

Pacific Journal of Mathematics

ASYMPTOTIC PROPERTIES OF DERIVATIVES OF STATIONARY MEASURES

SHU-TEH CHEN MOY

ASYMPTOTIC PROPERTIES OF DERIVATIVES OF STATIONARY MEASURES

SHU-TEH C. MOY

1. Introduction. Let X be a non-empty set and \mathcal{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 $n = 0, \pm 1, \pm 2, \dots$ and the infinite product σ -algebra $\mathcal{F} = \prod_{n=-\infty}^{\infty} \mathcal{S}_n$ where $\mathcal{S}_n = \mathcal{S}$ for $n = 0, \pm 1, \pm 2, \dots$. Elements of Ω are bilateral infinite sequences $\{\dots, x_{-1}, x_0, x_1, \dots\}$ with $x_n \in X$. Let us denote the elements of Ω by w . If $w = \{\dots, x_{-1}, x_0, x_1, \dots\}$ x_n is called the n th 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 n th 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 \mathcal{F} . For $n = 1, 2, \dots$ let $\Omega_n = \prod_{i=1}^n X_i$ where $X_i = X, i = 1, 2, \dots, n$ and $\mathcal{F}_n = \prod_{i=1}^n \mathcal{S}_i$ where $\mathcal{S}_i = \mathcal{S}, i = 1, 2, \dots, n$. Then $\Omega_1 = X$ and $\mathcal{F}_1 = \mathcal{S}$. Let $\mathcal{F}_{m,n}, m \leq n, n = 0, \pm 1, \pm 2, \dots$, be the σ -algebra of subsets of Ω consisting of sets of the form

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

Where $E \in \mathcal{F}_{n-m+1}$. Then $\mathcal{F}_{m,n} \subset \mathcal{F}_{m,n+1} \subset \mathcal{F}$. Let $\mu_{m,n}, \nu_{m,n}$ be the contractions of μ, ν , respectively to $\mathcal{F}_{m,n}$. 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, \dots, x_n and shall be designated by $f_{m,n}(x_m, \dots, x_n)$. Since $f_{m,n}(x_m, \dots, x_n)$ is positive with ν -probability one $1/f_{m,n}(x_m, \dots, x_n)$ is well defined with ν -probability one. We shall let the function $1/f_{m,n}(x_m, \dots, x_n)$ take on the value 0 when $f_{m,n}(x_m, \dots, x_n) \leq 0$. Thus $1/f_{m,n}(x_m, \dots, x_n)$ is well defined everywhere. In fact $1/f_{m,n}(x_m, \dots, x_n)$ 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) $1/f_{m,n}(x_m, \dots, x_n)$ converges with ν -probability one as $n \rightarrow \infty$. 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 \mathcal{S} , any σ -algebra of subsets of X . It will be proved that $n^{-1} \log f_{m,n}(x_m, \dots, x_n)$ converges as $n \rightarrow \infty$ 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 $X, \mathcal{S}, \Omega, \mathcal{F}, \Omega_n, \mathcal{F}_n, \mathcal{F}_{m,n}, \mu_{m,n}, \nu_{m,n} f_{m,n}(x_m, \dots, x_n)$ be as in §1. $x_n, n = 0, \pm 1, \pm 2, \dots$, are considered as functions or random variables on Ω to X . Notations for conditional probabilities 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 $A \in \mathcal{S}, m < n, n = 0 \pm 1, \pm 2, \dots$,

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

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

$$\begin{aligned} I_E(w) &= 1 \text{ if } w \in E \\ &= 0 \text{ if } w \notin E. \end{aligned}$$

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

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

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

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

On the other hand

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

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

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

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

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

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

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 $n = 0, \pm 1, \pm 2, \dots$ then $\nu_{m, n}$ is absolutely continuous with respect to $\mu_{m, n}$ for $n = 0, \pm 1, \pm 2, \dots$ and $m \leq n$ with*

$$(5) \quad \begin{aligned} f_{m, n}(x_m, \dots, x_n) &= f_{m, m+1}(x_m, x_{m+1}) \frac{f_{m+1, m+2}(x_{m+1}, x_{m+2})}{f_{m+1, m+1}(x_{m+1})} \\ &\quad \dots \frac{f_{n-1, n}(x_{n-1}, x_n)}{f_{n-1, n-1}(x_{n-1})} \end{aligned}$$

with μ -probability one.

Proof. We shall prove the theorem for the case that $m = 1, n = 2, 3, \dots$. 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 $\nu_{1, k}$ ($k \geq 2$) is absolutely continuous with respect to $\mu_{1, k}$ and $f_{1, k}(x_1, \dots, x_k)$ is given by (5) with μ -probability one. For any $A \in \mathcal{S}, B \in \mathcal{F}_k$

$$\begin{aligned} & \nu[x_{k+1} \in A, (x_1, \dots, x_k) \in B] \\ &= \int_{[(x_1, \dots, x_k) \in B]} P_{\nu}[x_{k+1} \in A | x_1, \dots, x_k] d\nu. \end{aligned}$$

Since ν is Markovian and by (4)

$$\begin{aligned} & \nu[x_{k+1} \in A, (x_1, \dots, x_k) \in B] \\ &= \int_{[(x_1, \dots, x_k) \in B]} P_{\nu}[x_{k+1} \in A | x_k] d\nu \end{aligned}$$

$$\begin{aligned}
&= \int_{[(x_1, \dots, x_k) \in B]} \frac{E_\mu[I_{x_{k+1} \in A} f_{k, k+1}(x_k, x_{k+1}) | x_k]}{f_{k, k}(x_k)} d\nu \\
&= \int_{[(x_1, \dots, x_k) \in B]} \frac{E_\mu[I_{x_{k+1} \in A} f_{k, k+1}(x_k, x_{k+1}) | x_k]}{f_{k, k}(x_k)} f_{1, k}(x_1, \dots, x_k) d\mu.
\end{aligned}$$

Since μ is Markovian

$$\begin{aligned}
&E_\mu[I_{x_{k+1} \in A} f_{k, k+1}(x_k, x_{k+1}) | x_k] \\
&= E_\mu[I_{x_{k+1} \in A} f_{k, k+1}(x_k, x_{k+1}) | x_1, \dots, x_k]
\end{aligned}$$

with μ -probability one. Hence

$$\begin{aligned}
&\nu[x_{k+1} \in A, (x_1, \dots, x_k) \in B] \\
&= \int_{[(x_1, \dots, x_k) \in B]} E_\mu\left[I_{x_{k+1} \in A} \frac{f_{k, k+1}(x_k, x_{k+1})}{f_{k, k}(x_k)} f_{1, k}(x_1, \dots, x_k) | x_1, \dots, x_k\right] d\mu \\
&= \int_{[(x_1, \dots, x_k) \in B]} I_{x_{k+1} \in A} f_{1, k}(x_1, \dots, x_k) \frac{f_{k, k+1}(x_k, x_{k+1})}{f_{k, k}(x_k)} d\mu.
\end{aligned}$$

Hence

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

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

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

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

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

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 \mathcal{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 $A \in \mathcal{S}$ and any n

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

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

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

(8) $E[g(x_{n-1}, x_n) | x_{n-1}] = T^{n-m}E[g(x_{m-1}, x_m) | x_{m-1}]$ with probability one for any real valued \mathcal{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 $n = 0, \pm 1, \pm 2, \dots$ 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 $m \leq n, n = 0, \pm 1, \pm 2, \dots$ and*

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

with μ -probability one.

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

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

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

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

Hence by (10) and (8)

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

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

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

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

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

with μ -probability one. Hence

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

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 $f_{1 \ n-1}(x_1(w), \dots, x_{n-1}(w)) = 0$, hence

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

with μ -probability one. Thus by (6)

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

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

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

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*

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

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

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

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

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 $A \in \mathcal{S}$

$$\nu[x_1 \in A] = \int_{[x_1 \in A]} f_{11}(x_1) d\mu = \int_{[x_1 \in A]} f_{12}(x_1, x_2) d\mu,$$

hence

$$E_\mu[f_{12}(x_1, x_2) | x_1] = f_{11}(x_1).$$

Since $\int |\log f_{12}(x_1, x_2)| d\nu < \infty$ hence

$$\int |f_{12}(x_1, x_2) \log f_{12}(x_1, x_2)| d\mu = \int |\log f_{12}(x_1, x_2)| d\nu < \infty.$$

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

$$(15) \quad E_\mu[L\{f_{12}(x_1, x_2)\} | x_1] \geq L\{f_{11}(x_1)\}.$$

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

$$\int |L\{f_{11}(x_1)\}| d\mu = \int |\log f_{11}(x_1)| d\nu < \infty$$

and

$$\int \log f_{12}(x_1, x_2) d\nu - \int \log f_{11}(x_1) d\nu = a \geq 0.$$

Now by Theorem 2

$$\log f_{1n}(x_1, \dots, x_n) = \log f_{11}(x_1) + \sum_{i=2}^n \{\log f_{12}(x_{i-1}, x_i) - \log f_{11}(x_{i-1})\}.$$

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

$$\int g d\nu = \int [\log f_{12}(x_1, x_2) - \log f_{11}(x_1)] d\nu = a \geq 0.$$

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

$$\log f_{m+1}(x_m, x_{m+1}) - \log f_m(x_m) = \log f_{12}(x_1, x_2) - \log f_{11}(x_1)$$

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 $A \in \mathcal{S}$ such that

$$(16) \quad \nu\{P_\nu[x_2 \in A | x_1] \neq P_\mu[x_2 \in A | x_1]\} > 0$$

then ν is singular with respect to μ .

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

$$(17) \quad \mu[f_{11}(x_1) \neq f_{12}(x_1, x_2)] > 0.$$

For, if $f_{11}(x_1) = f_{12}(x_1, x_2)$ with μ -probability one then by Lemma 1

$P_\nu[x_2 \in A | x_1]f_{11}(x_1) = P_\mu[x_2 \in A | x_1]f_{11}(x_1)$ with μ -probability one. Thus $P_\nu[x_2 \in A | x_1] = P_\mu[x_2 \in A | x_1]$ with ν -probability one for every $A \in \mathcal{S}$. Now the function $L(\xi) = \xi \log \xi$ is strictly convex, hence it follows from (17) that

$$a = \int [L\{f_{12}(x_1, x_2)\} - L\{f_{11}(x_1)\}]d_\mu > 0.$$

Applying Theorem 3 $f_{1n}(x_1, \dots, x_n) \rightarrow \infty$ with ν -probability one as $n \rightarrow \infty$. Hence $1/f_n(x_1, \dots, x_n) \rightarrow 0$ with ν -probability one as $n \rightarrow \infty$. Let \mathcal{F}' be the σ -algebra generated by $\bigcup_{n=1}^\infty \mathcal{F}_{1n}$ and μ', ν' be the contractions of μ, ν to \mathcal{F}' respectively. Since $1/f_{1n}(x_1, \dots, x_n)$ is the derivative of ν_{1n} -continuous part of μ_{1n} with respect to ν_{1n} , $1/f_{1n}(x_1, \dots, x_n)$ converges with ν -probability one as $n \rightarrow \infty$ to the derivative of ν' -continuous part of μ' with respect to ν' ([2], pp. 343). Now $1/f_{1n}(x_1, \dots, x_n)$ 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 \mathcal{F} . If $\nu_{n-k, n}$ is absolutely continuous with respect to $\mu_{n-k, n}$ for $n = 0, \pm 1, \pm 2, \dots$, then $\nu_{m, n}$ is absolutely continuous with respect to $\mu_{m, n}$ for $n = 0, \pm 1, \pm 2, \dots$ and $m \leq n$ with

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

with μ -probability one.

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

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

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

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

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

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

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 $A \in \mathcal{S}$ such that

$$(20) \quad \nu\{[P_\nu[x_{k+1} \in A | x_1, \dots, x_k] \neq P_\mu[x_{k+1} \in A] | x_1, \dots, x_k]\} > 0$$

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 \mathcal{S} be the σ -algebra of all subsets of X . Let ν be any stationary probability measure on \mathcal{S} and μ be the measure on \mathcal{S} such that $\mu[X_m = a_0, X_{m+1} = a_1, \dots, X_n = a_{n-m}] = K^{-(n-m+1)}$ for any integers m, n and a_0, a_1, \dots, a_{n-m} in X . μ may be described as the equally distributed independent measure on \mathcal{S} . Then $n^{-1} \log f_{1, n}(x_1, \dots, x_n)$ converges as $n \rightarrow \infty$ 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, \dots and ν be a stationary probability measure on \mathcal{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 \mathcal{F} with stationary transition probabilities. Let $p(a_i, a_j)$ be the value of $P_\mu[x_1 = a_j | x_0 = a_i]$ when $x_0 = a_i$. Let ν_{1n} be absolutely continuous with respect to μ_{1n} for $n = 1, 2, \dots$. If*

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

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

Proof. Let

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

and

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

Then

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

with μ -probability one and

$$p(a_i | x_{n-1}, \dots, x_1) = \frac{\nu_n(x_1, \dots, x_{n-1}, a_i)}{\nu_{n-1}(x_1, \dots, x_{n-1})}$$

with ν -probability one and

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

with μ -probability one. Hence

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

with ν -probability one and

$$\begin{aligned} (21) \quad \log \frac{f_{1n-1}(x_1, \dots, x_n)}{f_{1n-1}(x_1, \dots, x_{n-1})} &= \sum_{i=1}^{\infty} \log P_\nu[x_n = a_i | x_{n-1}, \dots, x_1] I_{x_n = a_i} \\ &\quad - \log p(x_{n-1}, x_n) \\ &= T^n g_n \end{aligned}$$

with ν -probability one where

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

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

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

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

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

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

uniformly in n . Hence

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

converges in $L_1(\nu)$ to

$$\sum_{i=1}^{\infty} -L\{P_\nu[x_0 = a_i | x_{-1}, x_{-2}, \dots]\} \text{ as } n \rightarrow \infty. \text{ Now}$$

$$\int -\sum_{i=1}^{\infty} \log P_\nu[x_0 = a_i | x_{-1}, \dots, x_{-(n-1)}] I_{x_0=a_i} d\nu \\ = \int -\sum_{i=1}^{\infty} L\{P_\nu[x_0 = a_i | x_{-1}, \dots, x_{-(n-1)}]\} d\nu \text{ and} \\ \int -\sum_{i=1}^{\infty} \log P_\nu[x_0 = a_i | x_{-1}, x_{-2}, \dots] I_{x_0=a_i} d\nu \\ = \int -\sum_{i=1}^{\infty} L\{P_\nu[x_0 = a_i | x_{-1}, x_{-2}, \dots]\} d\nu, \text{ hence}$$

$$(23) \quad \lim_{n \rightarrow \infty} \int -\sum_{i=1}^{\infty} \log P_\nu[x_0 = a_i | x_{-1}, \dots, x_{-(n-1)}] I_{x_0=a_i} d\nu \\ = \int -\sum_{i=1}^{\infty} \log P_\nu[x_0 = a_i | x_{-1}, x_{-2}, \dots] I_{x_0=a_i} d\nu.$$

(23) together with the facts that the sequence

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

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

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

are non negative with ν -probability one imply that

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

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

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

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 $1/n(h + Th + \dots + T^n h)$ as $n \rightarrow \infty$. Now by (21)

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

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

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 $\{1/n \log f_{1,n}(x_1, \dots, x_n)\}$ is equal with ν probability one to

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

which is greater or equal to

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

Hence by (21)

$$\begin{aligned} \bar{h} &\geq \int \sum_{i=1}^{\infty} \log P_{\nu}[x_0 = a_i | x_{-1}, x_{-2}] I_{x_0=a_i} d\nu - \int \log p(x_{-1}, x_0) d\nu \\ &= \int \log f_{13}(x_1, x_2, x_3) d\nu - \int \log f_{12}(x_1, x_2) d\nu. \end{aligned}$$

However $\int \log f_{13}(x_1, x_2, x_3) d\nu - \int \log f_{12}(x_1, x_2) d\nu = 0$ if and only if

$$(24) \quad \mu[f_{12}(x_1, x_2) \neq f_{13}(x_1, x_2, x_3)] = 0.$$

(24) implies that

$$P_{\nu}[x_3 \in A | x_1, x_2] = P_{\mu}[x_3 \in A | x_1, x_2]$$

with ν -probability one for any $A \in \mathcal{S}$. This is impossible since μ is Markovian and ν is not. Hence $\bar{h} > 0$ with ν -probability one. Hence $f_{1n}(x_1, \dots, x_n) \rightarrow \infty$ 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 Andersen 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

PACIFIC JOURNAL OF MATHEMATICS

EDITORS

DAVID GILBARG

Stanford University
Stanford, California

F. H. BROWNELL

University of Washington
Seattle 5, Washington

A. L. WHITEMAN

University of Southern California
Los Angeles 7, California

L. J. PAIGE

University of California
Los Angeles 24, California

ASSOCIATE EDITORS

E. F. BECKENBACH

T. M. CHERRY

D. DERRY

E. HEWITT

A. HORN

L. NACHBIN

M. OHTSUKA

H. L. ROYDEN

M. M. SCHIFFER

E. SPANIER

E. G. STRAUS

F. WOLF

SUPPORTING INSTITUTIONS

UNIVERSITY OF BRITISH COLUMBIA
CALIFORNIA INSTITUTE OF TECHNOLOGY
UNIVERSITY OF CALIFORNIA
MONTANA STATE UNIVERSITY
UNIVERSITY OF NEVADA
NEW MEXICO STATE UNIVERSITY
OREGON STATE COLLEGE
UNIVERSITY OF OREGON
OSAKA UNIVERSITY
UNIVERSITY OF SOUTHERN CALIFORNIA

STANFORD UNIVERSITY
UNIVERSITY OF TOKYO
UNIVERSITY OF UTAH
WASHINGTON STATE COLLEGE
UNIVERSITY OF WASHINGTON
* * *

AMERICAN MATHEMATICAL SOCIETY
CALIFORNIA RESEARCH CORPORATION
HUGHES AIRCRAFT COMPANY
SPACE TECHNOLOGY LABORATORIES
NAVAL ORDNANCE TEST STATION

Mathematical papers intended for publication in the *Pacific Journal of Mathematics* should be typewritten (double spaced), and the author should keep a complete copy. Manuscripts may be sent to any one of the four editors. All other communications to the editors should be addressed to the managing editor, L. J. Paige at the University of California, Los Angeles 24, California.

50 reprints per author of each article are furnished free of charge; additional copies may be obtained at cost in multiples of 50.

The *Pacific Journal of Mathematics* is published quarterly, in March, June, September, and December. The price per volume (4 numbers) is \$12.00; single issues, \$3.50. Back numbers are available. Special price to individual faculty members of supporting institutions and to individual members of the American Mathematical Society: \$4.00 per volume; single issues, \$1.25.

Subscriptions, orders for back numbers, and changes of address should be sent to Pacific Journal of Mathematics, 2120 Oxford Street, Berkeley 4, California.

Printed at Kokusai Bunken Insatsusha (International Academic Printing Co., Ltd.), No. 6, 2-chome, Fujimi-cho, Chiyoda-ku, Tokyo, Japan.

PUBLISHED BY PACIFIC JOURNAL OF MATHEMATICS, A NON-PROFIT CORPORATION

The Supporting Institutions listed above contribute to the cost of publication of this Journal, but they are not owners or publishers and have no responsibility for its content or policies.

Pacific Journal of Mathematics

Vol. 10, No. 4

December, 1960

M. Altman, <i>An optimum cubically convergent iterative method of inverting a linear bounded operator in Hilbert space</i>	1107
Nesmith Cornett Ankeny, <i>Criterion for rth power residuacity</i>	1115
Julius Rubin Blum and David Lee Hanson, <i>On invariant probability measures I</i>	1125
Frank Featherstone Bonsall, <i>Positive operators compact in an auxiliary topology</i>	1131
Billy Joe Boyer, <i>Summability of derived conjugate series</i>	1139
Delmar L. Boyer, <i>A note on a problem of Fuchs</i>	1147
Hans-Joachim Bremermann, <i>The envelopes of holomorphy of tube domains in infinite dimensional Banach spaces</i>	1149
Andrew Michael Bruckner, <i>Minimal superadditive extensions of superadditive functions</i>	1155
Billy Finney Bryant, <i>On expansive homeomorphisms</i>	1163
Jean W. Butler, <i>On complete and independent sets of operations in finite algebras</i>	1169
Lucien Le Cam, <i>An approximation theorem for the Poisson binomial distribution</i>	1181
Paul Civin, <i>Involutions on locally compact rings</i>	1199
Earl A. Coddington, <i>Normal extensions of formally normal operators</i>	1203
Jacob Feldman, <i>Some classes of equivalent Gaussian processes on an interval</i>	1211
Shaul Foguel, <i>Weak and strong convergence for Markov processes</i>	1221
Martin Fox, <i>Some zero sum two-person games with moves in the unit interval</i>	1235
Robert Pertsch Gilbert, <i>Singularities of three-dimensional harmonic functions</i>	1243
Branko Grünbaum, <i>Partitions of mass-distributions and of convex bodies by hyperplanes</i>	1257
Sidney Morris Harmon, <i>Regular covering surfaces of Riemann surfaces</i>	1263
Edwin Hewitt and Herbert S. Zuckerman, <i>The multiplicative semigroup of integers modulo m</i>	1291
Paul Daniel Hill, <i>Relation of a direct limit group to associated vector groups</i>	1309
Calvin Virgil Holmes, <i>Commutator groups of monomial groups</i>	1313
James Fredrik Jakobsen and W. R. Utz, <i>The non-existence of expansive homeomorphisms on a closed 2-cell</i>	1319
John William Jewett, <i>Multiplication on classes of pseudo-analytic functions</i>	1323
Helmut Klingen, <i>Analytic automorphisms of bounded symmetric complex domains</i>	1327
Robert Jacob Koch, <i>Ordered semigroups in partially ordered semigroups</i>	1333
Marvin David Marcus and N. A. Khan, <i>On a commutator result of Taussky and Zassenhaus</i>	1337
John Glen Marica and Steve Jerome Bryant, <i>Unary algebras</i>	1347
Edward Peter Merkes and W. T. Scott, <i>On univalence of a continued fraction</i>	1361
Shu-Teh Chen Moy, <i>Asymptotic properties of derivatives of stationary measures</i>	1371
John William Neuberger, <i>Concerning boundary value problems</i>	1385
Edward C. Posner, <i>Integral closure of differential rings</i>	1393
Marian Reichaw-Reichbach, <i>Some theorems on mappings onto</i>	1397
Marvin Rosenblum and Harold Widom, <i>Two extremal problems</i>	1409
Morton Lincoln Slater and Herbert S. Wilf, <i>A class of linear differential-difference equations</i>	1419
Charles Robson Storey, Jr., <i>The structure of threads</i>	1429
J. François Treves, <i>An estimate for differential polynomials in $\partial/\partial z_1, \dots, \partial/\partial z_n$</i>	1447
J. D. Weston, <i>On the representation of operators by convolutions integrals</i>	1453
James Victor Whittaker, <i>Normal subgroups of some homeomorphism groups</i>	1469