \in B]} T^{n-2} \left\{ \frac{E_\mu[I_{x_2 \in A} f_{1\,2}(x_1, x_2) \mid x_1]}{f_{1\,1}(x_1)} \right. f_{n-1\,n-1}(x_{n-1}) d\mu \\
&\quad = \int_{[x_{n-1} \in B]} \frac{E_\mu[I_{x_n \in A} f_{1\,2}(x_{n-1}, x_n) \mid x_{n-1}]}{f_{1\,1}(x_{n-1})} f_{n-1\,n-1}(x_{n-1}) d\mu \\
&\quad = \int_{[x_{n-1} \in B]} I_{x_n \in A} f_{n-1\,n-1}(x_{n-1}) \frac{f_{1\,2}(x_{n-1}, x_n)}{f_{1\,1}(x_{n-1})} d\mu \\
&\quad = \int_{[x_n \in A, x_{n-1} \in B]} f_{n-1\,n-1}(x_{n-1}) \frac{f_{1\,2}(x_{n-1}, x_n)}{f_{1\,1}(x_{n-1})} d\mu\,.
\end{aligned}$$

Thus for any $E \in \mathscr{F}_{n-1\,n}$

$$\nu(E) = \int_E f_{n-1\,n-1}(x_{n-1}) \frac{f_{1\,2}(x_{n-1}, x_n)}{f_{1\,1}(x_{n-1})} d\mu . \tag{11}$$

Hence for any integer n, $\nu_{n-1\,n}$ is absolutely continuous with respect to $\mu_{n-1\,n}$ and Theorem 1 is applicable. (11) also implies that

$$f_{n-1\,n}(x_{n-1}, x_n) = f_{n-1\,n-1}(x_{n-1}) \frac{f_{1\,2}(x_{n-1}, x_n)}{f_{1\,1}(x_{n-1})} \tag{12}$$

with μ-probability one. Hence

$$\frac{f_{n-1\,n}(x_{n-1}, x_n)}{f_{n-1\,n-1}(x_{n-1})} = \frac{f_{1\,2}(x_{n-1}, x_n)}{f_{1\,1}(x_{n-1})} \tag{13}$$

with μ-probability one on the set $[f_{n-1\,n-1}(x_{n-1}) > 0]$. However, except that w belongs to a set of μ-probability $0, n > 1, f_{n-1\,n-1}(x_{n-1}(w)) = 0$ imply that $f_{1\,n-1}(x_1(w), \cdots, x_{n-1}(w)) = 0$, hence

$$f_{1\,n-1}(x_1, \cdots, x_{n-1}) \frac{f_{n-1\,n}(x_{n-1}, x_n)}{f_{n-1\,n-1}(x_{n-1})} = f_{1\,n-1}(x_1, \cdots, x_{n-1}) \frac{f_{1\,2}(x_{n-1}, x_n)}{f_{1\,1}(x_{n-1})}$$

with μ-probability one. Thus by (6)

$$f_{1\,n}(x_1, \cdots, x_n) = f_{1\,n-1}(x_1, \cdots, x_{n-1}) \frac{f_{1\,2}(x_{n-1}, x_n)}{f_{1\,1}(x_{n-1})} \tag{14}$$

with μ-probability one. Combining (12) (13) and by induction, if $n > 1$

$$f_{1\,n}(x_1, \cdots, x_n) = f_{1\,1}(x_1) \frac{f_{1\,2}(x_1, x_2)}{f_{1\,1}(x_1)} \cdots \frac{f_{1\,2}(x_{n-1}, x_n)}{f_{1\,1}(x_{n-1})}$$

with μ-probability one. Thus we have proved the theorem for the case that $m = 1$. For the general case the proof is similar.

THEOREM 3. *If μ has stationary transition probabilities and ν is stationary and if*

$$\int |\log f_{m\,m+1}(x_m, x_{m+1})| d\nu < \infty \text{ then}$$

$$\int |\log f_{m\,n}(x_m, \cdots, x_n)| d\nu < \infty \text{ for } n = m, m+1, m+2, \cdots$$

and $n^{-1} \log f_{m\,n}(x_m, \cdots, x_n)$ converges as $n \to \infty$ with ν-probability one and also in $L_1(\nu)$ to a function g with $\int g\, d\nu = a$ where

$$a = \int [\log f_{1\,2}(x_1, x_2) - \log f_{1\,1}(x_1)] d\nu \geqq 0$$

In particular, if ν is ergodic, $g = a$ with ν-probability one.

Proof. We shall first prove the theorem for the case that $m = 1$. Since for any $A \in \mathscr{S}$

$$\nu[x_1 \in A] = \int_{[x_1 \in A]} f_{1\,1}(x_1) d_\mu = \int_{[x_1 \in A]} f_{1\,2}(x_1, x_2) d\mu ,$$

hence

$$E_\mu[f_{1\,2}(x_1, x_2) \mid x_1] = f_{1\,1}(x_1) .$$

Since $\int | \log f_{1\,2}(x_1, x_2) | d\nu < \infty$ hence

$$\int | f_{1\,2}(x_1, x_2) \log f_{1\,2}(x_1, x_2) | d\mu = \int | \log f_{1\,2}(x_1, x_2) | d\nu < \infty .$$

The real valued function $L(\xi) = \xi \log \xi$ defined for all real $\xi \geqq 0$[$L(0)$ is taken to be 0] is convex. By Jensen's inequality for conditional expectations ([2], pp. 33)

$$E_\mu[L\{f_{1\,2}(x_1 x_2)\} \mid x_1] \geqq L\{f_{1\,1}(x_1)\} . \tag{15}$$

By (15) and the fact that $L(\xi)$ is a function bounded below by a constant, we have

$$\int | L\{f_{1\,1}(x_1)\} | d\mu = \int | \log f_{1\,1}(x_1) | d\nu < \infty$$

and

$$\int \log f_{1\,2}(x_1, x_2) d_\nu - \int \log f_{1\,1}(x_1) d_\nu = a \geqq 0 .$$

Now by Theorem 2

$$\log f_{1\,n}(x_1, \cdots, x_n) = \log f_{1\,1}(x_1) + \sum_{i=2}^{n} \{\log f_{1\,2}(x_{i-1}, x_i) - \log f_{1\,1}(x_{i-1})\} .$$

Since ν is stationary, $\log f_{1\,n}(x_1, \cdots, x_n)$ is ν-integrable. Applying the ergodic theorem $n^{-1} \log f_{1\,n}(x_1, \cdots, x_n)$ converges with ν-probability one and also in $L_1(\nu)$ to a function g with

$$\int g d\nu = \int [\log f_{1\,2}(x_1, x_2) - \log f_{1\,1}(x_1)] d\nu = a \geqq 0 .$$

For m being any integer, we only need to mentioned that by (13),

$$\log f_{m\,m+1}(x_m, x_{m+1}) - \log f_{m\,m}(x_m) = \log f_{1\,2}(x_1, x_2) - \log f_{1\,1}(x_1)$$

with ν-probability one and therefore the same conclusion follows with a similar proof.

COROLLARY 1. *Suppose μ, ν satisfy the hypothesis of Theorem 3 for $m = 1$. If ν is ergodic and if there is an $A \in \mathscr{S}$ such that*

$$\nu\{P_\nu[x_2 \in A \mid x_1] \neq P_\mu[x_2 \in A \mid x_1]\} > 0 \tag{16}$$

then ν is singular with respect to μ.

Proof. First we shall show that follows from (16)

$$\mu[f_{1\,1}(x_1) \neq f_{1\,2}(x_1, x_2)] > 0 \ . \tag{17}$$

For, if $f_{1\,1}(x_1) = f_{1\,2}(x_1, x_2)$ with μ-probability one then by Lemma 1
$P_\nu[x_2 \in A \mid x_1] f_{1\,1}(x_1) = P_\mu[x_2 \in A \mid x_1] f_{1\,1}(x_1)$ with μ-probability one. Thus $P_\nu[x_2 \in A \mid x_1] = P_\mu[x_2 \in A \mid x_1]$ with ν-probability one for every $A \in \mathscr{S}$. Now the function $L(\xi) = \xi \log \xi$ is strictly convex, hence it follows from (17) that

$$a = \int [L\{f_{1\,2}(x_1, x_2)\} - L\{f_{1\,1}(x_1)\}] d_\mu > 0 \ .$$

Applying Theorem 3 $f_{1\,n}(x_1, \cdots, x_n) \to \infty$ with ν-probability one as $n \to \infty$. Hence $1/f_n(x_1, \cdots, x_n) \to 0$ with ν-probability one as $n \to \infty$. Let $\mathscr{F}'$ be the σ-algebra generated by $\bigcup_{n=1}^\infty \mathscr{F}_{1\,n}$ and μ', ν' be the contractions of μ, ν to $\mathscr{F}'$ respectively. Since $1/f_{1\,n}(x_1, \cdots, x_n)$ is the derivative of $\nu_{1\,n}$-continuous part of $\mu_{1\,n}$ with respect to $\nu_{1\,n}$, $1/f_{1\,n}(x, \cdots, x_n)$ converges with ν-probability one as $n \to \infty$ to the derivative of ν'-continuous part of μ' with respect to ν' ([2], pp. 343). Now $1/f_{1\,n}(x_1, \cdots, x_n)$ converges to 0 with ν-probability one, hence the ν'-continuous part of μ' is 0 and μ', ν' are mutually singular. Hence μ, ν are mutually singular.

3. Extension to k-Markovian measures. The results of the preceding section can be extended to k-Markovian measures immediately. We shall state the theorems only since the proofs in the preceding section with obvious modifications apply as well.

THEOREM 4. *Let μ, ν be any two k-Markovian measures on $\mathscr{F}$. If $\nu_{n-k\,n}$ is absolutely continuous with respect to $\mu_{n-k,\,n}$ for $n = 0, \pm 1, \pm 2, \cdots$, then $\nu_{m\,n}$ is absolutely continuous with respect to $\mu_{m\,n}$ for $n = 0, \pm 1, \pm 2, \cdots$ and $m \leqq n$ with*

$$f_{m\,n}(x_m, \cdots, x_n) = f_{m\,m+k}(x_m, \cdots, x_{m+k}) \frac{f_{m+1,m+1+k}(x_{m+1}, \cdots, x_{m+1+k})}{f_{m+1,m+k}(x_{m+1}, \cdots, x_{m+k})} \cdots \frac{f_{n-k\,n}(x_{n-k}, \cdots, x_n)}{f_{n-k\,n-1}(x_{n-k}, \cdots, x_{n-1})} \tag{18}$$

with μ-probability one.

THEOREM 5. *Let μ, ν be two k-Markovian measures on $\mathscr{F}$ with stationary transition probabilities. If $\nu_{n-k+1,n}$ is absolutely continuous with respect to $\mu_{n-k+1,n}$ for $n = 0, \pm 1, \pm 2, \cdots$ and $\nu_{1\,k+1}$ is absolutely continuous with respect to $\mu_{1\,k+1}$ then $\nu_{m\,n}$ is absolutely continuous with respect to $\mu_{m\,n}$ for $n = 0, \pm 1, \pm 2, \cdots, m \leqq n$ and*

$$(19)\quad f_{m\,n}(x_m, \cdots, x_n) = f_{m\,m+k-1}(x_m, \cdots, x_{m+k-1}) \frac{f_{1\,k+1}(x_{m+1}, \cdots, x_{m+k+1})}{f_{1,k}(x_{m+1}, \cdots, x_{m+k})} \frac{f_{1\,k+1}(x_{n-k}, \cdots, x_n)}{f_{1\,k}(x_{n-k}, \cdots, {}_{n-1})}$$

with μ-probability one.

THEOREM 6. *Let μ, ν be two k-Markovian measures such that ν is stationary and μ has stationary transition probabilities. If*

$$\int |\log f_{m\,m+k}(x_m, \cdots, x_{m+k})|\, d\nu < \infty$$

then $\int |\log f_{m\,n}(x_m, \cdots, x_n)|\, d\nu < \infty$ for $n = m, m+1, m+2, \cdots$ and $n^{-1}\log f_{m\,n}(x_m, \cdots, x_n)$ converges as $n \to \infty$ with ν-probability one to a function g with $\int g d\nu = a \geqq 0$ where

$$a = \int |\log f_{1\,k+1}(x_1, \cdots, x_{k+1}) - \log f_{1\,k}(x_1, \cdots, x_k)|\, d\nu \geqq 0\ .$$

In particular, if ν is ergodic, $g = a$ with ν-probability one.

COROLLARY 2. *Suppose μ, ν satisfy the hypothesis of Theorem 6 for $m = 1$. If ν is ergodic and if there is a set $A \in \mathscr{S}$ such that*

$$(20)\qquad \nu\{[P_\nu[x_{k+1} \in A \mid x_1, \cdots, x_k] \neq \mathrm{P}_\mu[x_{k+1} \in A] \mid x_1, \cdots, x_k]\} > 0$$

Then ν is singular with respect to μ.

4. A generalization of McMillan's theorem. In the setting of this paper, McMillan's Theorem may be stated as the following. Let X be a finite set ,of K points and $\mathscr{S}$ be the σ-algebra of all subsets of X. Let ν be any stationary probability measure on $\mathscr{F}$ and μ be the measure on $\mathscr{F}$ such that $\mu[X_m = a_0, X_{m+1} = a_1, \cdots, X_n = a_{n-m}] = K^{-(n-m+1)}$ for any intergers m, n and $a_0, a_1 \cdots a_{n-m}$ in X. μ may be described as the equally distributed independent measure on $\mathscr{F}$. Then $n^r f_{1\,n}(x_1, \cdots, x_n)$ converges as $n \to \infty$ in $L_1(\nu)$. In particular, if ν is ergodic, the limit function is equal to $\log K - H$ with ν-probability one where H is the entropy of ν measure [4]. We shall generalize this theorem to the case that X is countable and μ is Markovian with stationary transition probabilities.

THEOREM 7. *Let the totality of elements of X be $a_1, a_2, \cdots$ and ν be a stationary probability measure on $\mathscr{F}$ such that $\int -\log \nu_1(x_1) d\nu < \infty$ where ν_1 is the function defined on X by $\nu_1(a_i) = \nu[x_1 = a_i]$. Let μ be a Markovian measure on $\mathscr{F}$ with stationary transition probabilities. Let $p(a_i, a_j)$ be the value of $P_\mu[x_1 = a_j \mid x_0]$ when $x_0 = a_i$. Let ν_{1n} be absolutely continuous with respect to μ_{1n} for $n = 1, 2, \cdots$. If*

$$\int -\log p(x_1, x_2) d\nu < \infty$$

and $\int |\log f_{11}(x_1)| d\nu < \infty$ then $\int |\log f_{1n}(x_1, \cdots, x_n)| d\nu < \infty$ for $n = 1, 2, \cdots$ and $n^{-1} \log f_{1n}(x_1, \cdots, x_n)$ converges as $n \to \infty$ in $L_1(\nu)$. In particular, if ν is ergodic, the limit is equal to a constant with ν-probability one.

Proof. Let

$$\nu_n(a_{i_1}, a_{i_2}, \cdots, a_{i_n}) = \nu[x_1 = a_{i_1}, x_2 = a_{i_2}, \cdots, x_n = a_{i_n}]$$

and

$$\mu_n(a_{i_1}, a_{i_2}, \cdots, a_{i_n}) = \mu[x_1 = a_{i_1}, x_2 = a_{i_2}, \cdots, x_n = a_{i_n}] .$$

Then

$$f_{1n}(x_1, \cdots, x_n) = \frac{\nu_n(x_1, \cdots, x_n)}{\mu_n(x_1, \cdots, x_n)}$$

with μ-probability one and

$$P_\nu[x_n = a_i \mid x_{n-1}, \cdots, x_1] = \frac{\nu_n(x_1, \cdots, x_{n-1}, a_i)}{\nu_{n-1}(x_1, \cdots, x_{n-1})}$$

with ν-probability one and

$$P_\mu[x_n = a_i \mid x_{n-1}] = \frac{\mu_n(x_1, \cdots, x_{n-1}, a_i)}{\mu_n(x_1, \cdots, x_{n-1})}$$

with μ-probability one. Hence

$$\frac{f_{1n}(x_1, \cdots, x_n)}{f_{1\,n-1}(x_1, \cdots, x_{n-1})} = \sum_{i=1}^{\infty} \frac{P_\nu[x_n = a_i \mid x_{n-1}, \cdots, x_1]}{P_\mu[x_n = a_i \mid x_{n-1}]} I_{x_n = a_i}$$

with ν-probability one and

$$\log \frac{f_{1\,n-1}(x_1, \cdots, x_n)}{f_{1\,n-1}(x_1, \cdots, x_{n-1})} = \sum_{i=1}^{\infty} \log P_\nu[x_n = a_i \mid x_{n-1}, \cdots, x_1] I_{x_n = a_i} \tag{21}$$

$$-\log p(x_{n-1}, x_n)$$

$$= T^n g_n$$

with ν-probability one where

$$g_n = \sum_{i=1}^{\infty} \log P_\nu[x_0 = a_i \mid x_{-1}, \cdots, x_{-(n-1)}] I_{x_0 = a_i} - \log p(x_{-1}, x_0) . \tag{22}$$

We know that $P_\nu[x_0 = a_i \mid x_{-1}, \cdots, x_{-(n-1)}]$ converges with ν-probability one as $n \to \infty$ to $P_\nu[x_0 = a_i \mid x_{-1}, x_{-2}, \cdots]$ by Doob's Martingale Convergence Theorem. Hence $L\{P_\nu[x_0 = a_i \mid x_{-1}, \cdots, x_{-(n-1)}]\}$ converges with ν-probability one to $L\{P_\nu[x_0 = a_i \mid x_{-1}, x_{-2}, \cdots]\}$. But $L(\xi)$ is a bounded function for $0 \leqq \xi \leqq 1$, hence $L\{P_\nu[x_0 = a_i \mid x_{-1}, x_{-(n-1)}\}$ are uniformly bounded with ν-probability one. Hence $L\{P_\nu[x_0 = x_i \mid x_{-1}, \cdots, x_{-(n-1)}]\}$ also converges in $L_1(\nu)$ to $L\{P_\nu[x_0 = a_i \mid x_{-1}, x_{-2}, \cdots]\}$ as $n \to \infty$. Now by Jensen's inequality $\int - L\{P_\nu[x_0 = a_i \mid x_{-1}, \cdots, x_{-(n-1)}]\} d\nu \leqq - L\{P_\nu[x_0 = a_i]\}$. Since

$$\sum_{i=1}^{\infty} - L\{P_\nu[x_0 = a_i]\} = \int - \log \nu_1(x_0) d\nu < \infty$$

$$\sum_{i=1}^{m} - L\{P_\nu[x_0 = a_i \mid x_{-1}, \cdots, x_{-(n-1)}]\}$$

converges in $L_1(\nu)$, as $m \to \infty$, to

$$\sum_{i=1}^{\infty} - L\{P_\nu[x_0 = a_i \mid x_{-1}, \cdots, x_{-(n-1)}]\}$$

uniformly in n. Hence

$$\sum_{i=1}^{\infty} - L\{P_\nu[x_0 = a_i \mid x_{-1}, \cdots, x_{-(n-1)}$$

converges in $L_1(\nu)$ to

$$\sum_{i=1}^{\infty} - L\{P_\nu[x_0 = a_i \mid x_{-1}, x_{-2}, \cdots]\} \text{ as } n \to \infty. \quad \text{Now}$$

$$\int - \sum_{i=1}^{\infty} \log P_\nu[x_0 = a_i \mid x_{-1}, \cdots, x_{-(n-1)}] I_{x_0 = a_i} d\nu$$
$$= \int - \sum_{i=1}^{\infty} L\{P_\nu[x_0 = a_i \mid x_{-1}, \cdots, x_{-(n-1)}]\} d\nu \text{ and}$$

$$\int - \sum_{i=1}^{\infty} \log P_\nu[x_0 = a_i \mid x_{-1}, x_{-2}, \cdots] I_{x_0 = a_i} d\nu$$
$$= \int - \sum_{i=1}^{\infty} L\{P_\nu[x_0 = a_i \mid x_{-1}, x_{-2}, \cdots]\} d\nu, \text{ hence}$$

$$\lim_{n \to \infty} \int - \sum_{i=1}^{\infty} \log P_\nu[x_0 = a_i \mid x_{-1}, \cdots, x_{-(n-1)}] I_{x_0 = a_i} d\nu = \int - \sum_{i=1}^{\infty} \log P_\nu[x_0 = a_i \mid x_{-1}, x_{-2}, \cdots] I_{x_0 = a_i} d\nu . \tag{23}$$

(23) together with the facts that the sequence

$$\left\{-\sum_{i=1}^{\infty}\log P_{\nu}[x_0 = x_i \mid x_{-1}, \cdots, x_{-(n-1)}]I_{x_0=a_i}\right\}$$

is also convergent with ν-probability one and that the functions

$$-\sum_{i=1}^{\infty}\log P_{\nu}[x_0 = x_i \mid x_{-1}, \cdots, x_{-(n-1)}]I_{x_0=a_i}$$

are non negative with ν-probability one imply that

$$\sum_{i=1}^{\infty} P_{\nu}[x_0 = a_i \mid x_{-1}, \cdots, x_{-(n-1)}]I_{x_0=a_i}$$

converges as $n \to \infty$ in $L_1(\nu)$ to

$$\sum_{i=1}^{\infty} P_{\nu}[x_0 = a_i \mid x_{-1}, x_{-2}, \cdots]I_{x_0=a_i} .$$

Thus we have $\{g_n\}$ to be an $L_1(\nu)$ convergent sequence. Let the limit of the sequence *be* h. Let $\bar{h}$ be the $L_1(\nu)$ limit of $1/n(h + Th + \cdots + T^n h)$ as $n \to \infty$. Now by (21)

$$\log f_{1\,2}(x_1, \cdots, x_n) = \log f_{1\,1}(x_1) + \sum_{i=2}^{n} T^i g_i. \quad \text{Thus}$$

$$\begin{aligned}
\int\left|\frac{1}{n}\log f_{1\,n}(x_1, \cdots, x_n) - \bar{h}\right|d\nu
&\leqq \frac{1}{n}\int |\log f_{1\,1}(x_1)|\,d\nu + \int\left|\frac{1}{n}\left(\sum_{i=2}^{n} T^i g_i - \sum_{i=2}^{n} T^i h\right)\right|d\nu \\
&\quad + \int\left|\frac{1}{n}\sum_{i=2}^{n} T^i h - \bar{h}\right|d\nu \\
&= \frac{1}{n}\int\left|\log f_{1\,1}(x_1)\right|d\nu + \frac{1}{n}\sum_{i=2}^{n}\int |g_i - h|\,d\nu \\
&\quad + \int\left|\frac{1}{n}\sum_{i=2}^{n} T^i h - \bar{h}\right|d\nu \to 0 \text{ as } n \to \infty .
\end{aligned}$$

COROLLARY 3. *Under the hypothesis of Theorem* 7, *if* ν *is ergodic and not Markovian then* ν *is singular to* μ.

Proof. If ν is ergodic then the $L_1(\nu)$ limit, $\bar{h}$, of $\{1/n \log f_{1\,n}(x_1, \cdots, x_n)\}$ is equal with ν probability one to

$$\int\sum_{i=1}^{\infty} L\{P_{\nu}[x_0 = a_i \mid x_{-1}, x_{-2}, \cdots]\}d\nu - \int \log p(x_{-1}, x_0)d\nu$$

which is greater or equal to

$$\int \sum_{i=1}^{\infty} L\{P_\nu[x_0 = a_i \mid x_{-1}, x_{-2}]\} d\nu - \int \log p(x_{-1}, x_0) d\nu .$$

Hence by (21)

$$\begin{aligned} \bar{h} &\geqq \int \sum_{i=1}^{\infty} \log P_\nu[x_0 = a_i \mid x_{-1}, x_{-2}] I_{x_0=a_i} d\nu - \int \log p(x_{-1}, x_0) d\nu \\ &= \int \log f_{1\,3}(x_1, x_2, x_2) d\nu - \int \log f_{1\,2}(x_1, x_2) d\nu . \end{aligned}$$

However $\int \log f_{1\,3}(x_1, x_2, x_3) d\nu - \int \log f_{1\,2}(x_1, x_2) d\nu = 0$ if and only if

(24) $$\mu[f_{1\,2}(x_1, x_2) \neq f_{1\,3}(x_1, x_2, x_3)] = 0 .$$

(24) implies that

$$P_\nu[x_3 \in A \mid x_1, x_2] = P_\mu[x_3 \in A \mid x_1, x_2]$$

with ν-probability one for any $A \in \mathscr{S}$. This is impossible since μ is Markovian and ν is not. Hence $\bar{h} > 0$ with ν-probability one. Hence $f_{1\,n}(x_1, \cdots, x_n) \to \infty$ with ν probability one and ν is singular to μ by the same argument used in the proof in Corollary 1.

The extensions of Theorem 7 and Corollary 3 to k-Markovian μ is obvious.

References

1. Erik Sparre Anderson and Borge Jessen, *Some limit theorems in an abstract set*, Danske Vid. Selsk. Nat.-Fys. Medd. **22**, no. 14 (1946). **22**, No. 14 (1946).
2. J. L. Doob, *Stochastic Processes*, John Wiley and Sons, Inc., New York.
3. Amiel Feinstein, *Foundations of Information Theory*, McGraw-Hill Inc. New York, Toronto, London.
4. B. McMillan, *The basic theorems of information theory*, Annals of Math. Statistics, **24** (1953), 196–219.
5. C. E. Shannon, *The mathematical theory of communication*, Bell Syst. Techn. Journ. **27** (1948), 379–423, 623–456.

WAYNE STATE UNIVERSITY